The classical genetic and genomic approach to the pathogenesis of primary ciliary dyskinesia

Geremek, Maciej

The classical genetic and genomic approach to the pathogenesis of primary ciliary dyskinesia

Maciej Geremek

RIJKSUNIVERSITEIT GRONINGEN

The classical genetic and genomic approach to the pathogenesis of primary ciliary dyskinesia

Proefschrift

ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
woensdag 13 juni 2012
om 11.00 uur

door

Maciej Geremek

geboren op 20 april 1976
te Warszawa, Polen

Promotores:	Prof. dr. C. Wijmenga Prof. dr. M. Witt
Beoordelingscommissie:	Prof. dr. C.M.A. van Ravenswaaij-Arts Prof. dr. H.J.M. Groen Prof. dr. E.H.H.M. Rings Prof. dr. R.J. Sinke

Table of contents

Chapter 1

Introduction

I. The aims and outline of the thesis

The number of diseases caused by mutations in a single gene is estimated to be over 6,000. The Online Mendelian Inheritance in Man (OMIM) database includes 2,715 phenotypes with a known molecular basis.[1] Thus, the genetic basis of at least half of the known Mendelian disorders remains a mystery. Technological and scientific progress has recently enabled researchers to focus on complex disorders that are more common and therefore represent a bigger social problem. Although single gene diseases are rare, in total they affect around 1% of the population. Knowing the genetic basis of these diseases will help to diagnose them, to design evidence-based therapeutic approaches, and to perform clinical trials with new therapies.

Primary ciliary dyskinesia (PCD) is a rare genetic disease characterized by recurrent respiratory tract infections, bronchiectasis and infertility. Pulmonary symptoms occur in PCD because of a lack of an efficient mucociliary clearance caused by the kinetic dysfunction of respiratory cilia. Male infertility is caused by immotility of flagella in spermatozoids. Female subfertility is caused by dysfunction of motile cilia in the fallopian tubes, which help to move the ovum toward the uterus. The disease is genetically heterogeneous - mutations in many different genes have a similar functional effect on the motility of cilia and cause identical or similar phenotypes. Despite the fact that mutations in twelve genes have already been found in PCD patients, the genetic background is still unknown in most cases.

We undertook two different strategies in an attempt to elucidate the pathogenesis of PCD (figure 1). We applied a classical genetic approach to look for new PCD loci and genes using linkage analysis and candidate gene sequencing. We also used a functional genomics approach to find new disease-associated pathways and functionally related genes in a genome-wide gene expression experiment.

The primary research strategies for identifying genes causing Mendelian disorders have been positional cloning, physical mapping, and candidate-gene

sequencing. Linkage analysis reveals polymorphic genetic markers that are inherited through family pedigrees together with the disease. We performed linkage analysis on several small families with PCD and tried to deal with the genetic heterogeneity by phenotypical stratification of the patients. Most of the families were of Polish origin. We found significant linkage to chromosome 15q for a subtype of PCD called Kartagener syndrome. We followed up this finding by sequencing candidate genes in the region. The linkage analysis is described in chapter 2, while the fine mapping of the linkage region and the candidate-gene sequencing is described in chapter 3.

Mendelian diseases are caused by mutations in a single gene. However, the mutations can affect the function of many other genes by altering the regulatory networks at the genomic level. The complexity of the ciliary structure and the multitude of genes involved in its functioning makes it extremely difficult to prioritize the candidate disease genes. We performed a pilot gene expression study in bronchial biopsies from PCD patients to find out if expression analysis can be useful in a search for new ciliary genes or pathways important to the disease patho-mechanisms. We found significant down-regulation of ciliary genes in these biopsies. Moreover, we saw that ciliary genes have a highly correlated expression pattern across disease samples. We extracted the pattern by clustering analysis and used it to identify new ciliary genes, representing new PCD candidate genes spread throughout the genome. The the clustering analysis of the ciliary genes correlated expression is described in chapter 4, and the case-control comparison of the gene expression profiles in chapter 5.

The main obstacle to finding frequently mutated PCD genes, seems to be the disease's extensive genetic heterogeneity in the absence of phenotypical diversity. The final chapter discusses this problem in the light of new technological developments in genomics. I also outline some topics for future research and the possible clinical relevance of medical genetics research for PCD patients.

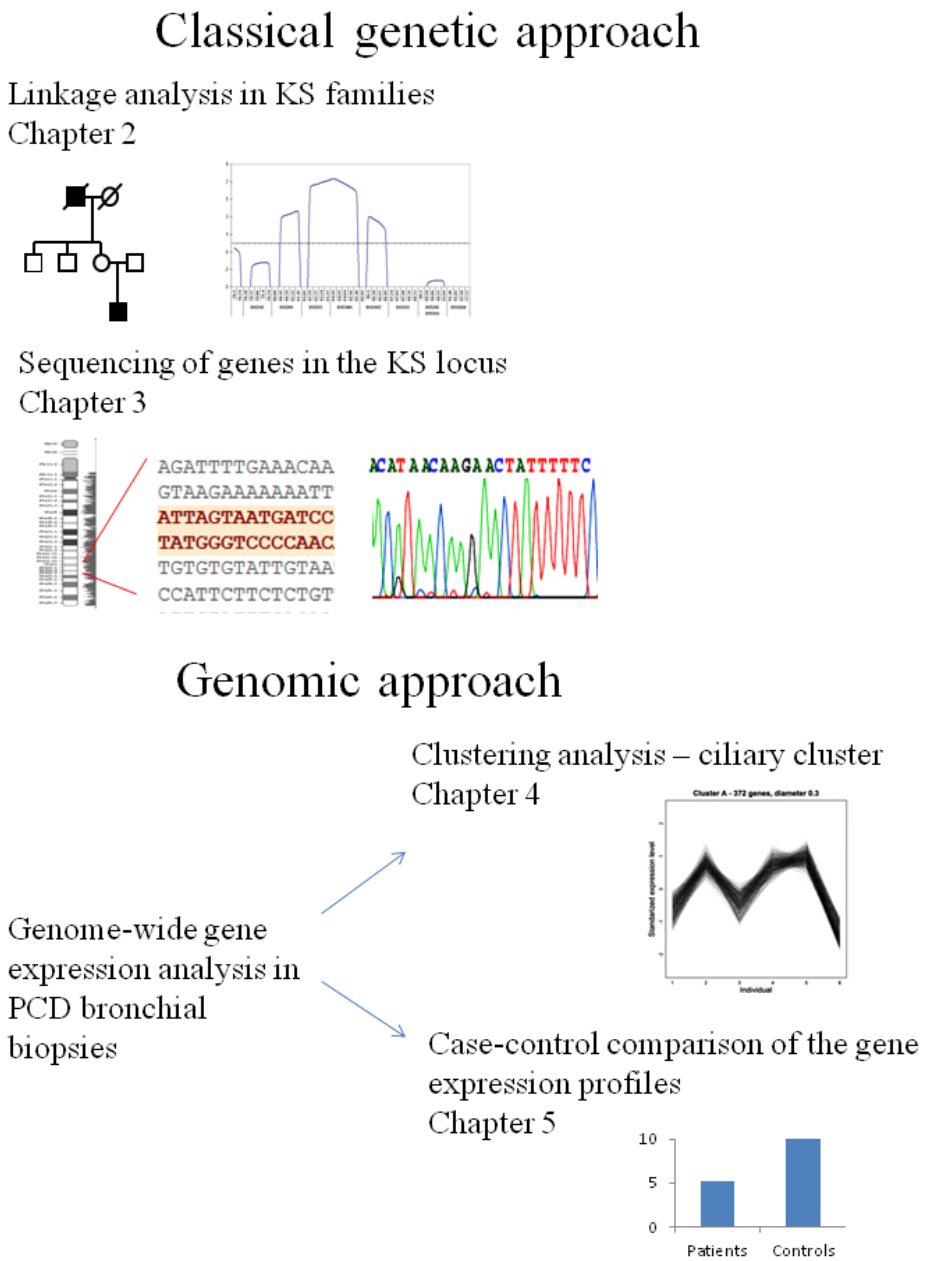

Figure 1. Outline of this thesis

Reference

1. Online Mendelian Inheritance in Man, OMIM (TM). McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD). World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/

II. Primary ciliary dyskinesia: genes, candidate genes and chromosomal regions

Maciej Geremek, Michał Witt

J Appl Genet 2004;45:347-61

Abstract

Primary ciliary dyskinesia (PCD) is a multisystem disease characterized by recurrent respiratory tract infections, sinusitis, bronchiectasis and male subfertility, associated in about 50% patients with *situs inversus totalis* (the Kartagener syndrome). The disease phenotype is caused by ultrastructural defects of respiratory cilia and sperm tails. PCD is a heterogenetic disorder, usually inherited as an autosomal recessive trait. So far, muta- tions in two human genes have been proved to cause the disease. However, the pathogenetics of most PCD cases remains unsolved. In this review, the disease pathomechanism is discussed along with the genes that are or may be involved in the pathogenesis of primary ciliary dyskinesia and the Kartagener syndrome.

Introduction

Primary ciliary dyskinesia (PCD), formerly known as the immotile cilia syndrome (ICS), belongs to a relatively small group of human genetic disorders caused by dysfunctions of particular organelles, analogous to mitochondrial, lysosomal or peroxisomal diseases. In 1933, four cases of coexisting *situs inversus*, bronchiectasis, and chronic sinusitis were described.[1] Later this set of symptoms was named the Kartagener triad or Kartagener syndrome. In the 1970's it was found that the mobility of flagella in the sperm and cilia in the respiratory endothelium of males with the Kartagener syndrome was impaired, causing both chronic bronchopulmonary symptoms and male infertility. This mobility impairment was due to ciliary ultrastructure defects that could be observed under the electron microscope.[2, 3] Afzelius et al.[2] coined the term "immotile cilia syndrome", which has been replaced by the currently used "primary ciliary dyskinesia".

All forms of PCD are characterized by dysmotility or immotility of cilia/flagella in the airway epithelial cells, spermatozoa, and other ciliated cells of

the body.[4, 5, 6, 7] To differentiate this inherited syndrome from the ultrastructural ciliary abnormalities observed during or after injuries (e.g. respiratory infections), the term "secondary ciliary dyskinesia" (SCD) is used for the acquired forms of the disorder.[8, 9] The Kartagener syndrome (KS), where *situs inversus*, a mirror reversal of the usual left-right asymmetry of the abdominal and thoracic internal organ locations coexists with other symptoms, is considered a subtype of PCD.[4]

Cilia

Cilia and flagella, organelles extending from the cell surface with which they share the cell membrane, are present on protozoan, animal and some plant cells. Flagella provide motility to unicellular organisms or cells such as spermatozoa. Cilia propel fluid over the epithelia of various organs.

The main body of the organellum, called the axoneme, consists of 9 peripheral microtubular doublets surrounding a central pair of microtubules (Figure 1). Microtubular doublets are composed of heterodimers of and tubulin assembled into 13 and 11 protofilaments (in microtubules A and B, respectively). The 13 protofilaments of the central microtubular pair are spatially oriented with respect to the central pair of the adjacent cilia. Radial spokes connect peripheral and central microtubules, while peripheral doublets are connected with each other by nexin links. Tubulins in the microtubules are accompanied by microtubule-associated proteins (MAPs). The most extensively studied MAPs are dyneins, which form inner (IDA) and outer (ODA) dynein arms extending from microtubul A of each doublet. In *Chlamydomonas reinhardtii* IDAs are arranged along the longitudinal axis of microtubule A, with a 96 nm long repeat pattern. Seven distinct IDA isoformes containing eleven dynein heavy chains, numerous intermediate and light chains are known. In the same alga a single isoform of ODA containing three heavy, two intermediate and eight light chains, is repeated every 24 nm along the axoneme. The axoneme at its basis is connected with the kinetosome (basal body) consisting of nine microtubular triplets, two of them continuous with axonemal microtubule A and B. The tubulin-dynein complex, typical for the inner structure

of cilia and flagella, constitutes a functional equivalent of the myofibrillar actin-myosin complex. In the presence of ATP the dyneins, attached on one end to microtubule A, bind to microtubule B of the adjacent doublet and move towards its minus end, providing a mechanism for sliding between the two doublets.

Single cilia lacking the central pair of microtubules (the 9 + 0 structure), called also primary cilia or monocilia, are present on the surface of ganglion cells, the pineal gland, the adrenal gland, the pancreas, the liver, the kidney, the heart, cartilages, the connective tissue, the dermis, the epidermis and some fetal cells. Monocilia act as sensory organelles, displaying receptor molecules (e.g. kidney principal cells) and initiate fluid flow (e.g. in the embryonal node).

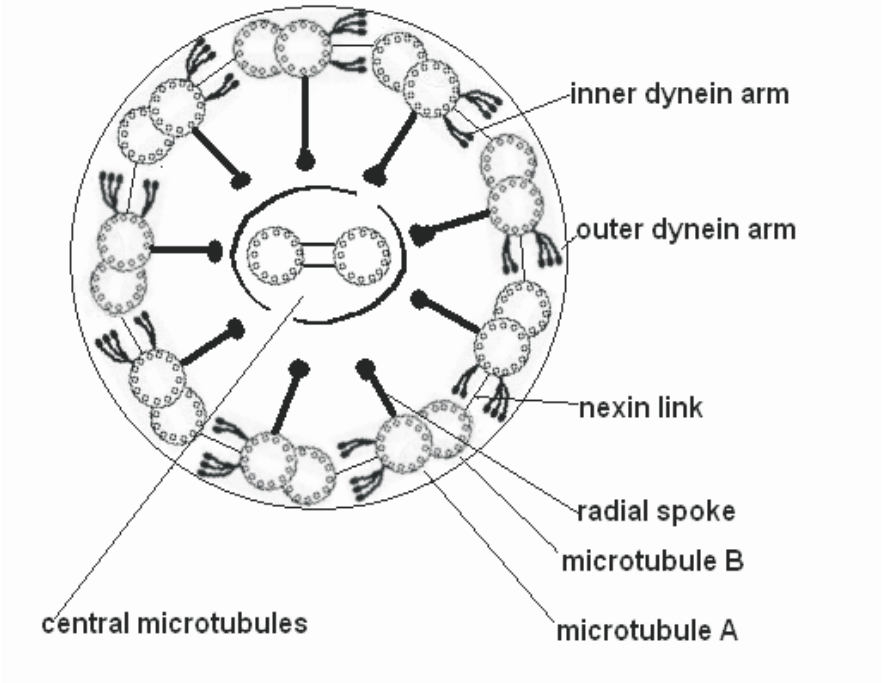

Figure 1. Schematic representation of a cross section of a ciliary axoneme

Cilia and flagella are not only motile but also dynamic in terms of their structure. Within six hours, steady state flagella of *Chlamydomonas reinhardtii* exchange nearly 20% of their proteins. In a process called intraflagellar transport,

axonemal structural proteins are transported from the cytoplasm toward the tip of a flagellelar microtubule, where the assembly process occurs. A number of other proteins, including kinesin II (heterotrimeric kinesin, a member of the KIF3 motor protein family) are responsible for this transfer. The components moved by kinesin require recycling to the flagellar base and this reverse process requires cytoplasmic dynein.[10]

LR determination

Left-right (LR) asymmetry determination is believed to occur in four steps: (1) initial breaking of the bilateral symmetry of the embryo, (2) the establishment of asymmetric gene expression within the embryonic organizer (node), (3) the transfer of the asymmetrical signals from the embryonic organizer to the lateral plate mesoderm, (4) the transfer of this spatial information to organs being positioned.[11]

The node (organizer), a separate structure at the distal tip of the gastrulating embryo, is the mammalian equivalent of the Hensen node in chickens and the Spemann organizer in *Xenopus*. It remains in direct contact with the ectoand endoderm and is a major organizing center that regulates pattern formation in early-stage embryos. The node is essential for both gastrulation and the body axes formation. The "organizing" function is restricted to the endodermal, ventral cells of the node. Each ventral node cell carries a single cilium (monocilium) pointing from the embryo into the extracellular fluid.[12]

Monocilia generate the vortical leftward flow of the extraembryonic fluid in the node region. This flow is supposed to asymmetrically distribute hypothetical morphological inducers, thus providing a mechanism for the initial breaking of symmetry, triggering a different activation of signaling pathways and gene expression on the left and right side of the node.[13, 14, 15] During the following steps, the asymmetrical information, encoded by the differential pattern of gene expression, is spread to the lateral plate.[16, 17, 18]

A complicated and still not entirely understood network of protein interactions is responsible for preventing the contra-lateral spread of the asymmetric signaling cascade in the embryo. Maintaining the embryonic midline barrier provides an additional mechanism stabilizing the asymmetrical gene expression. A differential expression pattern of several factors has been identified in different regions of organ primordia, pointing to the asymmetrical gene expression underlying organ development.[11]

Another, two–cilia model of asymmetry determination has been recently proposed.[19,20] In this model, the node contains two distinct classes of the primary cilia: flow-generating motile cilia and flow-detecting mechanosensory cilia. Based on the similarity with the kidney mechanosensorycilia, where polycystin-1 and polycystin-2[21,22,23] have been shown to form membrane calcium channels, ciliary signaling in the embryonic node could be coupled to intracellular calcium signaling. At the end of this signal transduction pathway, specific factors would be released on the left side of the node, where the flow remains the strongest.

Data supporting the ciliary fluid flow hypothesis come from the studies of murine *inversus viscerum* (*iv*) mutants, harboring mutations in a ciliary component, the left-right dynein (the protein product of *lrd*).[18] The *iv* mice have immotile nodal monocilia, fail to generate the nodal flow and exhibit situs randomization.[15] However, other murine mutants, called the *inversion of embryonic turning* (*inv*), cannot be explained by a "single-cilia" model, since they exhibit nearly 100% incidence of *situs inversus* and the direction of the nodal flow is not altered in those mutants.[15, 24] Inversin (the protein product of *inv*) has been localized in the centrioles, in the primary cilium and in the anaphase-promoting complex, and proposed to link the mechanosensory function of the monocilia with the cell cycle.

In KS families affected siblings with *situs inversus* and *situs solitus* can be observed, indicating that the determination of the LR asymmetry in those patients is random.

Epidemiology and genetics of PCD/KS

Estimates of PCD prevalence range from 1/15,000 to 1/60,000 live births[5]; other estimates place PCD prevalence at the level of 1/12,500.[25, 26] Since about half of the PCD cases display *situs inversus*, the incidence of KS is estimated to range between 1/30,000 and 1/120,000 live births (or 1/25,000). It is certain that the disease frequency is markedly underestimated, especially when PCD patients with the usual visceral arrangement (CDO, ciliary disfunction only) are concerned. The absence of the mirror image arrangement of the organs, renders diagnosis more difficult and CDO patients are often diagnosed later, if at all.

Inheritance in most cases appears to be autosomal recessive,[2, 4, 5] although pedigrees showing autosomal dominant or X-linked modes of inheritance have also been reported.[27, 28, 29]

All the symptoms of PCD are caused by functional defects of the cilia, classified as dysmotility or immotility. Eighteen different ultrastructural defects underlying PCD have been characterized, among them: the most common outer dynein arm defects, inner dynein arm defects, microtubule transpositions, radial spokes defects, and the absence of nexin links. In 10-20% of patients no ultrastructural defects can be found in electron microscopy, although their cilia are immotile.

PCD is the first human disorder known to be caused by a primary dysfunction of the ciliary apparatus. Recently, two other disorders have been pathomechanically linked to the epithelial cilia. Polycystic kidney disease is caused by the impairment of intraflagellar transport in the non–motile primary cilia of kidney epithelial cells.[30] Outer segments of retinal photoreceptors are formed from the primary cilia. In a mature photoreceptor, a short fragment of a cilium called the connecting cilium, remains between the outer and inner segment of cones and rods. In a process, similar to the intraflagellar transport, opsin is moved from the inner segments of the rod and cones toward their outer segments via the connecting

cilium; its defect is believed to be the main cause of RP.[31, 32] The genetic basis of a variety of defects affecting ciliary structure and function in PCD is not clear. Approximately 250 different polypeptides have been identified in the ciliary axoneme of lower organisms. At least the same number of proteins can be expected in the axonemes of humans.[33] It is rather unlikely that mutations of as many as 250 different genes coding for various ciliary proteins cause the same or similar pathologic consequences of the ciliary dysfunction. If that was true, one might expect the incidence of PCD to be much higher than actually reported.[3] It is possible that many ciliary protein gene mutations may be lethal even if heterozygous. On the other hand, mutations in other genes may not affect ciliary function at all, perhaps due to the redundancy of functions of common gene families such as dyneins. In any case, the low incidence of KS and PCD suggests that perhaps mutations in only a small number of these genes underlie the diseases. Two strategies can be applied to map genes causing PCD. One, a candidate gene approach, consists in studying genes selected based on their purported ciliary function or animal/plant models of the disease. The second strategy, linkage studies, consist in looking for genomic markers linked with the disease in PCD families (Table 1).

Table 1. PCD genes, candidate genes and chromosomal regions indicated by different mapping strategies

Chromosomal region	Gene	Disease involvement	Animal model	Linkage	References
1p35.1	*HP28*		*Chlamydomonas reinhardtii p28*		61
2p14-q14.3	*CFC1*	heterotaxy			75
2q33.1	*DNAH7*				60
2q34	*hPF20*		*Chlamydomonas reinhardtii p20*		62
3p21.3	*DNAH1*		murine *Mdhc7* mutants		58
4q*	-			PCD	94
5p15-p14	*DNAH5*	PCD	murine *Mdnah5* mutants	PCD	45
5q31	*KIF3A*		murine *Kif3a* null mutants		14
6p21.3	HLA region *(TUBB, MLN, KNSL2, HSET)*			PCD	67, 70
6q25-q27	*TCTE3*		*Chlamydomonas reinhardtii LC2*		63

7p15	*DNAH11*	PCD	murine *iv/iv* mutants		49, 51
7q33-q34	-		murine *hop* mutants		34
8q*	-			PCD	94
9p13-21	*DNAI1*	PCD	*Chlamydomonas reinhardtii IC78*		41, 42
9q31	*INV*	Nephronophthis	murine *inv* mutants		80, 81
10p*	-			PCD	94
10q24	*DPCD*		murine *Poll* mutants		43
11q*	-			PCD	94
13q*	-			PCD	94
13q12.1	*Tg737*		murine models of PKD with SI		83
14q32	*EMAPL*	Usher syndrome 1a			94, 65
15q13.1-15.1	-			PCD	93
15q24	-			KS	95
16p12.1-12.2	-			PCD	93
16pter*	-			PCD	94
17p12 (17q25)	*DNAH9*		murine *iv/iv* mutants		53
17q*	-			PCD	94
17q22-q25	*HFH-4*		murine *Foxj1-/-* mutants		89
17q25	*DNAI2*		*Chlamydomonas reinhardtii IC69*		44
19q13.3	*RSHL1*				74
	ZIC3			PCD	91, 92, 94
20	*KIF3B*		murine *Kif3b* null mutants		13
Xp11.4	*RPGR*			PCD with RP	72, 73

*results of a genome-wide low density linkage analysis

Animal models of PCD/KS

A number of animal models of PCD have been identified, including examples in the mouse (*iv/iv, inv/inv, hop*) ,[24, 34, 35] rat (WIC-Hyd rat),[36] dog[37, 38] and pig[39]. Both dogs (springer spaniels, English setters, Newfoundlands) and pigs have a relatively high incidence of a respiratory disease with the phenotype almost identical to the primary ciliary dyskinesia in humans, but in both species hydrocephalus is an additional prominent component of the syndrome. *Chlamydomonas reinhardtii*, a

unicellular alga with two flagella containing an axonemal structure similar to that of the human respiratory cilia and sperm tails, became a good model for the studies of mutational events related to ciliary ultrastructural abnormalities.[40] Studies on some of these models helped to focus attention on particular regions of the human genome that contain homologues of the genes involved in animal PCD models, thus creating a list of candidate genes and possible chromosomal locations to be taken into account in genetic mapping of human PCD. So far, eleven genes been screened for mutation (Table 2) and two genes have been identified, in which mutations underlying the ultrastructural phenotype were found in PCD patients (*DNAI*, *DNAH5*); the third gene should be considered as a possible candidate (*DNAH11*).

Table 2. PCD candidate genes screened for mutation.

Gene	Mutations detected	Ciliary localization	L-R asymmetry	Ultrastructural defect	Reference
HP28	ND	Inner dynein arms	NK	NK	61
DNAH5	in 8 PCD\KS families	Outer dynein arms	SS, SI	Outer dynein arms absence	46
DNAI1	in 1 PCD family in 2 PCD\KS families in 3 KS families	Outer dynein arms	SS,SI	Outer dynein arms absence	41, 42, 43
DNAH9	ND	NK	NK	NK	53
DNAH11	in 1 CF/PCD patient	Outer dynein arms	SI	none	49
hPF20	ND	Central complex microtubules	NK	NK	62
TCTE3	ND	Outer dynein arms	NK	NK	63
HFH-4	ND	NK	NK	NK	89
DPCD	ND	NK	NK	NK	43
DNAH7	ND	Inner dynein arms	NK	NK	60
DNAI2	ND	Outer dynein arms	NK	NK	44

SI – *situs inversus*; SS – *situs solitus*; ND – not detected; NK – not known

DNAI1 (chromosome 9p13-21)

DNAI1, a human homologue of the *Chlamydomonas reinhardtii* gene *IC78*, was the very first human gene, in which mutations causing the ultrastructural phenotype were described in PCD patients.[41] *DNAI1*, located at 9p13-21, is highly

expressed in the trachea and testes and is composed of 20 exons. This gene encodes a dynein intermediate chain type 1, found in the outer dynein arms (ODA). Originally two loss-of-function mutations of this gene have been identified in a patient with PCD characterized by immotile cilia lacking ODA: (i) maternal insertion of four nucleotides in exon 5, causing a frameshift; and (ii) paternal T insertion at +3 position of the intron 1 donor site resulting in the absence of splicing. Both mutations generate severely truncated polypeptides lacking 85% and 95% of the DNAI1 protein, respectively. However, linkage between this gene and PCD has been excluded in five other families, demonstrating locus heterogeneity of the disease. Further studies of a number of KS families[42] and PCD/KS families with ODA defects[43] revealed four other mutations of the *DNAI1* gene: (iii) transition G to A (position 1714or5) in exon 16, resulting in a G515S substitution, removing highly conserved glycine, a potential N-myristoylation site; (iv) 12 bp deletion in exon 17, resulting in protein truncation, removing a highly conserved phenylalanine at position 556; (v) G1874C transversion leading to a missense mutation (W568S); (vi) G1875A transition leading to a nonsense mutation (W568X). The last two mutations were present at the conserved tryptophan residue in the WD-repeat region which is predicted to lead to an abnormal folding of the DNAI1. The splice defect caused by the T insertion in intron 1 is the most common mutation present on one allele in all the cases analysed so far. Mutations of the *DNAI1* cause either *situs solitus* (normal) or *situs inversus*.

The other human gene coding for dynein intermediate chain 2 (*DNAI2*), located at 17q25, with a similar pattern of tissue specificity of expression, showed no mutations of the coding sequence in a series of PCD patients investigated so far.[44]

***DNAH5* (chromosome 5p15-5p14)**

A total genome scan, performed in a large consanguineous family of Arab origin with the Kartagener syndrome and the absence of outer dynein arms, pointed to a region on chromosome 5p15-5p14 as being linked to the disease. A murine

chromosomal region on chromosome 15, which is syntenic to human chromosome 5p and 8q, contains a gene for a heavy chain of axonemal dynein. Further experiments have led to the isolation of the human gene *DNAH5*, coding for an axonemal dynein heavy chain 5. Homology search revealed a high sequence conservation between human DNAH5 protein and *Chlamydomonas reinhardtii* axonemal dynein heavy chain gamma.[45] Sequencing of the overlapping cDNA fragments from the human testes and trachea resulted in the characterization of the 13872 nucleotide open reading frame encoding 4624 aminoacids. *DNAH5* is composed of 79 exons and is expressed in the lung, the kidney, and, at a lower level, in the brain, heart and testes.

Sequencing of the entire coding region in several PCD families (selected following the analysis of a haplotype composed of intragenic polymorphic markers) revealed four homozygous mutations and six heterozygous mutations in eight families; in two cases, no mutations were revealed on the second allele.[46] Seven nucleotide changes introduced a premature stop codon, resulting in peptides lacking motor domains and microtubule binding domains; in addition, two missense mutations in the conserved domains and one splice-site mutation were found. From seventeen individuals carrying mutations in *DNAH5*, seven displayed *situs inversus*. The occurrence of individuals with *situs inversus* and *situs solitus* within the same family strengthens the hypothesis of the left-right asymmetry randomization in patients with ciliary defects.

Dnahc5, a murine equivalent of human DNAH5, was shown by *in situ* hybridization to be present in the embryonic ventral node between 7th and 8th day after conception; this confirms its role in the early stages of embryogenesis. DNAH5 is the first axonemal protein confirmed to play an essential role in both embryonic cilia and lung cilia functioning.[46]

Recently, a genotype-phenotype correlation for this gene was reported. Results of the electron microscopy of the respiratory cilia from individuals harboring a premature translation termination mutation in *DNAH5* were compared with

patients with a splice-site mutation in *DNAH5*. In the former case, cilia displayed a complete absence of ODA. Ultrastructural findings in the individuals with splice site mutations showed only a partial loss of ODA.[47]

DNAH11 (chromosome 7p21)

The gene for a heavy dyneine chain of the outer dynein arms maps to the 7p15 region[48], and clones containing sequences homologous to the dynein gene family map to 7q21-q22. The third gene documented in the literature as likely to be causative of PCD, *DNAH11* coding for human dynein heavy chain 11 mapped to chromosome 7p21. Its homologue is mutated in the mouse *iv/iv* model of *situs inversus*.[49] The mutation analysis of all the 40 exons of the human *DNAH11* gene was performed and a homozygous nonsense mutation R2852X was identified in exon 52. The mutation was detected in a patient previously reported as having uniparental disomy of chromosome 7 (UPD7).[50] The patient expressed complete *situs inversus*, motionless respiratory cilia and cystic fibrosis (due to the homozygous F508del mutation of the CFTR gene located on human chromosome 7). His mother was not a carrier of the CF mutation and both chromosomes 7 were inherited from his father (a documented CF carrier), as false maternity was excluded. Electron microscopy analysis in this patient revealed normal cilia and dynein arms, which is compatible with the nonsense mutation in the middle of the motor domain of a heavy dynein chain. The truncated mutant DNAH11 protein could theoretically be incorporated in the dynein arms but is inactive due to the lack of the microtubule-binding domain. A question remains whether this particular patient really suffered from PCD or all the respiratory symptoms caused by CF, with *situs inversus* were only incidentally co-occurring in this case.

Other genes/regions coding for axonemal structure proteins

Linkage data obtained for PCD families recruited in Poland, strongly excluded a putative KS gene (at least for a high proportion of families studied) from all or most of chromosome 7.[51] PCD families in that study were subclassified either as

KS families (if at least one PCD-affected member exhibited *situs inversus*) or as ciliary dysfunction only (CDO) families (if none of the affected members exhibited *situs inversus*). The KS and CDO families were analysed separately because of the possibility that different molecular pathologies could underlie these subtypes of PCD. Such a hypothesis is consistent with the fact that mice carrying the *hop* mutation that maps to a region syntenic to human 7q33-q34 exhibit CDO but not *situs inversus*. The results provided a weak suggestion of linkage in only the CDO subset of the PCD families. These linkage data concords with the existence of the *DNAH11* mutation in the UPD7 patient described above, with PCD and normal *situs*.[49] Human chromosome 7q33-q34 is syntenic to a fragment of mouse chromosome 6 containing the *hop* mutation (previously named *hpy*). In the homozygous form, this mutation causes a dynein-related defect in the cilia and flagella, similar to that seen in PCD and resulting e.g. in male infertility. In contrast to all the other mouse models of PCD, these mutants do not show laterality defects.[34, 52]

In other studies, no mutations were detected in the *DNAH9*[53] and in six other axonemal dynein heavy chain genes.[54] The former was mapped by FISH and radiation hybrid analysis to 17p12[55, 56]; the alternative map position of this locus is 17q25.[57] A mutation search performed in two families with the concordant inheritance of *DNAH9* alleles in the affected individuals has not revealed any pathogenic mutations.[53]

Disruption, by homologous recombination, of the mouse dynein heavy chain 7 gene (*Mdhc7*), a homologue of the human *DNAH1*, results in male infertility due to dramatically reduced sperm motility and in a ciliary beat frequency reduction, but not in defects in the axonemal structure of the tracheal cilia or sperm flagella.[58] *DNAH1* was mapped to 3p21.3, which is syntenic to mouse chromosome 14 region B-C containing the *Mdhc7* gene.[59]

A comparative biochemical analysis of cilia from healthy subjects and a PCD patient with the absence of inner dynein arms has lead to the characterization

of *DNAH7* as a part of IDA. Northern and Western blots indicated that both *DNAH7* mRNA and the protein itself were present in cells from this IDA-deficient patient. However, contrary to the healthy controls, the patient's cells stained with the anti-DNAH7 antibody showed an increase of the specific signal in the cytoplasm and its lack in cilia. The analysis of the *DNAH7* sequence from the PCD patient has not revealed any mutations. It was hypothesized that PCD in this patient might result from a defect in the DNAH7 transport from the cytoplasm to the cilia.[60]

Mutations in the *Chlamydomonas p28* gene that encodes a dynein light chain, lead to the absence of inner dynein arms in the flagellar axonemes of this model organism. The analysis of *HP28,* a human *p28* ortholog located at 1p35.1, has not resulted in the detection of any mutations in PCD patients studied.[61]

Mutants of *Chlamydomonas reinhardtii* called *pf20* have totally paralyzed flagella and display a complete lack of the central complex. One transcript variant from the human homologue, *hPF20* gene, is expressed exclusively in the testes. Genomic localization of *hPF20* was assigned to chromosome 2q34. No mutations have been revealed in the screening performed in four patients with PCD, *situs solitus* and the central complex defect, and in one infertile patient with the central complex defect and the absence of inner dynein arms in the sperm.[62]

Mutants of *Chlamydomonas reinhardtii* called *LC2* lack outer dynein arms and show a slow swimming phenotype. *LC2* codes for a light dynein chain. Its murine homologue, *Tcte3,* was originally identified in the testes as a candidate gene for the involvement in the non-Mendelian transmission of mouse chromosome 17 variant forms, known as t-haplotypes. In this variant, *Tcte3* had an aberrant sequence and sperm motility was reduced. Human *TCTE3,* located on 6q25-q27, is expressed in ciliated tissues. Mutation screening of *TCTE3* in patients with PCD have not revealed any sequence abnormalities.[63]

EMAPL, the echinoderm microtubule-associated-protein-like gene, is located on human chromosome 14q32. Mutations in this gene have been shown to

cause the Usher syndrome type 1A; axonemes of the respiratory cilia in these patients sometimes exhibit ultrastructural defects similar to those observed in PCD.[64, 65]

The *HLA* region of chromosome 6p contains the tubulin gene (*TUBB*)[66] and was suggested to be linked to PCD in one study.[67] However, the motilin gene (*MLN*), which also resides in this region, was not supported as a candidate for the involvement in PCD.[68, 69] Another candidate located in the HLA region is the kinesin-like 2 gene (*KNSL2*).[70] Kinesins are involved in ciliary beating[71] and this gene is located at the centromeric end of the major histocompatibility complex. Four single-base substitutions were detected in the *HSET* gene (kinesin-related protein), mapped also to 6p21.3, in two PCD families studied.[70]

RPGR (retinitis pigmentosa GTPase regulator) protein is a structural component of the connecting cilium in the retina but is also expressed in the testes and lungs. A mutation in the *RPGR* gene located on Xp21.1, responsible for 70% of the X-linked retinitis pigmentosa, has been reported to be associated in one family with retinitis pigmentosa (RP), hearing loss, and bronchosinusitis.[72] Preliminary results might suggest theinvolvement of *RPGR* in the pathogenesis of particular cases of PCD/RP comorbidity.[28, 73]

Genes/regions involved in L-R asymmetry

Candidate genes for PCD include a whole group of genes involved in the determination of the left/right asymmetry of internal organs. Left-right (L-R) axis abnormalities or laterality defects are relatively common in humans (1 in 8,500 live births). It has to be emphasized, however, that the L-R asymmetry defects that are not linked with bronchopulmonary symptoms do not classify as PCD or KS.

Only few genes associated with a small percentage of human situs defects have been identified. These include *ZIC3*,[74] *CFC1* (encoding the CRYPTIC protein) ,[75] *LEFTB* (formerly *LEFTY2*)[76, 77] and *ACVR2B* (encoding activin receptor IIB).[78] Mutations in *ZIC3* (chromosome 19q13.3-qter) and *CFC1* (chromosome

2p14-q14.3) were found in individuals exhibiting heterotaxic phenotypes (randomized organ positioning).[74, 75]

It has been shown that proteins KIF3A and KIF3B of the kinesin superfamily determine the L-R asymmetry by participating in the assembly of the nodal cilia that produce the leftward nodal flow of an extra embryonic fluid in mouse embryos.[13, 14, 71] Targeted mutagenesis of KIF3A and KIF3B kinesins results in the failure to assemble ciliary axonemes and causes a bilateral expression of factors normally expressed only on one side of the node.[79] Murine embryos lacking KIF3A die after ten days. Electron microscopy shows the loss of cilia, normally present on the embryonic ventral node, and a bilateral expression of Pitx2, normally expressed only on the left side of the node. Mutant embryos display morphological and asymmetrical defects. Cardiac looping is normal, reversed or retarded.[31] The *KIF3B* null mutants, totally lacking the KIF3B protein, do not survive beyond midgestation. Early mouse embryos exhibit growth retardation, pericardial sac ballooning and neural tube disorganization. Left-right asymmetry is randomized in the heart loop and the direction of embryonic turning. Lefty-2, normally present on the left side of the node, is expressed bilaterally or absent. The embryonic ventral node lacks cilia, but basal bodies are present.[13] Genes of these proteins are located on human chromosomes 5q31 (*KIF3A*) and 20 (*KIF3B*).

Mouse mutant *inv* (*inversion of embryonic turning*) is the only known murine mutation that causes almost 100% of homozygous embryos to develop *situs inversus*.[24] These embryos do not survive and they have cardiovascular, renal and pancreatic defects; however, their respiratory cilia show no structural or activity abnormalities. The gene causing the *inv* mutation has been cloned; the protein, called inversin, contains ankyrin-like repeats and potential nuclear localization signals.[80] This gene reverses the left-right pattern of expression of the *nodal*, *lefty-1* and *lefty-2* genes, suggesting that inversin acts upstream of these TGF-beta-like genes, which are also involved in the left-right asymmetry.[71] In addition to the aberrant shape of the *inv/inv* node itself, the defective monocilia

rotate more rapidly than they normally do. Inversin has been shown to be strongly expressed in the primary cilia of the renal ducts. The human inversin gene (*INV*) is located on chromosome 9q31. Mutations in the *INV* have been detected in patients suffering from nephronophthisis (NPHP2), an autosomal recessive cystic kidney disease with or without *situs inversus*.[81] In a study of heterotactic patients, a German family of Turkish origin has been described with all, both affected and non-affected, individuals heterozygous for a mutation in the donor splicing site of intron 12 in the *INV* gene, resulting in two different aberrant splicing isoforms. No complementing mutation was found in the affected; their phenotype has been therefore explained either by a randomization of the LR asymmetry, or by the di- or trigenic mode of inheritance.[82]

A transgenic mouse harboring an insertional mutation in the *Tg737* gene displays a phenotype resembling the human recessive autosomal polycystic kidney disease.[30] *Tg737* encodes a tetratricorepeat protein Polaris. Embryos of knockout mice lacking Polaris have a defective LR development, with *situs* randomization and frequently the reverse direction of heart looping.[83] The mutation is lethal: embryos do not survive longer than ten days. The *Chlamydomonas* homologue of Polaris functions in intraflagellar transport (IFT).[84] The human Polaris homologue is encoded by the gene located on chromosome 13q12.1. The analysis of the human *TG737* gene sequence in patients with the autosomal recessive form of the polycystic kidney disease did not reveal any mutations.[85]

The mouse mutants *iv/iv* have laterality defects in 50% of offspring, but ciliary abnormalities have not been described.[86] It has been shown that mutations of the *lrd* gene (encoding left/right dynein) are the cause of the *iv* and *legless* (*lgl*) mouse mutants, in which embryonic ventral node cilia are affected causing a randomization of the L-R
asymmetry.[71, 87] The region, to which the gene is likely to map in humans, is on chromosome 7q. This finding definitely confirms the role of the dynein protein family in both ciliary structure/function and lateralization, shown also for *DNAH5* in humans.[46]

A link between ciliary motility and lateralization has been also provided by mouse mutants defective in genes required for ciliary growth and/or beating. A randomization of the L-R asymmetry due to a complete absence of cilia was observed in mice harborin a mutation in the hepatocyte nuclear factor 3/forkhead homologue-4 *(Hfh-4,* alias *Foxj1*) gene. Homozygous *Foxj1* -/- mutants displayed a complete absence of respiratory cilia and spermatic flagella, and 50% of the animals revealed heterotaxy and *situs inversus totalis*. In addition, the expression of the left-right dynein (*lrd*) was not observed in the embryonic lungs of *Foxj1* mutants, suggesting that *Foxj1* may act by regulating the expression of dyneins.[88] The human homologue, *HFH-4*, also called *FKHL13* and recently renamed *Foxj1*, has been mapped to chromosome 17q22-q25. Members of the forkhead/winged-helix family of transcription factors, to which HFH-4 belongs, have specific tissue expression. HFH-4 has been detected in the epithelial cells of the lung, oviduct, choroid plexus and the ependyma of the brain and seminiferous tubules of the testes. Further studies with anti-HFH-4 antibodies revealed that the protein is localized in the ciliated cells of those tissues, but exclusively in the mobile cilia. During normal ciliogenesis the appearance of cilia follows *HFH-4* expression, both in the mouse and the human. Mutation screening of the two exons and of intron-exon junctions of the *Foxj1* in PCD families have not confirmed the role of HFH-4 in PCD pathogenesis.[89]

Targeted mutagenesis of the murine gene coding for DNA polymerase caused hydrocephalus, *situs inversus*, chronic sinusitis and male infertility.[90] The analysis of the genomic region surrounding the polymerase gene revealed the presence of another gene, *Dpcd,* on the opposite strand. The human homologue of this gene, *DPCD,* is expressed during ciliogenesis. No disease causing mutations were found in this gene in the PCD patients scanned.[43]

Other PCD/KS-linked chromosomal regions and genes

A high-density genomic scan of chromosome 15, using microsatellite markers specific for this chromosomal region, performed on a large number of the

Kartagener syndrome (KS) and ciliary dysfunction only (CDO) families of Polish origin,[51] localized a putative *KS* gene to the chromosomal region of 15q24. In CDO families, linkage for this chromosomal location was negative, demonstrating locus heterogeneity dependent on the presence or absence of *situs inversus* among PCD individuals.

A total genome scan in five PCD families of Arabic origin (a defect of ODA, four with reported consanguinity) defined a critical region of a possible PCD locus location on chromosome 19q13.3.[91] Linkage was not found for this region in two of these families, again confirming locus heterogeneity. Chromosome 19q13.3–qter is a gene-rich region containing at least two candidates for a PCD locus. One is the radial spokehead-like 1 gene (*RSHL1*), residing at the myotonic dystrophy locus. *RSHL1*, expressed in adult testes, codes for a radial spokehead-like protein characterized by high homology to proteins of the sea urchins and *Chlamydomonas*, important in the normal ciliary or flagellar action, including that of sea urchin spermatozoa.[92] The other one, the *ZIC3* locus, codes for the zinc finger transcription factor and is mutated in subjects with the X-linked and sporadic laterality defects.[74]

Another genome-wide linkage scan was performed in PCD families from two different isolated populations: from the Faroe Islands and from the Druze community in northern Israel. Subsequent fine mapping pointed to two PCD-linked loci, located on chromosomes 16p12.1-12.2 and 15q13.1-15.1 in the Faroe Islands and the Israeli Druze population, respectively. The Faroe Island subset consisted of families with the absence of outer dynein arms; only one patient exhibited *situs inversus*. In the Druze families, several patients exhibited *situs inversus* and a partial loss of inner dynein arms.[93]

A genome-wide linkage search performed in 31 PCD families failed to identify the major disease locus, confirming the common conviction on extensive locus heterogeneity.[94] Potential genomic regions harboring PCD loci were localized on chromosomes: 3p, 4q, 5p, 7p, 8q, 10p, 11q, 13q, 15q, 16p, 17q and 19q. Linkage

analyses in PCD families with a documented dynein arm deficiency provided suggestive evidence for a linkage to chromosomal regions 8q and 16pter, while analyses carried out exclusively in the Kartagener syndrome families pointed towards 8q and 19q as possible locations of the *KS* locus.

Conclusions

It is clear that PCD is genetically heterogeneous, with different genes involved in different families, depending, in part, on whether *situs inversus* is present (i.e., KS versus CDO subtypes of PCD). The results obtained so far confirm that PCD is not likely to be caused by mutations of hundreds of different genes, even though cilia themselves have so many structural, assembly and regulatory proteins. Future priorities are to identify the genes that reside in the mapped regions, and to identify new genes by combinations of linkage and linkage disequilibrium gene mapping strategies and by mutation screening of appropriate candidate genes.

On the other hand, PCD is not the only disease where defects of the ciliary structure and function may be involved. PKD and RP may serve as examples. Although clinical features of those disorders are different and concern different organs, in some patients a PCD-related phenotype is also present, as in case of RP with bronchopulmonary symptoms or PKD with *situs inversus*. It is not known whether and how the genes mutated in those diseases affect ciliary function(s); one possibility could be that the tissue specific expression of some genes at different stages of ontogenesis results in tissue specific effects of their mutations, leading to clinical manifestations as different disorders.

Acknowledgements

This article has been written within the project 3PO5E 019 23 of the State Committee for Scientific Research (KBN). M.G. is supported by the International

PhD Program of the University of Utrecht. A critical reading of the manuscript by Prof. Ewa Ziętkiewicz is gratefully acknowledged.

References

1. Kartagener M. Zur Pathogenese der Bronchiektasien bei Situs viscerum inversus. *Beitr Klin Tuberk* 1933; 83: 489 –501.
2. Afzelius BA. A human syndrome caused by immotile cilia. *Science* 1976; 193:317-9.
3. Eliasson R, Mossberg B, Camner P, Afzelius BA. The immotile-cilia syndrome. A congenital ciliary abnormality as an etiologic factor in chronic airway infections and male sterility. *N Engl J Med* 1977; 297: 1–5.
4. McKusick V. Kartagener syndrome (244400) and Immotile Cilia Syndrome (242650). In: Online Mendelian Inheritance of Man, ed. 2002; http://www.ncbi.nlm.nih.gov/omim/.
5. Afzelius BA, Mossberg B. Immotile-cilia syndrome (primary ciliary dyskinesia), including Kartagener syndrome. In: Scriver C, Beaudet AL, Sly W, Valle D, eds. The metabolic and molecular bases of inherited diseases. 7th ed. New York: McGraw-Hill 1995: 3943–3954.
6. Schidlow DV. Primary ciliary dyskinesia (the immotile cilia syndrome). *Ann Allergy* 1994; 73: 457–470.
7. Afzelius BA. Genetics and pulmonary medicine. 6. Immotile cilia syndrome: past, present, and prospects for the future. *Thorax* 1998; 53:894-7.
8. Sleigh MA. Primary ciliary dyskinesia. *Lancet* 1981; 2: 476.
9. Jorissen M, Willems T, van der Schueren B, Verbeken E. Secondary ciliary dyskinesia is absent after ciliogenesis in culture. *Acta Otorhinolaryngol Belg* 2000; 54(3): 333–342.
10. Song L, Dentler WL. Flagellar protein dynamics in Chlamydomonas. *J Biol Chem* 2001; 276: 29754–29763.

11. Capdevila J, Vogan KJ, Tabin CJ, Izpisua Belmonte JC. Mechanisms of left-right determination in vertebrates. *Cell* 2000; 101:9–21.
12. Bellomo D, Lander A, Harragan I, Brown NA. Cell proliferation in mammalian gastrulation: the ventral node and notochord are relatively quiescent. *Dev Dyn* 1996; 205: 471–485.
13. Nonaka S, Tanaka Y, Okada Y *et al*. Randomizaton of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. *Cell* 1998; 95: 829–837.
14. Takeda S, Yonekawa Y, Tanaka Y *et al*. Left-right asymmetry and kinesin superfamily protein KIF3 new insights in determination of laterality and mesoderm induction by kif3A-/mice analysis. *J Cell Biol* 1999; 145: 825–836.
15. Okada Y, Nonaka S, Tanaka Y *et al*. Abnormal nodal flow precedes situs inversus in iv and inv mice. *Mol Cell* 1999; 4: 459–468.
16. Levin M, Johnson RL, Stern CD *et al*. A molecular pathway determining left-right asymmetry in chick embryogenesis. *Cell* 1995; 82: 803–814.
17. Collignon J, Varlet I, Robertson EJ. Relationship between asymmetric nodal expression and the direction of embryonic turning. *Nature* 1996; 381: 155–158.
18. Lowe LA, Supp DM, Sampath K *et al*. Conserved left-right asymmetry of nodal expression and alterations in murine situs inversus. *Nature* 1996; 381: 158–161.
19. McGrath J, Somlo S, Makova S *et al*. Two populations of node monocilia initiate left-right asymmetry in the mouse. *Cell* 2003; 114: 61–73.
20. Tabin CJ, Vogan KJ. A two-cilia model for vertebrate left-right axis specification. *Genes Dev*; 2000; 17: 1–6.
21. Hanaoka K, Qian F, Boletta A *et al*. Co-assembly of polycystin-1 and -2 produced unique cation-permeable currents. *Nature* 2000; 408: 990–994.
22. Koulen P, Cai Y, Geng L *et al*. Polycystin-2 is an intracellular calcium release channel. *Nat Cell Biol* 2002; 4: 191–197.

23. Pazour GJ, San Agustin JT, Follit JA *et al*. Polycystin-2 localizes to kidney cilia and the ciliary level is elevated in orpk mice with polycystic kidney disease. *Curr Biol* 2002; 12: R378–R380.
24. Yokoyama T, Copeland N, Jenkins N *et al*. Reversal of left-right asymmetry: a situs inversus mutation. *Science* 1993; 260: 679–682.
25. Kroon AA, Heij JM, Kuijper WA *et al*. Function and morphology of respiratory cilia in situs inversus. *Clin Otolaryngol* 1991 16: 294–297.
26. Teknos TN, Metson R, Chasse T *et al*. New developments in the diagnosis of Kartagener's syndrome. *Otolaryngol Head Neck Surg* 1997; 116: 68–74.
27. Narayan D, Krishnan SN, Upender M *et al*. Unusual inheritance of primary ciliary dyskinesia (Kartagener syndrome). *J Med Genet* 1994; 31: 493–496.
28. Krawczynski MR, Dmenska H, Witt M.Apparent X-linked primary ciliary dyskinesia associated with retinitis pigmentosa and a hearing loss. *J Appl Genet* 2004; 45:107–110.
29. Krawczynski MR, Witt M. PCD and RP: X-linked inheritance of both disorders? *Pediatr Pulmonol* 2004; 38: 88–89.
30. Moyer JH, Lee-Tischler MJ, Kwon HY *et al*. Candidate gene associated with a mutation causing recessive polycystic kidney disease in mice. *Science* 1994; 264: 1329–1333.
31. Marszalek JR, Liu X, Roberts EA, Chui D *et al*. Genetic evidence for selective transport of opsin and arrestin by kinesin-II in mammalian photoreceptors. *Cell* 2000; 102: 175–187.
32. Wolfrum U, Schmitt A. Rhodopsin transport in the membrane of the connecting cilium of mammalian photoreceptor cells. *Cell Motil Cytoskeleton* 2000; 46: 95–107.
33. Luck DJL, Huang B, Piperno G. Genetic and biochemical analysis of the eukaryotic flagellum. *Soc Exp Biol Symp* 1982; 35: 399–407.
34. Handel MA. Allelism of hop and hpy. *Mouse News Lett* 1985; 72: 124.
35. Brueckner M, D'Eustachio P, Horwich AL. Linkage mapping of a mouse gene, "iv", that controls left-right asymmetry of the heart and viscera. *Proc Natl Acad Sci* USA 1989; 86: 5035–5038.

36. Torikata C, Kijimoto C, Koto M. Ultrastructure of respiratory cilia of WIV-Hyd male rats. An animal model for human immotile cilia syndrome. *Am J Pathol* 1991; 138: 341–347.
37. Edwards DF, Patton CS, Bemis DA. Immotile cilia syndrome in three dogs from a litter. J *Am Vet Med Assoc* 1983; 183: 667–672.
38. Watson PJ, Herrtege ME, Sargan D. Primary ciliary dyskinesia in Newfoundland dogs. *Vet Rec* 1998; 143: 484.
39. Roperto F, Galati O, Rossacco P. Immotile cilia syndrome in pigs. A model for human disease. *Am J Pathol* 1993; 143: 634–647.
40. Blair DF, Dutcher SK. Flagella in prokaryotes and lower eukaryotes. *Curr Opin Genet Dev* 1992; 2: 756–767.
41. Pennarun G, Escudier E, Chapelin C *et al*. Loss-of-function mutations in a human gene related to Chlamydomonas reinhardti dynein IC78 result in primary ciliary dyskinesia. *Am J Hum Genet* 1999; 65: 1508–1519.
42. Guichard C, Harricane M-C, Lafitte J-J *et al*. Axonemal dynein intermediate-chain gene (DNAI1) mutations result in situs inversus and primary ciliary dyskinesia (Kartagener syndrome). *Am J Hum Genet* 2001; 68: 1030–1035.
43. Zariwala M, O'Neal WK, Noone PG *et al*. Investigation of the possible role of a novel gene, DPCD, in primary ciliary dyskinesia. *Am J Respir Cell Mol Biol* 2004; 30: 428–34.
44. Pennarun G, Chapelin C, Escudier E *et al*. The human dynein intermediate chain 2 gene (DNAI2): cloning, mapping, expression pattern, and evaluation as a candidate for primary ciliary dyskinesia. *Hum Genet* 2000; 107: 642–649.
45. Omran H, Haffner K, Volkel A *et al*. Homozygosity mapping of a gene locus for primary ciliary dyskinesia on chromosome 5p and identification of the heavy dynein chain DNAH5 as a candidate gene. *Am J Respir Cell Mol Biol* 2000; 23: 696–702.

46. Olbrich, H., Haffner, K., Kispert *et al*. Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. *Nature Genet* 2002; 30: 143-144.
47. Kispert A, Petry M, Olbrich H *et al*. Genotype-phenotype correlations in PCD patients carrying DNAH5 mutations. *Thorax* 2003; 58: 552–554.
48. Kastury K, Taylor WE, Gutierrez M *et al*. Chromosomal mapping of two members of the human dynein gene family to chromosome regions 7p15 and 11q13 near the deafness loci DFNA 5 and DFNA 11. *Genomics* 1997; 44: 362–364.
49. Bartoloni L, Blouin JL, Pan Y *et al*. Mutations in the DNAH11 (axonemal heavy chain dynein type 11) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia. *Proc Natl Acad Sci* USA 2002; 99: 10282–10286.
50. Pan Y, McCaskill CD, Thompson KH *et al*. Paternal isodisomy of chromosome 7 associated with complete situs inversus and immotile cilia. *Am J Hum Genet* 1998; 62: 1551–1555.
51. Witt M, Wang Y-F, Wang S *et al*. Exclusion of chromosome 7 for Kartagener syndrome but suggestion of linkage in families with other forms of primary ciliary dyskinesia. *Am J Hum Genet* 1999; 64: 313–318.
52. Johnson DR, Hunt DM. Hop-sterile, a mutant gene affecting sperm tail development in the mouse. *J Embryol Exp Morphol* 1971; 25: 223–236.
53. Bartoloni L, Blouin J-L, Maiti AK *et al*. Axonemal beta heavy chain dynein DNAH9: cDNA sequence, genomic structure, and investigation of its role in primary ciliary dyskinesia. *Genomics* 2001; 72: 21–33.
54. Poller WC, Knoll R, Fechner H *et al*. Molecular genetic analysis of the axonemal dynein heavy chain gene family in patients with Kartagener syndrome. *Am Thoracic Soc* 95th Intern Conf 1999.
55. Bartoloni L, Blouin J-L, Sainsbury AJ *et al*. Assignment of the human dynein heavy chain gene DNAH17L to human chromosome 17p12 by in situ hybridization and radiation hybrid mapping. *Cytogenet Cell Genet* 1999; 84: 188–189.

56. Maiti AK, Mattei M-G, Jorissen M *et al*. Identification, tissue specific expression, and chromosomal localization of several human dynein heavy chain genes. *Eur J Hum Genet* 2000; 8: 923–932.
57. Milisav I, Jones MH, Affara NA. Characterization of a novel human dynein-related gene that is specifically expressed in testis. *Mammalian Genome* 1996; 7: 667–672.
58. Neesen J, Kirschner R, Ochs M *et al*. Disruption of an inner arm dynein heavy chain gene results in asthenozoospermia and reduced ciliary beat frequency. *Hum Mol Genet* 2001; 10: 1117–1128.
59. Neesen J, Koehler MR, Kirschner R *et al*. Identification of dynein heavy chain genes expressed in human and mouse testis: chromosomal localization of an axonemal dynein gene. Gene 1997; 200: 193–202.
60. Zhang YJ, O'Neal WK, Randell SH *et al*. Identification of dynein heavy chain 7 as an inner arm component of human cilia that is synthesized but not assembled in a case of primary ciliary dyskinesia. *Biol Chem* 2002; 277: 17906–17915.
61. Pennarun G, Bridoux AM, Escudier E *et al*. The human HP28 and HFH4 genes: evaluation as candidate genes for primary ciliary dyskinesia. *Am Thoracic Soc* 97th Intern Conf 2001.
62. Pennarun G, Bridoux AM, Escudier E *et al*. Isolation and expression of the human hPF20 gene orthologous to Chlamydomonas PF20: evaluation as a candidate for axonemal defects of respiratory cilia and sperm flagella. *Am J Respir Cell Mol Biol* 2002; 26: 362–370.
63. Neesen J, Drenckhahn JD, Tiede S *et al*. Identification of the human ortholog of the t-complex-encoded protein TCTE3 and evaluation as a candidate gene for primary ciliary dyskinesia. *Cytogenet Genome Res* 2002; 98: 38–44.
64. Bonneau D, Raymond F, Kremer C *et al*. Usher syndrome type I associated with bronchiectasis and immotile nasal cilia in two brothers. *J Med Genet* 1993; 30: 253–254.

65. Eudy JD, Edmonds M, Yao SF *et al*. Isolation of a novel human homologue of the gene coding for echinoderm microtubule-associate protein (EMAP) from the Usher syndrome type 1a locus at 14q32. *Genomics* 1997; 42: 104–106.
66. Volz A, Weiss E, Trowsdale J, Ziegler A. Presence of an expressed -tubulin gene (TUBB) in the HLA class I region may provide the genetic basis for HLA-linked microtubule dysfunction. *Hum Genet* 1994; 93: 42–46.
67. Bianchi E, Savasta S, Calligaro A *et al*. HPA haplotype segregation and ultrastructural study in familial immotile-cilia syndrome. *Hum Genet* 1992; 89: 270–274.
68. Gasparini P, Griffa A, Oggiano N *et al*. Immotile Cilia Syndrome: A recombinant Family at HLA-Linked Gene Locus. *Am J Med Genet* 1994; 49: 450–451.
69. Gasparini P, Grifa A, Savasta S *et al*. The motilin gene: subregional localisation, tissue expression, DNA polymorphisms and exclusion as a candidate gene for the HLA-associated immotile cilia syndrome. *Hum Genet* 1994; 94: 671–674.
70. Janitz K, Wild A, Beck S *et al*. Genomic organization of the HSET locus and the possible association of HLA-linked genes with immotile cilia syndrome (ICS). *Immunogenetics* 1999; 49: 644–652.
71. Supp DM, Potter S, Brueckner M. Molecular motors: the driving force behind mammalian left-right development. *Trends in Cell Biology* 2000; 10: 41–45.
72. Zito I, Downes SM, Patel RJ *et al*. RPGR mutation associated with retinitis pigmentosa, impaired hearing, and sinorespiratory infections. *J Med Genet* 2003; 40: 609–615.
73. Moore A, Escudier E, Roger G *et al*. RPGR, a gene mutated in patients with a complex phenotype associating X-linked primary ciliary dyskinesia and retinitis pigmentosa. *J Med Genet* 2006; 43:326-33.
74. Gebbia M, Ferrero GB, PiliaG *et al*. X-linked situs abormalities result from mutations in ZIC3. *Nature Genet* 1997; 17: 305–308.

75. Bamford RN, Roessler E, Burdine RD *et al*. Loss-of-function mutations in the EGF-CFC gene CFC1 are associated with human left-right laterality defects. *Nature Genet* 2000; 26: 365–369.
76. Meno C, Shimono A, Saijoh Y, Yashiro K, Mochida K, Ohishi S et al. Lefty-1 is required forleft-right determination as a regulator of lefty-2 and nodal. *Cell* 1998; 94: 287–297.
77. Meno C, Gritsman K, Ohishi S *et al*. Mouse Lefty-2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. *Mol Cell* 1999; 4: 287–298.
78. Kosaki R, Gebbia M, Kosaki K *et al*. Left-right axis malformations associated with mutations in ACVR2B, the gene for human activin receptor type IIB. *Am J Med Genet* 1999; 82: 70–76.
79. Morris RL, Scholey JM. Heterotrimeric kinesin-II is required for the assembly of motile 9+2 ciliary axonemes in sea-urchin embryos. *J Cell Biol* 1997; 138: 1009–1022.
80. Morgan D, Turnpenny L, Goodship J *et al*. Inversin, a novel gene in the vertebrate left-right axis pathway, is partially deleted in the inv mouse. *Nature Genet* 1998; 20: 149–156.
81. Otto EA, Schermer B, Obara T *et al*. Mutations in INVS encoding inversin cause nephronophthisis type 2, linking renal cystic disease to the function of primary cilia and left-right axis determination. *Nat Genet* 2003; 34: 413–420.
82. Schon P, Tsuchiya K, Lenoir D *et al*. Identification, genomic organization, chromosomal mapping and mutation analysis of the human INV gene, the ortholog of a murine gene implicated in left-right axis development and biliary atresia. *Hum Genet* 2002; 110: 157–65.
83. Murcia NS, Richards WG, Yoder BK *et al*. The Oak Ridge Polycystic Kidney Disease (orpk) disease gene is required for left-right axis determination. *Development* 2000; 127: 2347–2355.

84. Pazour GJDB, Vucica Y, Seeley ES *et al*. Chlamydomonas IFT88 and its mouse homologue, polycystic kidney disease gene tg 737, are required for assembly of cilia and flagella. *J Cell Biol* 2000; 151: 709–718.
85. Onuchic LF, Schrick JJ, Ma J *et al*. Sequence analysis of the human hTg737 gene and its polymorphic sites in patients with autosomal recessive polycystic kidney disease. *Mamm Genome* 1995; 6: 805–808.
86. Brueckner M, D'Eustachio P, Horwich AL. Linkage mapping of a mouse gene, "iv", that controls left-right asymmetry of the heart and viscera. *Proc Natl Acad Sci* USA 1989; 86: 5035–5038.
87. Supp DM, Witte DP, Potter SS, Brueckner M. Mutation of an axonemal dynein affects left-right asymmetry in inversus viscerum mice. *Nature* 1997; 389: 963–966.
88. Chen J, Knowles HJ, Hebert JL, Hackett BP. Mutation of the mouse hepatocyte nuclear factor/forkhead homologue 4 gene results in an absence of cilia and random left-right asymmetry. *J Clin Invest* 1998; 102: 1077–1082.
89. Maiti AK, Bartoloni L, Mitchison HM *et al*. No deleterious mutations in the FOXJ1 (alias HFH-4) gene in patients with primary ciliary dyskinesia (PCD). Cytogenet Cell Genet 2000; 90: 119–122.
90. Kobayashi Y, Watanabe M, Okada Y *et al*. Hydrocephalus, situs inversus, chronic sinusitis, and male infertility in DNA polymerase lambda-deficient mice: possible implication for the pathogenesis of immotile cilia syndrome. *Mol Cell Biol* 2002; 22: 2769–76
91. Meeks M, Walne A, Spiden S *et al*. A locus for primary ciliary dyskinesia maps to chromosome 19q. *J Med Genet* 2000; 37: 241–244.
92. Eriksson M, Ansved T, Anvret M, Carey N. A mammalian radial spokehead-like gene, RSHL1, at the myotonic dystrophy-1 locus. *Biochem Biophys Res Commun* 2001; 281: 835–841.
93. Jeganathan D, Chodhari R, Meeks M *et al*. Loci for primary ciliary dyskinesia map to chromosome 16p12.1-12.2 and 15q13.1-15.1 in Faroe Islands and Israeli Druze genetic isolates. *J Med Genet* 2004; 41: 233–40.

94. Blouin JL, Meeks M, Radhakrishna U *et al*. Primary ciliary dyskinesia: a genome-wide linkage analysis reveals extensive locus heterogeneity. *Eur J Hum Genet* 2000; 8: 109–118.

III. Update of the article "Primary ciliary dyskinesia: genes, candidate genes and chromosomal regions"

Since the introductory article has been written, several developments in PCD genetics and biology of cilia occurred.

Several approaches have been taken to characterize the ciliary genome and proteome. Liquid chromatography-mass spectrometry of human epithelial cilia identified 164 axonemal proteins.[1] A mass spectrometry study of flagella of the unicellular alga *Chlamydomonas reinhardtii* identified 360 proteins very likely to be involved in cilia formation, and 292 that are probably involved.[2] In addition, while comparing the genomes of non-ciliated organisms to ciliated ones, 688 and 183 cilia-related genes in *Chlamydomonas reinhardtii*[3] and in *Drosophila melanogaster*,[4] respectively, has been predicted. The presence of X-box promoter elements, which are targets of cilia-related transcription factors, has led to the identification of additional ciliary genes in *Caenorhabditis elegans*.[5, 6] The results of these different studies have been assembled in online databases: the Ciliome database[7] and the Cilia Proteome database.[8] The cilium seems to be a far more complicated organelle than it was previously assumed. Thousands of proteins are involved in its functioning. Moreover, the classical division into motile cilia and the sensory primary cilia is blurred since motile cilia perform a variety of mechanosensory or chemosensory functions and primary cilia located on the embryonic node are motile.[9, 10]

In the recent years several new disease causing genes have been described. *TXNDC3* encodes a thioredoxin–nucleoside diphosphate kinase, a member of a large family of enzymatic proteins that function as general protein disulfide reductases. TXNDC3 protein is the human ortholog of the sea urchin IC1 gene encoding a component of sperm outer dynein arms. Duriez and colleagues screened for mutations in *TXNDC3* in forty seven PCD patients and identified a heterozygous T>A transversion in exon 15, causing a nonsense mutation at codon 426 (c.1277T>A, p.Leu426X) in one individual.[11] A heterozygous C>T transition in

intron 6 (c.271–27C>T) turned out to be the second disease causing allele. The transition is present in 1% of control population. This SNP modifies the ratio of two physiological isoforms generated by alternative splicing of *TXNDC3*. So far this patient represents the only case of PCD associated with mutations in *TXNDC3*.

A whole-genome linkage scan in a consanguineous Iranian Jewish kindred by use of single nucleotide polymorphisms arrays detected linkage to chromosome 17q25.1. A region of homozygosity shared between the four affected offspring was identified across the *DNAI2* gene. Sequencing of *DNAI2* exons revealed a mutation affecting the donor splice site of exon 11. A subsequent mutation screen in 105 unrelated PCD families revealed two families with disease causing mutations in *DNAI2*. Some of these patients displayed *situs inversus* and the ultrastructural imaging of cilia showed outer dynein arms defects.[12]

Castleman and colleagues performed a linkage study in seven consanguineous families with PCD and central-microtubular-pair abnormalities. One locus on chromosome 6p21.1 was identified in two families with intermittent absence of the central-pair structure in electron microscopy of cilia. Candidate genes sequencing revealed mutations in *RSPH9*. Another locus was mapped on chromosome 6q22.1 in five families with complete absence of the central pair in electron microscopy of cilia. Mutations were subsequently identified in *RSPH4A*. Both *RSPH9* and *RSPH4A* encode proteins, which on the basis of homology with proteins of the biflagellate alga *Chlamydomonas reinhardtii,* are predicted to be components of the axonemal radial spokes. All the patients had a normal visceral laterality.[13]

A Japanese rice fish called Medaka *(Oryzias latipes),* when having a mutation in the *Ktu* gene develop laterality defects, polycystic kidney disease, and have impaired sperm motility. Cilia and flagella from the mutant fish show defects or loss of outer and inner dynein arms. Ktu/PF13 protein is involved in pre-assembly of dynein arm complexes in the cytoplasm. A mutation screen of the *Ktu* gene in 112

PCD pedigrees detected two homozygous mutations in two consanguineous families. In both families *situs inversus* has been observed.[14]

Two groups independently reported mutations in *LRCC50*, coding for another protein involved in pre-assembly of dynein arms in the cytoplasm. Duquesnoy and colleagues followed a candidate-gene approach based on data from a mutant strain of *Chlamydomonas reinhardtii* carrying an ODA7 defect. Mutations in human orthologoue of ODA7, the *LRCC50* gene, were found in four families with a PCD phenotype, characterized in a ciliary ultrastructure by the absence of both dynein arms. *Situs inversus* was present in three patients.[15] Loges and colleagues performed total-genome scans by using single-nucleotide polymorphism (SNP) arrays in seven consanguineous PCD families with combined ODA and IDA defects. In one of these families they found four regions with homozygosity by descent and the highest LOD score. By analysis of candidate genes mutations in *LRCC50* were detected including one large deletion.[16]

The highly inbred structure of the Old English Sheepdog population allowed identification of mutations in *ccdc39* gene in dogs suffering from chronic airway inflammation. Sequencing of the human ortholog revealed loss-of-function mutations in 19 out of 50 selected families with absence of inner dynein arms coexisting with various ultrastructural defects. *Situs inversus* was present in 45% of cases.[17]

Mouse and zebrafish *ccdc40* mutants display laterality and cilia defects. Sequencing analysis of *CCDC40* in a cohort of 26 PCD families revealed mutations in 17 patients from 15 families. The electron microscopy of cilia showed misplacement of central pair microtubules and defects in several axonemal structures excluding outer dynein arms.[18]

Orofaciodigital syndrome type I (OFD1) is an X-linked dominant disease characterized by malformations of the face, oral cavity, and digits. Mutations in *OFD1*, encoding a protein localized at the centrosome and basal body of primary

cilia, cause orofaciodigital syndrome. A frameshift mutation in *OFD1* was found in an extended family in which X-linked intellectual disability cosegregated with PCD. High-speed video microscopy examination of nasal epithelium confirmed impaired ciliary motility.[19]

Homozygosity mapping and subsequent sequencing analysis performed in two consanguineous Bedouin families identified mutation in *DNAL1* encoding the ortholog of the Chlamydomonas axonemal dynein light chain 1. All the three patients had *situs inversus.* Electron microscopy of cilia showed absence of outer dynein arms.[20]

Ciliopathies are an emerging group of pleiotropic, clinically overlapping human disorders caused by mutations in ciliary genes.[21] Primary ciliary dyskinesia involves the dysfunction of motile cilia. The remaining ciliopathies are caused by mutations in genes important for functioning of primary cilia, however in some of these diseases motile cilia can also be affected. The spectrum of clinical symptoms of ciliopathies is broad and the genetic heterogeneity of these diseases is high.

Autosomal dominant polycystic kidney disease (ADPKD) is caused by mutations in *PKD1* and *PKD2* genes, which encode polycystin 1 and polycystin 2 respectively.[22] ADPKD causes chronic renal failure in adults. It was the first disease linked to sensory cilia dysfunction. Primary cilia in renal tubules detect flow by bending and transmit a calcium-mediated signal to epithelial cells.

Mutations in more than a dozen genes can cause nephronophthisis, an autosomal recessive kidney disease leading to renal failure.[23] Some patients present with extrarenal symptoms: retinal degeneration (Senior-Loken syndrome), mental retardation, cerebellar ataxia, bone anomalies or liver involvement.

Bardet-Biedl syndrome (BBS) is a rare, genetically heterogeneous, autosomal recessive ciliopathy with a varied clinical presentation, including retinal degeneration, polycystic kidneys, truncal obesity, polydactyly, hypogonadism,

intellectual disability, diabetes mellitus, and congenital heart disease. There are at least 12 BBS genes encoding proteins of the basal bodies and the axonemes.

Meckel-Gruber syndrome (MKS) is a fatal ciliopathy, characterized by cystic renal disease, polydactyly, hepatic developmental defects, pulmonary hypoplasia, posterior encephalocele, and other central nervous system abnormalities.[24] Mutations in six genes have been found in MKS families.

Joubert syndrome is characterized by distinctive cerebellar and brain stem malformation called the molar tooth sign, hypotonia, and developmental delays. This phenotype can be accompanied by retinal degeneration, renal disease, occipital encephalocele, hepatic fibrosis, and polydactyly.[25] To date, mutations in 14 genes have been identified in individuals with a Joubert syndrome and related disorders.

Orofaciodigital syndrome type 1 is an X-linked dominant disorder caused by mutations in the *OFD1* gene encoding a centrosomal protein.[26] The disease is lethal in males. Females with *OFD1* mutations present with oral, dental, facial, and digital anomalies. Affected individuals can also have renal cysts and CNS abnormalities.

There is a significant genetic and clinical overlap between ciliopathies. For example, Meckel-Gruber syndrome and nephronophthisis can be caused by mutations in *NPHP1*. The *CEP290* gene encodes a centrosomal protein involved in ciliary assembly and ciliary trafficking. Mutations in *CEP290* have been found in Bardet-Biedl syndrome, Joubert syndrome, Leber congenital amaurosis, Meckel-Gruber syndrome, and Senior-Loken syndrome.[27] *Situs inversus* can be present in patients with PCD, nephronophthisis, Bardet-Biedl syndrome and Joubert syndrome. Biliary atresia is a neonatal disorder characterized by fibroinflammatory obliteration of the biliary tract. Approximately 20% of the patients have left-right laterality defects and abnormalities of cholangiocyte cilia.[28] The PCD phenotype can be associated with biliary atresia.[29] Retinal degeneration is a common feature of several ciliopathies related to primary cilia. Mutations in the retinitis pigmentosa GTPase regulator gene (*RPGR*) have been found in patients presenting with retinal

degeneration and the primary ciliary dyskinesia phenotype.[30] A syndrome characterized by respiratory features consistent with PCD, intellectual disabilities, and macrocephaly can be caused by mutation in *OFD1*, providing another example of an overlap between motile and primary cilia related ciliopathies.[31]

References

1. Ostrowski LE, Blackburn K, Radde KM *et al*. A proteomic analysis of human cilia: identification of novel components. *Mol Cell Proteomics* 2002; 1:451-65.
2. Pazour GJ, Agrin N, Leszyk J, Witman GB. Proteomic analysis of a eukaryotic cilium. *J Cell Biol* 2005; 170:103-13.
3. Li JB, Gerdes JM, Haycraft CJ *et al*. Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* 2004; 117:541-52.
4. Avidor-Reiss T, Maer AM, Koundakjian E *et al*. Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell* 2004; 117:527-39.
5. Blacque OE, Perens EA, Boroevich KA *et al*. Functional genomics of the cilium, a sensory organelle. *Curr Biol* 2005; 15:935-41.
6. Efimenko E, Bubb K, Mak HY *et al*. Analysis of xbx genes in *C. elegans*. *Development* 2005; 132:1923-34.
7. Inglis PN, Boroevich KA, Leroux MR. Piecing together a ciliome. *Trends Genet* 2006; 22:491-500.
8. Gherman A, Davis EE, Katsanis N. The ciliary proteome database: an integrated community resource for the genetic and functional dissection of cilia. *Nat Genet* 2006; 38:961-2.
9. Bloodgood RA. Sensory reception is an attribute of both primary cilia and motile cilia. *J Cell Sci* 2010; 123:505-9.

10. Badano JL, Mitsuma N, Beales PL, Katsanis N. The ciliopathies: an emerging class of human genetic disorders. *Annu Rev Genomics Hum Genet* 2006; 7:125-48.
11. Duriez B, Duquesnoy P, Escudier E *et al*. A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia. *Proc Nat Acad Sci* 2007; 104:3336-3341.
12. Loges NT, Olbrich H, Fenske L *et al*. DNAI2 mutations cause primary ciliary dyskinesia with defects in the outer dynein arm. *Am J Hum Genet* 2008; 83:547-558.
13. Castleman VH, Romio L, Chodhari R *et al*. Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with central-microtubular-pair abnormalities. *Am J Hum Genet* 2009; 84:197-209.
14. Omran H, Kobayashi D, Olbrich H *et al*. Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. *Nature* 2008; 456:611-616.
15. Duquesnoy P, Escudier E, Vincensini L *et al*. Loss-of-function mutations in the human ortholog of Chlamydomonas reinhardtii ODA7 disrupt dynein arm assembly and cause primary ciliary dyskinesia. *Am J Hum Genet* 2009; 85:890-896.
16. Loges NT, Olbrich H, Becker-Heck A *et al*.. Deletions and point mutations of LRRC50 cause primary ciliary dyskinesia due to dynein arm defects. *Am J Hum Genet* 2009; 85:883-889.
17. Merveille AC, Davis EE, Becker-Heck A *et al*. CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. *Nat Genet* 2011; 43:72-8.
18. Becker-Heck A, Zohn IE, Okabe N *et al*. The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. *Nat Genet* 2011; 43:79-84.
19. Budny B, Chen W, Omran H et al. A novel X-linked recessive mental retardation syndrome comprising macrocephaly and ciliary dysfunction is allelic to oral-facial-digital type I syndrome. Hum Genet. 2006; 120:171-8.

20. Mazor, M., Alkrinawi, S., Chalifa-Caspi et al. Primary ciliary dyskinesia caused by homozygous mutation in DNAL1, encoding dynein light chain 1. Am. J. Hum. Genet. 2011; 88:599-607, 2011.
21. Ferkol TW, Leigh MW. Ciliopathies: the central role of cilia in a spectrum of pediatric disorders. *J Pediatr* 2012; 160:366-71.
22. Hildebrandt F, Otto E. Cilia and centrosomes: a unifying pathogenic concept for cystic kidney disease? *Nat Rev Genet* 2005; 6:928-40.
23. Hildebrandt F, Attanasio M, Otto E. Nephronophthisis: disease mechanisms of a ciliopathy. *J Am Soc Nephrol* 2009; 20:23-35.
24. Dawe HR, Smith UM, Cullinane AR *et al*. The Meckel-Gruber Syndrome proteins MKS1 and meckelin interact and are required for primary cilium formation. Hum Mol Genet 2007; 16:173-86.
25. Doherty D. Joubert syndrome: insights into brain development, cilium biology, and complex disease. *Semin Pediatr Neurol* 2009; 16:143-54.
26. Singla V, Romaguera-Ros M, Garcia-Verdugo JM, Reiter JF. Ofd1, a human disease gene, regulates the length and distal structure of centrioles. *Dev Cell* 2010; 18:410-24.
27. Coppieters F, Lefever S, Leroy BP, De Baere E. CEP290, a gene with many faces: mutation overview and presentation of CEP290base. *Hum Mutat* 2010; 31:1097-108.
28. Chu AS, Russo PA, Wells RG. Cholangiocyte cilia are abnormal in syndromic and non-syndromic biliary atresia. *Mod Pathol* 2012 Feb 3.
29. Gershoni-Baruch R, Gottfried E, Pery M *et al*. Immotile cilia syndrome including polysplenia, situs inversus, and extrahepatic biliary atresia. *Am J Med Genet* 1989; 33:390-3.
30. Moore A, Escudier E, Roger G *et al*. RPGR is mutated in patients with a complex X linked phenotype combining primary ciliary dyskinesia and retinitis pigmentosa. *J Med Genet* 2006; 43:326-33.

31. Budny B, Chen W, Omran H *et al*. A novel X-linked recessive mental retardation syndrome comprising macrocephaly and ciliary dysfunction is allelic to oral-facial-digital type I syndrome. *Hum Genet* 2006; 120:171-8.

Part I

Classical genetic approach

Chapter 2

Linkage analysis localizes a Kartagener syndrome gene to a 3.5 cM region on chromosome 15q24-25

Maciej Geremek, Ewa Ziętkiewicz, Scott R. Diehl, Behrooz Z. Alizadeh, Cisca Wijmenga, Michał Witt

J Med Genet 2006;43:e1

Abstract

Background: Primary ciliary dyskinesia (PCD) is a genetic disorder caused by ciliary immotility/dysmotility due to ultrastructural defects of the cilia. Kartagener syndrome (KS), a subtype of PCD, is characterised by situs inversus accompanying the typical PCD symptoms of bronchiectasis and chronic sinusitis. In most cases, PCD is transmitted as an autosomal recessive trait, but its genetic basis is unclear due to extensive genetic heterogeneity.

Methods: In a genome-wide search for PCD loci performed in 52 KS families and in 18 PCD families with no situs inversus present (CDO, ciliary dysfunction-only), the maximal pairwise LOD score of 3.36 with D15S205 in the KS families indicated linkage of a KS locus to the long arm of chromosome 15. In the follow-up study, 65 additional microsatellite markers encompassing D15S205 were analysed.

Results: A maximal pairwise LOD score of 4.34 was observed with D15S154, further supporting linkage of the KS, but not the CDO, families to 15q24–25. Analysis of heterogeneity and haplotypes suggested linkage to this region in 60% of KS families.

Conclusions: Reinforced by the results of multipoint linkage, our analyses indicate that a major KS locus is localised within a 3.5 cM region on 15q, between D15S973 and D15S1037.

Introduction

Primary ciliary dyskinesia (PCD), formerly known as immotile cilia syndrome (MIM 244400; 242650), is a systemic disease caused by inherited dysfunction of the ciliary apparatus.[1,2] Analogous to mitochondrial, lysosomal, and peroxisomal diseases, PCD therefore belongs to a group of disorders involving cellular organelles.[3] The clinical consequences of PCD cover a wide spectrum and mainly affect the lower and upper airways and the male reproductive system. In general,

prognosis remains good, but morbidity can be considerable if the disease is not correctly managed.[4] All forms of PCD are characterised by dysmotility or immotility of cilia in airway epithelial cells, spermatozoa, and other ciliated cells of the body.[3, 5, 6]

Estimates of PCD incidence range from 1/16 000 to 1/60 000 live births.[5] Half of all PCD cases are classified as Kartagener syndrome (KS, 1/32 000 and 1/120 000 live births), where bronchiectasis and chronic sinusitis are also accompanied by situs inversus (SI), a reversal of the usual left-right asymmetry of the abdominal and thoracic organ locations.

All forms of PCD are characterised by dysmotility or immotility of cilia in airway epithelial cells, spermatozoa, and other ciliated cells of the body.[3, 5, 6] The cases of ciliary dysmotility or immotility reported to date are associated with ultrastructural defects of the cilia, such as a total or partial absence of dynein arms (70–80% of all electron microscopy detectable defects), defects of radial spokes or nexin links, and general axonemal disorganisation with microtubular transposition.[5] Discoordination of ciliary function can also be caused by random orientation of the cilia.[7] Inheritance of PCD in most families is autosomal recessive, although pedigrees showing autosomal dominant or X-linked modes of inheritance have also been reported.[8–10] Although within families PCD appears to be transmitted via a single locus, a broad spectrum of ciliary ultrastructural defects clearly suggests that PCD is genetically heterogeneous.

The molecular basis of PCD is far from being clear. Up to 250 different polypeptides have been identified within the ciliary axoneme of the model unicellular alga *Chlamydomonas reinhardtii*; at least the same number of proteins can be expected in the axonemes of humans.[11] It is rather unlikely that mutations within as many as 250 different genes coding for various ciliary proteins cause the same or similar pathologic consequences of ciliary dysfunction. If this were true, one might expect the incidence of PCD to be much higher than it actually is.[3] It is possible that mutations in some ciliary protein gene(s) may be lethal even if

heterozygous, while mutations in other genes may not affect ciliary function at all. Hence, one would expect involvement of a limited number of genes in the pathogenesis of this disorder. This potential reduction of locus heterogeneity provides cautious optimism for the success of gene mapping efforts.

So far, only two known human genes, *DNAI1* and *DNAH5*, have a documented causative effect in PCD. *DNAI1*, located in 9p13–21, codes for the intermediate chain of the axonemal dynein and is an ortholog of the IC78 dynein gene in *Chlamydomonas*.[12–14] *DNAH5*, located in 5p15–p14, codes for the heavy chain of the axonemal dynein and is related to the *Chlamydomonas* gene coding for the dynein γ-heavy chain.[15] Mutations in both these genes cause an outer dynein arms defect phenotype. The relatively broad spectrum of low frequency mutations in both these genes enhances the genetic heterogeneity of PCD. Together, mutations in *DNAI1* and *DNAH5* have been shown to account for ~24% of PCD.[16]

Obviously, mutations in genes other than *DNAI1* and *DNAH5* must be responsible for most PCD.

In the past, a number of genome-wide scans in PCD families have been performed, resulting in several chromosomal regions being considered as candidates for PCD gene(s) locations. A genome-wide, low density linkage search performed in 31 PCD families revealed potential PCD loci on 3p, 4q, 5p, 7p, 8q, 10p, 11q, 13q, 15q, 16p, 17q, and 19q, but failed to identify a major disease locus, confirming extensive locus heterogeneity.[17] A genome scan in five PCD families of Arabic origin (four with reported consanguinity) indicated a possible PCD locus on 19q13.3.[18] Another genome-wide scan and subsequent fine mapping in Faroe Island and Israeli Druze families pointed to two PCD linked loci, on 16p12.1–12.2 and 15q13.1–15.1, respectively.[19]

In our previous study on Polish PCD families, linkage analysis performed separately in families with KS and families with CDO (ciliary dysfunction-only, that is PCD not accompanied by situs inversus), indicated that these two subtypes of the

disorder are caused by genes located in different genomic regions.[20] We excluded linkage of a KS locus to chromosome 7,[20] contrary to the suggestion of other authors.[21] At the same time, however, our results provided suggestive evidence of linkage to 7p15 in the non-KS subset of Polish PCD families. Subsequently, *DNAH11*, the gene for the axonemal dynein heavy chain, has been mapped to the 7p15 region,[22] but its involvement in the pathogenesis of PCD remains unclear.

Here we report the chromosomal localisation of a putative KS locus based on linkage analysis in Polish PCD families. The results of both pairwise and multipoint analyses suggest that ~60% of the KS families are linked to a 3.5 cM region on 15q24–25. In contrast, the CDO families are excluded from linkage to chromosome 15, thus demonstrating locus heterogeneity for PCD in general.

Methods

We recruited a total of 70 PCD families, 66 from Poland and four from Slovakia. The families studied in this report did not come from a genetic isolate. Each family had at least one member diagnosed as having PCD with no other major anomalies or dysmorphologies present. In 60% of the families, the clinical diagnosis was confirmed by transmission electron microscopy (EM) analysis of bronchial cilia ultrastructure; of 46 PCD patients analysed by EM, 32 displayed outer and inner dynein arms defects, 11 outer dynein arms defects, and three inner dynein arms defects.

Families were classified as KS if at least one PCD affected member exhibited situs inversus (52 families), or as CDO if none of the affected members exhibited situs inversus (18 families). The mode of inheritance in all families studied was consistent with autosomal recessive transmission (that is, both sexes affected at comparable frequencies, no affected parents or distant relatives). In the KS families, 60 affected subjects (including eight independent pairs of affected siblings and one affected trio) and 139 unaffected family members were genotyped. One KS family with a consanguinity loop, initially included in the study group, was excluded from

further calculations after haplotype analysis. In the CDO families, 21 affected subjects (including three independent pairs of affected siblings) and 39 unaffected members were analysed.

A whole genome scan with 462 microsatellite markers (11–32 per chromosome) was performed. Markers were distributed with a mean intermarker distance of ~8 cM. There were nine intervals between adjacent markers greater than 20 cM; the three largest intermarker intervals were 35, 26, and 26 cM. In the extended analysis of chromosome 15q, 62 marker loci encompassing D15S205 were added, with a mean heterozygosity of 0.7 and intermarker distances of 0–2 cM. Markers were assigned genetic map positions using the Decode map.[23] For markers positioned at exactly the same location or not included in the Decode genetic map,[23] the latest Ensembl assembly (v27.35a.1) was used to determine marker order and to estimate genetic map distances for multipoint linkage analyses.

All PCR amplified microsatellite markers were genotyped using fluorescence based, semi-automated DNA sizing technology,[24] based on Applied Biosystems 373 Automated DNA Sequencers and GENESCAN and GENOTYPER software (Applied Biosystems, Perkin Elmer, Foster City, CA).

Pairwise LOD score analyses were performed using FASTLINK.[25] Multipoint LOD scores allowing for locus heterogeneity were calculated using HOMOG on data generated by LINKMAP.[26] Multipoint LOD scores for markers on chromosome 15 were calculated using GENEHUNTER.[27] We assumed a recessive mode of inheritance with full penetrance, 0.0001 phenocopies, and a PCD disease allele frequency of 0.005.[20] These assumptions were designed to yield a population prevalence for the disease consistent with that reported for PCD. A model with no phenocopies was used to determine heterogeneity in the analysed KS families. Haplotypes were generated using GENEHUNTER software and verified manually.

Results

To assess the statistical power of our collection of PCD families, we calculated the maximum pairwise LOD score for a theoretical marker locus with nine alleles and a heterozygosity of 0.73, linked at different θ's to the disease locus. The results indicated that linkage would be detected resulting in a LOD score >3.3 (at θ=0 and assuming locus homogeneity) with both the entire collection of 70 PCD families and with the subset of 52 KS families, but not with the 18 CDO families (the expected maximal LOD score in the CDO subset was below the value of 2, reflecting the low power of detecting linkage in those families).

A genome-wide scan was performed with 462 microsatellite markers. In the 52 KS families, one single marker locus on the long arm of chromosome 15 (D15S205) revealed a LOD score >2; the maximal LOD score value for D15S205 was 3.36, at θ=0.05. As expected, none of the markers showed pairwise LOD scores >2 in the set of 18 CDO families; combining both subsets (that is, the KS and CDO families) resulted in pairwise LOD scores below the level of statistical significance. This preliminary analysis was followed by typing the 52 KS families with 62 additional microsatellite markers spanning ~20 cM surrounding D15S205, roughly corresponding to 21 Mb of the total physical distance. The highest pairwise LOD score value of 4.34 (at θ=0) was obtained for D15S154 located on 15q24–25 (Table 1).

To examine whether all the analysed KS families were likely to be linked to this chromosome 15q region, we performed tests of heterogeneity (HOMOG software). The test results supported heterogeneity (H2: linkage, heterogeneity *v* H1: linkage, homogeneity; likelihood ratio 487.33) indicating that chromosome 15q is likely to be involved in 60% of the KS families. Six pedigrees out of 52 were indicated as being unlinked to the region. Analysis of haplotypes (not shown) revealed that in four of those pedigrees affected and unaffected siblings shared both parental haplotypes, while in two other pedigrees affected siblings had different

parental haplotypes. Another 15 pedigrees were not informative and were eliminated from further calculations. Analysis of the remaining 31 pedigrees confirmed linkage under homogeneity (H1: linkage, homogeneity *v* H0: no linkage; likelihood ratio 8.511×106).

Table 1. Pairwise linkage findings for the high density follow up on chromosome 15 in the KS families.

Marker[a]	cM	LOD-all[b]	Theta-all[b]	LOD-31fam[c]	Theta-31fam[c]
D15S967	75.67	0.702202	0.15	1.546949	0.1
D15S131	76.52	0.720699	0.1	2.470191	0.05
D15S204	77.60	1.466857	0.05	2.880023	0.05
D15S197	77.69	0.851782	0	2.018382	0
D15S124	78.37	0.278873	0.15	1.658671	0.05
D15S192	78.38	1.036385	0.1	2.333415	0.05
D15S818	79.03	1.016548	0.1	2.980691	0.05
D15S984	80.37	2.860638	0.05	3.80305	0.05
D15S989	83.75	0.445265	0.15	2.075293	0
D15S1005	84.19	2.339443	0.05	5.758866	0
D15S969	84.31	1.373751	0.05	3.693551	0
D15S1037	85.81	2.860638	0.05	3.80305	0.05
D15S1047	85.83	1.180228	0.05	2.358563	0.05
D15S211	85.84	2.248682	0.05	3.033275	0.05
D15S206	87.02	2.35253	0	2.024841	0.05
D15S115	87.24	3.641433	0	2.124683	0.1
D15S205	89.09	3.364761	0.05	4.176748	0.05
D15S154	89.10	4.344967	0	2.636754	0.05
D15S653	89.22	1.681151	0.05	1.591634	0.1

D15S539	89.22	1.651598	0.05	2.47482	0.05
D15S152	89.30	3.506446	0	2.372452	0.1
D15S972	89.31	2.727519	0.05	3.466998	0
D15S999	89.30	1.290613	0.05	2.626094	0.05
D15S979	93.21	2.15897	0.05	2.175463	0.1
D15S1045	94.83	2.358117	0	2.914872	0
D15S116	95.49	2.829704	0.05	2.067284	0.1
D15S127	97.26	0.775935	0.1	1.525221	0.1

[a]Only markers with LOD>1.5 are shown; [b] Calculated in 52 KS families; [c] Calculated in 31 KS families (without the 6 unlinked and 15 uninformative families).

Results of the pairwise LOD scores obtained for the 31 KS families "homogenised" for their linkage to 15q are shown in Table 1. The highest pairwise LOD score value (5.75) was obtained for D15S1005. The result of the multipoint linkage analysis performed on the homogenised set of 31 KS families (Figure 1) confirmed the pairwise LOD score results and excluded close linkage of more distantly located marker loci with KS. The multipoint LOD scores reduced abruptly at the genetic map positions 80.8 and 87.2 cM (at markers D15S991 and D15S115, respectively). The 95% confidence interval ($LOD_{max}-1$) localised the putative KS gene to a 3.5 cM region between D15S973 and D15S1037.

Interestingly, no extended common haplotype encompassing marker D15S1005 could be detected among the affected members of different KS families (Figure 2). Two longest common haplotypes, one encompassing three markers to the left and two right of D15S1005, and another encompassing five markers to the right of D15S1005, were shared by only two families each. Even if only three markers including D15S1005 were considered, a common haplotype was shared by at most four families. This

contradicts the notion of the ancestral haplotype carrying the major common mutation.

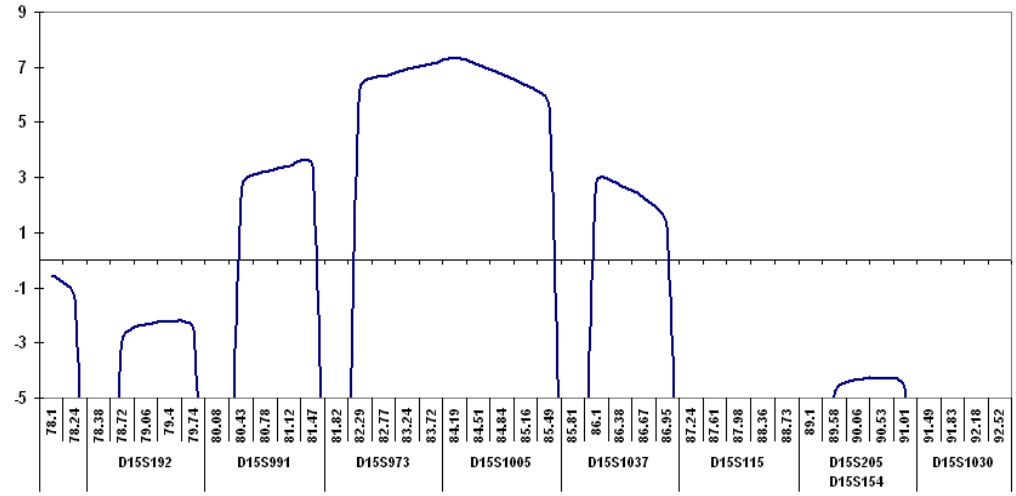

Figure 1. The multipoint LOD score plot for 31 KS families. Disease gene frequency 0.005; penetrance 1.0; phenocopies 0.0

Both pairwise and multipoint LOD scores calculated for the set of 18 CDO families were negative for the entire KS region, excluding close linkage of this region to PCD without situs inversus. Even when genetic heterogeneity within the CDO subgroup of PCD was assumed, allowing only some of the families to be linked to this region, the maximum pairwise LOD scores for all markers in this region were still non-significant (not shown).

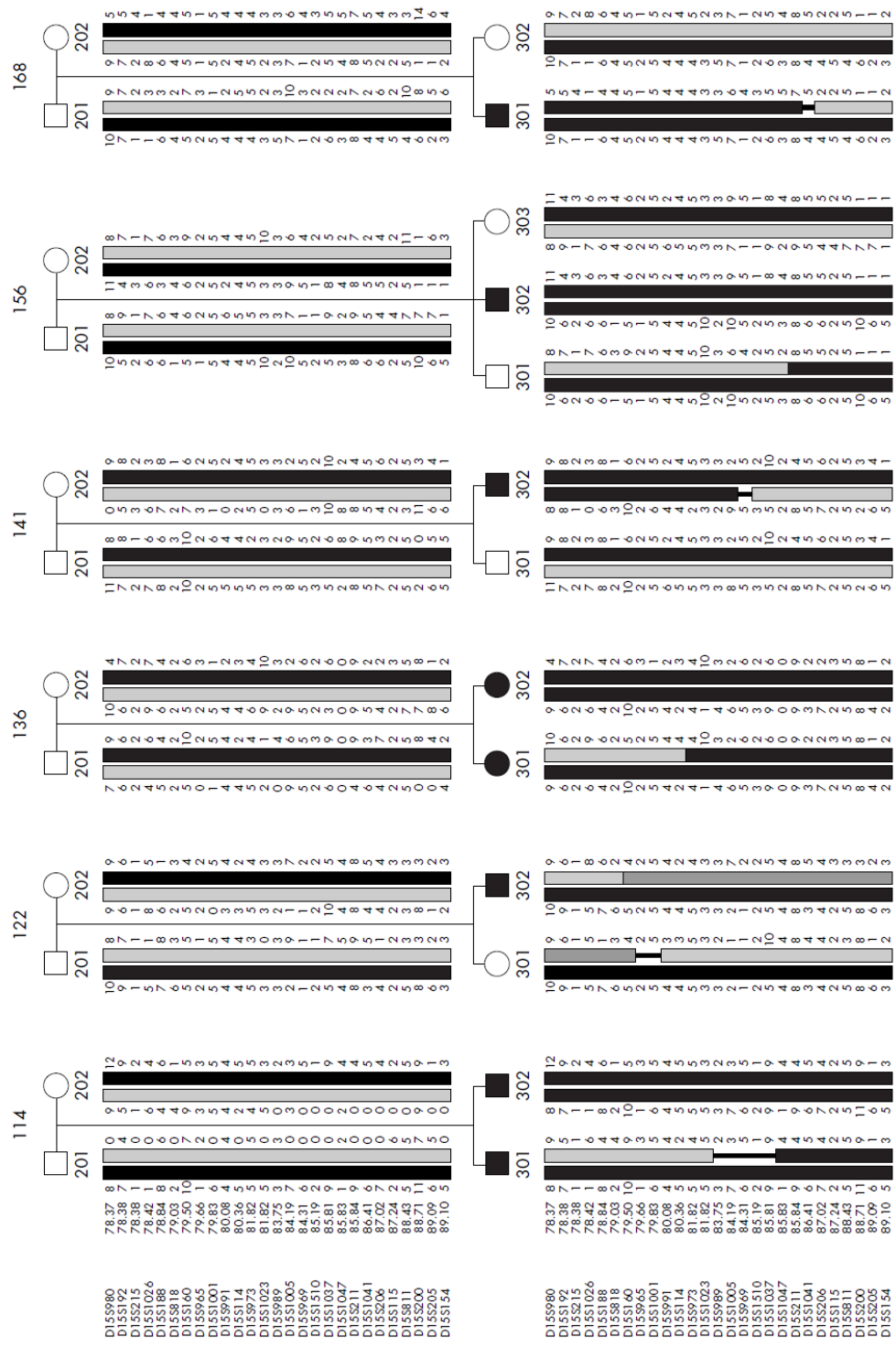

Figure 2. Segregation of haplotypes in six selected families, thus narrowing the region of interest

Discussion

In our previous study on Polish PCD families, we excluded linkage of a KS locus to chromosome 7,[20] contrary to the suggestion of other authors.[21] At the same time, however, our results provided suggestive evidence of linkage to 7p15 in the subset of PCD families without situs inversus. Our disparate results for linkage of PCD to chromosome 7 in the KS and CDO families[11] indicated that the genetic basis of these two forms of PCD might be different, consistent with some postulated animal models such as the *hop* mouse mutation, which exhibits CDO without situs inversus.[28] In a search for putative PCD genes on other candidate chromosome(s), we examined our collection of PCD families after splitting them into subsets of KS and CDO families. Both the genome scan and the follow-up linkage analyses confirmed linkage to chromosome 15q in the KS but not in the CDO families. Importantly, when both subsets of PCD families, that is, the KS and CDO families, were combined, the pairwise LOD scores for the markers that were significant in the KS families, dropped below the level of statistical significance. This indicated that the failure to find linkage in CDO families, rather than merely reflecting the low number of families, revealed the real lack of linkage of the analysed chromosome 15 region to CDO. These gene mapping findings are consistent with the hypothesis that there is no single gene of major effect that controls the risk for both forms of PCD. Had such a gene existed, the combined set of the KS and CDO families would have yielded stronger support for the linkage.

Linkage analysis in genetically heterogeneous recessive disorders presents a difficult task. Large pedigrees with many affected persons are rare and heterogeneity obscures the linkage signal by introducing noise from unlinked families. We have successfully applied a strategy that was used in mapping Bardet-Biedl syndrome, a heterogeneous recessive disorder with eight loci reported so far.[29, 30] Once the threshold of LOD score value showing significant linkage to chromosome 15q was reached and confirmed in the follow up analysis with more markers, we focused our efforts on homogenising the analysed set of KS families. Statistical analysis of

heterogeneity, performed under a more stringent disease model with full penetrance and no phenocopies, indicated six families with negative LOD scores in the D15S1005 region. Excluding those families resulted in the increased LOD scores for D15S1005 and its closest neighbours and enabled us to narrow the region containing the putative KS gene. As can be seen in Table 1, by removing the unlinked families we eliminated noise from our data: LOD scores maximise at $\theta=0$ for the immediate neighbours of D15S1005 and at $\theta=0.05$ for the more distant markers within 10 cM.

Another genome-wide scan and subsequent fine mapping in Israeli Druze PCD families pointed to a PCD locus at 15q13.1–15.1.[19] Although in Druze families several patients exhibited situs inversus and partial loss of inner dynein arms, the 15q region implicated in KS is located proximally from the region identified in this study.

The physical size of the KS candidate region defined by the physical map position of the markers D15S973 and D15S1037 is ~2.8 Mb. This candidate region contains 35 genes and coding sequences, but no known functional candidate genes with a clear involvement in the ciliary structure and/or relevant to KS pathogenesis were identified among them. However, a number of interesting candidates reside in this region. The gene coding for the cellular retinoic acid binding protein 1 (*CRABP1*), located at 15q24, is one of the potential candidates. Retinoic acid is an important factor during development of the organism.[31] It is involved in promoting development of ciliated cells in lungs,[32] and also plays a role in spermatogenesis; moreover, changes in its concentration in the murine embryo may lead to situs aberrations.[33] Other potential candidates include the *MESDC1* and *MESDC2* (mesodermal development candidates 1 and 2) genes located at 15q25: *MESDC1* might be involved in anterior-posterior axis development in mouse.[34] The potential role of these genes in the pathogenesis of KS remains to be elucidated.

In conclusion, our linkage analysis confirms the concept of locus heterogeneity in the pathogenesis of PCD, at least when its different forms – with

and without situs inversion – are considered. Most likely localisation of the purported KS locus is within the 3.5 cM region on 15q24–q25.

Acknowledgements

The authors thank all PCD families participating in this study for their invaluable cooperation, and J Pawlik (Rabka, Poland) and A Kapellerova (Bratislava, Slovakia) for providing study material and clinical data. The constant encouragement and support of Robert P Erickson (University of Arizona, Tucson) and the help of Giovanni Malerba (University of Verona) in the interpretation of statistical data are gratefully acknowledged.

References

1. Witt M. *DNAI1* and *DNAH5* and primary ciliary dyskinesia and Kartagener syndrome. In: *Inborn Errors of Development: The Molecular Basis of Clinical Disorders of Morphogenesis*; C.Epstein, R.P.Erickson, A.Wynshaw-Boris (eds), Oxford University Press, New York. 2004:934-945.
2. Geremek M, Witt M. Primary ciliary dyskinesia: genes, candidate genes and chromosomal regions. *J Appl Genet* 2004; 3:347-61.
3. McKusick V. ed. (2004) Kartagener syndrome (244400) In: Online Mendelian Inheritance of Man. http://www.ncbi.nlm.gov/htbin-post/Omim.
4. Bush A, Cole P, Hariri M *et al*. Primary ciliary dyskinesia: diagnosis and standards of care. *Eur Respir J* 1998; 12:482-488.
5. Afzelius BA, Mossberg B. Immotile-cilia syndrome (primary ciliary dyskinesia) including Kartagener syndrome. In: Scriver C, Beaudet AL. Sly W, Valle D (eds) The Metabolic and Molecular Bases of Inherited Diseases. 7th ed McGraw-Hill, New York, 1995: 3943-3954.
6. Schidlow DV. Primary ciliary dyskinesia (the immotile cilia syndrome). *Ann Allergy* 1994; 73:457-470.

7. Rutland J, de Iongh RU. Random ciliary orientation: a cause of respiratory tract disease. *New Eng J Med* 1990; 323:1681-1684.
8. Narayan D, Krishnan SN, Upender M *et al*. Unusual inheritance of primary ciliary dyskinesia (Kartagener syndrome). *J Med Genet* 1994; 31:493-496.
9. Krawczynski MR, Witt M PCD and RP: X-linked inheritance of both disorders?
   Pediatr Pulmonol 2004; 38:88-9.
10. Krawczynski MR, Dmenska H, Witt M. Apparent X-linked primary ciliary dyskinesia associated with retinitis pigmentosa and a hearing loss. *J Appl Genet* 2004; 45:107-10.
11. Luck DJL, Huang B, Piperno G. Genetic and biochemical analysis of the eukaryotic flagellum. *Soc Exp Biol Symp* 1982; 35:399-407.
12. Pennarun G, Escudier E, Chapelin *et al*. Loss-of-function mutations in a human gene related to *Chlamydomonas reinhardtii* dynein IC78 result in primary ciliary dyskinesia. *Am J Hum Genet* 1999; 65:1508-1519.
13. Guichard C, Harricane M-C, Lafitte J-J *et al*. Axonemal dynein intermediate-chain gene (DNAI1) mutations result in situs inversus and primary ciliary dyskinesia (Kartagener syndrome). *Am J Hum Genet* 2001; 68:1030-1035.
14. Zariwala M, Noone PG, Sannuti A *et al*. Germline mutations in an intermediate chain dynein cause primary ciliary dyskinesia. *Am J Respir Cell Mol Biol* 2001; 25:577-83.
15. Olbrich H, Häffner K, Kispert A *et al*. Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. *Nat Genet* 2002; 30:143-144.
16. El Zein L, Omran H, Bouvagnet P. Lateralization defects and ciliary dyskinesia: lessons from algae. *Trends Genet* 2003; 19:162-167.
17. Blouin JL, Meeks M, Radhakrishna U *et al*. Primary ciliary dyskinesia: a genome-wide linkage analysis reveals extensive locus heterogeneity. *Eur J Hum Genet* 2000; 8:109-18.

18. Meeks M, Walne A, Spiden S *et al*. A locus for primary ciliary dyskinesia maps to chromosome 19q. *J Med Genet* 2000; 37:241-244.
19. Jeganathan D, Chodhari R, Meeks M *et al*. Loci for primary ciliary dyskinesia map to chromosome 16p12.1-12.2 and 15q13.1-15.1 in Faroe Islands and Israeli Druze genetic isolates. *J Med Genet* 2004; 41:233-40.
20. Witt M, Wang Y-F, Wang S *et al*. Exclusion of chromosome 7 for Kartagener syndrome but suggestion of linkage in families with other forms of primary ciliary dyskinesia. *Am J Hum Genet* 1999; 64:313-318.
21. Pan Y, McCaskill CD, Thompson KH, Hicks *et al*. Paternal isodisomy of chromosome 7 associated with complete situs inversus and immotile cilia. *Am J Hum Genet* 1998; 62:1551-1555.
22. Bartoloni L, Blouin J-L, Pan Y, Gehrig C *et al*. Mutations in *DNAH11* (axonemal heavy chain dynein type 1) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia. *Proc Natl Acad Sci USA* 2002; 99:10282-10286
23. Kong A, Gudbjartsson DF, Sainz J *et al*. A high-resolution recombination map of the human genome. *Nat Genet* 2002; 3:241-7.
24. Ziegle JS, Su Y, Corcoran KP *et al*. Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* 1992; 14:1026-1031.
25. Schäffer AA. Faster linkage analysis computations for pedigrees with loops or unused alleles. *Hum Hered* 1996; 46:226-235.
26. Ott J. Analysis of human genetic linkage. 3rd rev ed Baltimore: Johns Hopkins
University Press 1999.
27. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and non-parametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; 58:1347-1363.
28. Handel MA. Allelism of *hop* and *hpy*. *Mouse News Lett* 1985; 72:124.

29. Leppert M, Baird L, Anderson KL *et al*. Bardet-Biedl syndrome is linked to DNA markers on chromosome 11q and is genetically heterogeneous. *Nature Genet* 1994; 7:108-112.
30. Katsanis N, Lewis RA, Stockton DW *et al*. Delineation of the critical interval of Bardet-Biedl syndrome 1 (BBS1) to a small region of 11q13, through linkage and haplotype analysis of 91 pedigrees. *Am J Hum Genet* 1999; 65:1672-9.
31. Smith SM, Dickman ED, Power SC, Lancman J. Retinoids and their receptors in vertebrate embryogenesis. *J Nutr* 1998; 128(2 Suppl):467S-470S.
32. Million K, Tournier F, Houcine O *et al*. Effects of retinoic acid receptor-selective agonists on human nasal epithelial cell differentiation. *Am J Respir Cell Mol Biol* 2001; 25:744-50.
33. Van Keuren ML, Layton WM, Iacob RA, Kurnit DM. Situs inversus in the developing mouse: proteins affected by the iv mutation (genocopy) and the teratogen retinoic acid (phenocopy). *Mol Reprod Dev* 1991; 29:136-44.
34. Hsieh J-C, Lee L, Zhang L *et al*. Mesd encodes the LRP5/6 chaperone essential for specification of mouse embryonic polarity. *Cell* 2003; 112:355-367.

Chapter 3

Sequence analysis of 21 genes located in the Kartagener syndrome linkage region on chromosome 15q

Maciej Geremek, Frederieke Schoenmaker, Ewa Zietkiewicz, Andrzej Pogorzelski, Scott Diehl, Cisca Wijmenga, Michal Witt

Eur J Hum Genet 2008;16:688-95

Abstract

Primary ciliary dyskinesia (PCD) is a rare genetic disorder, which shows extensive genetic heterogeneity and is mostly inherited in an autosomal recessive fashion. There are four genes with a proven pathogenetic role in PCD. DNAH5 and DNAI1 are involved in 28 and 10% of PCD cases, respectively, while two other genes, DNAH11 and TXNDC3, have been identified as causal in one PCD family each. We have previously identified a 3.5 cM (2.82 Mb) region on chromosome 15q linked to Kartagener syndrome (KS), a subtype of PCD characterized by the randomization of body organ positioning. We have now refined the KS candidate region to a 1.8 Mb segment containing 18 known genes. The coding regions of these genes and three neighboring genes were subjected to sequence analysis in seven KS probands, and we were able to identify 60 single nucleotide sequence variants, 35 of which resided in mRNA coding sequences. However, none of the variations alone could explain the occurrence of the disease in these patients.

Introduction

Primary ciliary dyskinesia (PCD, MIM no. 242650) is a rare genetic disorder caused – as far as is known – by mutations in genes encoding proteins important for ciliary beating.[1] The major clinical consequences fall into three categories: (1) in the respiratory tract, immotility or dyskinetic beating of cilia in epithelial cells impairs the mucociliary clearance, leading to recurrent infections, sinusitis and bronchiectasis; (2) in the urogenital tract, infertility occurs in some patients because of the dysmotility of sperm tails; (3) mirror reversal of body organ positioning (situs inversus) is seen in approximately half of PCD patients. It is believed to occur at random and to be caused by the immotility of primary cilia of cells on the ventral surface of the embryonic node. The presence of situs inversus in an affected family defines a PCD subtype, known as Kartagener syndrome (KS, MIM no. 244400).

For diagnostic purposes, cilia in bronchial scrapings can be visualized under a light microscope, whereas the clinical diagnosis of PCD is routinely verified by electron microscopy analysis of respiratory cilia ultrastructure. About 20 different ultrastructural defects have been described in PCD patients, with lesions of outer and/or inner dynein arms being the most frequent defects of the internal anatomy of cilia.[2, 3] Recently, it has been shown that immunofluorescence staining of ciliated epithelium with antibodies targeting DNAH5 can detect outer dynein arms defects and therefore aid diagnosis of PCD.[4]

Inheritance of PCD is autosomal recessive in most cases, although pedigrees with an X-linked mode of inheritance have also been described.[5, 6] To date, mutations in four genes have been found as causative for PCD, exclusively in patients with outer dynein arm defects. *DNAI1*, coding for intermediate dynein chain 1, was selected for mutation screening based on homology with the *Chlamydomonas reinhardtii* gene, which causes a slow-swimming phenotype when mutated,[7, 8] whereas *DNAH5*, coding for heavy dynein chain 5, was identified by homozygosity mapping in a large family and subsequent sequence analysis.[9] *DNAH5* and *DNAI1* are responsible for 28 and 10% of PCD cases, respectively,[10, 11] clearly indicating that other genes are involved in PCD/KS etiology. *DNAH11*,[12] coding for dynein heavy chain, and *TXNDC3*,[13] coding for a thioredoxin family member, were also found to be responsible for PCD, but so far each has been found in only one family.

Analysis of the ciliary proteome in unicellular alga *C. reinhardtii* led to the identification of a number of proteins potentially involved in cilia formation and maintenance: 360 proteins with high and 293 with moderate confidence.[14] Besides the known components of ciliary ultrastructure, there were also 90 signal transduction proteins, and many membrane proteins and metabolic enzymes. Given that our knowledge of ciliary genes comes mostly from studies of simple organisms and that the functional spectrum of the possible candidate genes is very wide, it is difficult to predict which of these should be considered as candidate genes for PCD.

On the basis of genetic linkage studies, extensive genetic heterogeneity of PCD is postulated. A total-genome scan performed in 31 multiplex families did not reveal any predominant locus, rather it showed several peaks with suggestive and indicative LOD scores.[15] Another genome scan, performed in five families of Arabic origin and with reported consanguinity, revealed linkage with chromosome 19q13.3 in only three of the families, thus confirming locus heterogeneity.[16] Two additional loci, on 16p12 and 15q, were indicated by studies on genetically isolated, but heterogeneous, populations from the Faroe Islands and the Israeli Druze.[17]

We earlier performed genome-wide linkage analysis in 52 families with KS and reported a region on chromosome 15q24–25, between D15S973 and D15S1037, linked to KS.[18] After obtaining a significant LOD score of 4.34 with D15S154 (using a strictly defined disease model), we defined a region of 3.5 cM (2.82 Mb) as containing the candidate gene.[18] In the present study, we have performed further fine-mapping of the region of interest and screened all 18 genes in the linked region and three neighboring genes for mutations in seven KS patients from the families that contributed most to the linkage results.[18]

Materials and methods

Biological material

DNA samples from the existing PCD collection[18] came from 31 Caucasian families (25 of Polish and 6 of Slovak origin) classified as KS families; none was from a genetically isolated population. Each family had at least one member diagnosed with PCD and exhibiting situs inversus, but with no other major anomalies or dysmorphologies present. The group studied consisted of 38 affected individuals and 99 unaffected family members. The primary complaints in these patients were derived from the respiratory tract and the clinical picture included symptoms of sinusitis, nasal polyps, bronchiectasis, recurrent infections of the upper respiratory tract. All cases were confirmed by a low concentration of nitric oxide (NO) measured in exhaled air.[19] NO was measured from the nasal cavity by a

chemiluminescence analyzer with a threshold value of 200 ppb for diagnosing PCD. In 60% of the families, the clinical diagnosis was confirmed by transmission electron microscopy analysis of bronchial cilia ultrastructure; the analysis revealed defects of outer or outer and inner dynein arms; a representative EM picture is shown in Supplementary figure 1. Direct microscopy of bronchial scrapings was used to confirm the clinical diagnosis of PCD in the patients that did not have electron microscopy imaging of cilia. KS inheritance in these pedigrees was in agreement with general assumptions for an autosomal recessive disease model. KS in these families has previously been shown to be linked with the region between D15S973 and D15S1037, assuming full penetrance, 0.0001 phenocopies and the disease allele frequency of 0.005. The genome-wide significant LOD score of 4.34 was obtained using a cohort of 52 families.[18] Pair-wise LOD score analyses were performed using the FASTLINK program, and multipoint LOD scores for markers on chromosome 15 were calculated using GENEHUNTER. Multipoint LOD scores allowing for locus heterogeneity were calculated using Simwalk2. The 95% confidence interval localized the KS locus to a 3.5 cM region between D15S973 and D15S1037.

Recombination mapping using microsatellite markers

To refine the candidate gene region between D15S973 and D15S1037, seven markers (D15S1027–380 kb–D15S524–33 kb–AFMA085AWB9–580 kb–D15S989–219 kb–D15S1005–58 kb–D15S969–695 kb–D15S551), were genotyped in all 38 affected individuals and 99 unaffected family members. Amplification products, obtained using 5′ end fluorescent-dye-labeled primers, were pooled and analyzed on an automated ABI Prism 3730 DNA sequencer (Applied Biosystems, Foster City, CA, USA). Genotypes were scored using GENMAPPER (v3.5NT) software. Inheritance was checked using PEDCHECK[20] and haplotypes in the 31 pedigrees were constructed from microsatellite data using SIMWALK2.[21]

Analysis of candidate genes

Sequence analysis was performed in seven affected individuals from the families with the highest LOD score for marker D15S1005.[18] Sequences of 135 exons and surrounding intron sequences of the 20 genes located in the region were obtained using Ensembl genome browser version 42 (Figure 1; Table 1). AL109678, not present in Ensembl, was also sequenced as it is an mRNA coding sequence annotated in the UCSC genome browser. At least 200 bp upstream of the transcription start site were also analyzed and the 3′UTR was sequenced in all genes except for *TMC3*, *KIAA1199*, *ARNT2* and *TMED3*. The Primer3 program was used for designing PCR primers yielding products of 400–500 bp;[22] primers were in the intronic sequences flanking the exons to be analyzed. PCR reactions were performed using standard protocols (data available upon request). Amplification products were purified, quantified on a 2% agarose gel, and diluted for direct sequencing on an automated ABI Prism 3730 DNA sequencer using BigDye Terminator Sequencing Standards (Applied Biosystems). Sequences were assembled using ContigExpress (Vector NTI suite v8.0, Informax Inc.).

Sequencing reactions were performed first with the forward primer, but if the quality of the sequencing electropherogram was unsatisfactory, or if single-nucleotide polymorphisms (SNPs) were found, they were repeated with the reverse primer. Each of the newly identified variants was validated through an independent PCR and sequencing reaction. The sequenced fragments were only included in the results if data were obtained for at least six of the seven patients. Seven reactions could not be amplified despite at least three attempts with two different sets of primers (nucleotides in mRNA sequence: BCL2A1, 602–710; FAM108C1, 1–343; C15ORF37, 421–640; TMED3, 260–299; ENSG00000180725, 457–618; FAH, 683–783, 991–1037). Reactions failed to use control as well as patients' DNA, indicating PCR problems related to the local sequence composition of those genomic fragments. All variations found in the protein-coding sequence and not previously described in public databases were sequenced in a control population group consisting of 48 unrelated, healthy individuals of Polish origin.

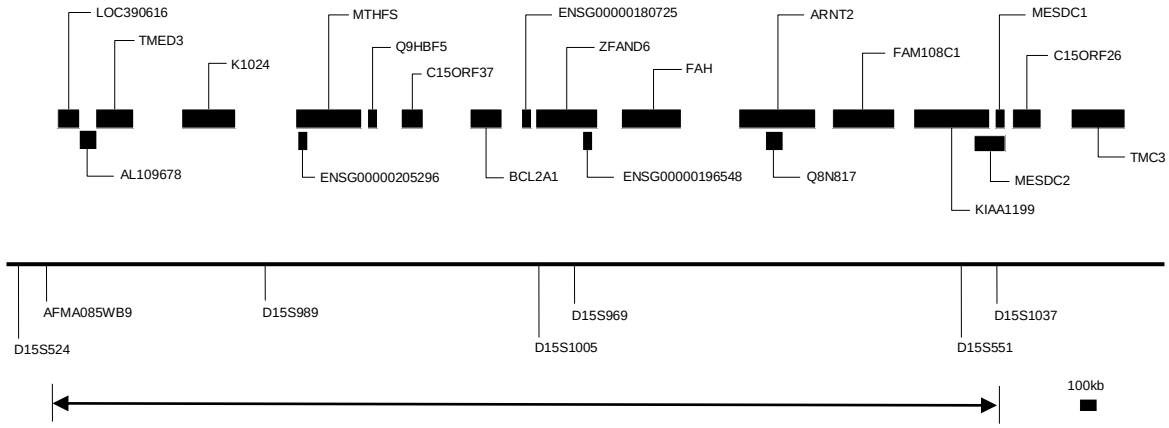

Figure 1. The location of 21 genes in the linkage region on chromosome 15q24-25. The minimal KS gene-containing region is indicated by an arrow.

Table 1. Genes included in the Kartagener syndrome sequencing study

Gene symbol	Gene name	GO description	# of exons	Transcript length (bp)
LOC390616	Unknown	Unknown	1	1608
AL109678	Unknown	Unknown	1	2540
TMED3	Transmembrane emp24 domain-containing protein 3 precursor (Membrane protein p24B)	protein transport; membrane; membrane; protein carrier activity	3	1420
K1024	UPF0258 protein KIAA1024	membrane; integral to membrane	4	6732
MTHFS	5-formyltetrahydrofolate cyclo-ligase (EC 6.3.3.2) (5,10-methenyl- tetrahydrofolate synthetase) (Methenyl-THF synthetase) (MTHFS)	magnesium ion binding; folic acid binding; ligase activity; metabolism	3	857
ENSG00000205296	Unknown	Unknown	2	153
Q9HBF5	Cervical cancer suppressor-1	Unknown	2	240
C15ORF37	C15orf37 protein	Unknown	1	2082
BCL2A1	Bcl-2-related protein A1 (Protein BFL-1) (Hemopoietic-specific early response protein) (Protein GRS)	regulation of apoptosis	3	624

ENSG00000180725	Unknown	Unknown	2	698
ZFAND6	Zinc finger A20 domain-containing protein 3 (Associated with PRK1 protein)	DNA binding; zinc ion binding; signal transduction; protein binding	7	1680
ENSG00000196548	Unknown	Unknown	3	123
FAH	Fumarylacetoacetase (EC 3.7.1.2) (Fumarylacetoacetate hydrolase) (Beta-diketonase) (FAA)	fumarylacetoacetase activity; magnesium ion binding; regulation of transcription, DNA-dependent; metabolism	15	1471
ARNT2	Aryl hydrocarbon receptor nuclear translocator 2 (ARNT protein 2)	transcription factor activity; signal transducer activity; aryl hydrocarbon receptor nuclear translocator activity; response to hypoxia; embryonic development	19	6550
Q8N817	CDNA FLJ40133 fis, clone TESTI2012231	Unknown	4	675
FAM108C1	Unknown	Unknown	3	2326
KIAA1199	KIAA1199	sensory perception of sound	29	6602
MESDC2	Mesoderm development candidate 2 (NY-REN-61 antigen)	mesoderm development	5	2938
MESDC1	Mesoderm development candidate 1	Unknown	1	2373
C15ORF26	Uncharacterized protein C15orf26	Unknown	7	1574
TMC3	Transmembrane channel-like protein 3	integral to membrane	19	2443

Results

Seven microsatellite markers located in the 3.5 cM region reported to be linked to KS were genotyped in 31 families that had positive LOD scores for chromosome 15q in our genome linkage scan[18] (Figure 1). Haplotype analysis was concordant with linkage to chromosome 15q24–25 (data not shown). A recombination was

mapped in family numbers 114 and 126 (Figure 2), 1 Mb more telomerically from the previous location defined by marker D15S937, thus narrowing the minimal gene-containing region down to 1.8 Mb, between markers AFMA085AWB9 and D15S1037.

The analysis of haplotype sharing between nonrelated KS patients was inconclusive. The most promising block, encompassing three markers (D15S524, AFMA085WB9 and D15S989) and spanning 0.6 Mb, was present on five nonrelated, affected chromosomes (families 103, 117, 136, 142 and 147) and on one untransmitted parental chromosome (data not shown).

Seven families with the highest positive LOD scores were selected for sequence analysis of all the genes located in the 1.8 Mb region (Figure 2). These families contributed 70% of the LOD score for marker D15S1005 on chromosome 15q,[18] and it was interesting that the *DNAI1* mutation analysis was negative in all seven families. The *DNAH5* contribution was excluded in four families (114, 117, 136 and 147; Figure 2), based on the analysis of SNPs located within the gene (discordant genotypes in the affected sibs), while in the three remaining families, the analyzed SNPs in *DNAH5* were uninformative (unpublished data). As the microsatellite data confirmed linkage of the disease to 15q24–25 in all seven families, only one affected individual, representative of each pedigree, was subjected to DNA sequencing. Supplementary table 1 summarizes the clinical findings of the seven families included for sequencing.

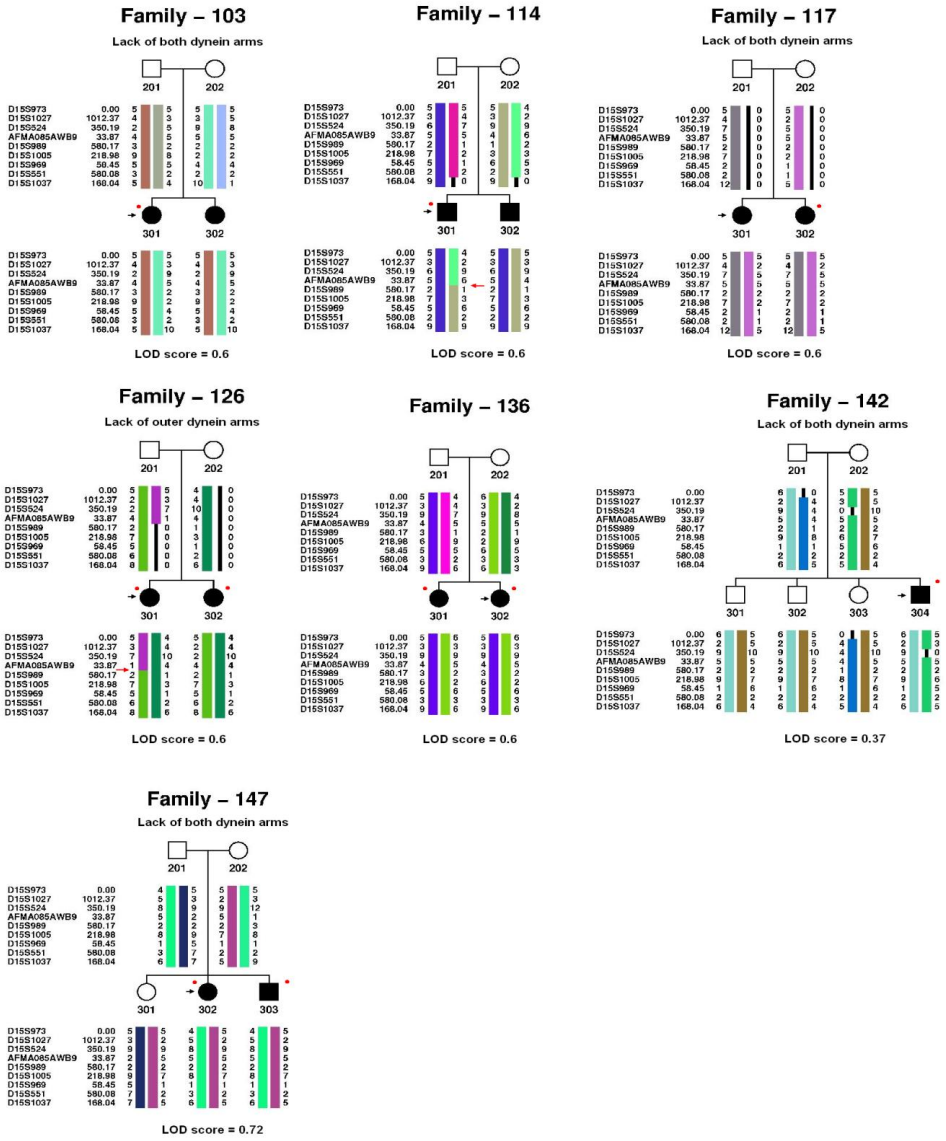

Figure 2. Pedigrees of the KS families included in the sequencing. The results of the pair-wise LOD score for marker D15S1005 are indicated below each family. The recombination (marked by a red arrow) defining the minimal KS gene-containing region can be observed in pedigrees #114 and #126. Individuals included in the sequencing are indicated by black arrows. Red dots indicate individuals exhibiting situs inversus. The numbers beside marker names indicate the intermarker distances in kilobase pairs. Missing data are denoted by zeros. The genotypes of the parents in pedigrees #117 and #126 were deduced from the genotypes of the siblings.

We analyzed the 18 coding sequences located between markers AFMA085WB9 and D15S1037 (Table 1). We also sequenced three genes located in the vicinity of the region, which could be functionally related to cilia: *MESDC2*, which is important in embryonic development; *C15ORF26*, which is expressed in lung and testis and displays homology to a flagellar protein identified in *Chlamydomonas*;[14] and *TMC3*, a membrane protein with partial homology to nephrocystine, as indicated by Blast search of the NCBI nr database (Figure 1). Except for *C15ORF26*, none of the genes have been reported to be a part of cilium.

In total, we analyzed 83 kb of genomic sequence in seven individuals (35 kb of coding sequences and 48 kb of intronic/flanking DNA). We found 60 SNPs, 45 of which were already present in public SNP databases. Twelve SNPs, including the four previously unknown ones, were located in the protein coding regions. None of the identified SNPs in the coding regions resulted in a stop codon change or a frame shift; six were amino-acid substitutions and seven were synonymous codon changes. Twenty-one SNPs, including seven novel ones, were located in the untranslated regions. Four SNPs were located in regions upstream of the transcription start site. Twenty-seven variations were located in noncoding regions, six of which were previously unknown. None of the SNPs found were closer than 10 bp to a splicing site. Variations in the protein-coding sequence, previously not described in SNP databases, were sequenced in 48 individuals of Polish origin to estimate population frequencies. Except for two SNPs (W399R in *ARNT2* and A506G in *AWP1*), all the variations were found in the healthy controls. A complete list of the SNPs found is given in Table 2 and the distribution of all SNPs in the patients can be found in Supplementary table 2.

Discussion

We have performed a follow-up study to our previous finding[18] of linkage to chromosome 15q24–25 in families with KS. Fine-mapping of the 3.5 cM (2.82 Mb) linked region in 31 KS families allowed us to narrow down the minimal candidate

gene region to 1.8 Mb. Analysis of haplotypes composed of microsatellite markers revealed that there was no extensive haplotype sharing among the affected individuals, arguing against the presence of only one or a few frequent mutations occurring on a common haplotype background. On the other hand, the absence of haplotype sharing does not exclude the possibility of the presence of a causative gene or genes in the linked region. The linkage analysis is independent of a founder effect. There are examples in literature of genes involved in rare recessive diseases, but not showing evidence of haplotype sharing. This happens because most of the patients are compound heterozygous for unique mutations in the given gene.[23]

Table 2. List of all sequence variations found in 7 Kartagener syndrome patients.

Gene	Variation	Amino acid	Exon\intron	ID in SNP database	Minor allele frequency in NCBI genome database 35	Minor allele frequency in 48 control individuals of Polish origin	Number of patient chr with the minor allele
LOC39016	c.735C>T	P245P	exon 1	rs8038778	T=0.367		7/14
TMED	c.659G>A	A172A	exon 3	rs906439		A=0.0	0/14
KIAA1024	c.847A>C	N258H[a]	exon 2	-		C=0.06	2/14
	c.918C>A	A281A	exon 2	-			1/14
	c.2569A>G	I832V[b]	exon 3	rs2297773	G=0.208		9/14
	c*1799A>C	-	3'UTR	rs17266017	C=0.375		2/14
	c.*3771_3772 insATG	-	3'UTR	-		Ins=0.06	2/14
BCL2A1	c.238G>A	C19Y[c]	exon 1	rs1138357	A=0.146		3/14
	c.299T>G	N39K[d]	exon 1	rs1138358	G=0.146		3/14
MTHFS	g.411A>G	-	Upstream	rs2865825	-		1/14
	g.421C>G	-	Upstream	rs2903105	-		3/14
	g.483delG	-	Upstream	rs35401897	-		2/14
C15ORF37	c.903C>A	P166Q[f]	exon 1	rs3803540	A=0.217		1/14
	c.*46G>T	-	3'UTR	rs17214656	T=0.167		2/14
	c.*124A>C	-	3'UTR	rs36070199	-		2/14
	c.*206A>G	-	3'UTR	rs12442408	G=0.250		1/14
	c.*413A>T	-	3'UTR	rs2733101	T=0.283		2/14
AWP1	c.295-90T>C	-	intron 2	rs1916048	C=0.217		5/12
	C574+11C>G	-	intron 5	rs1522636	G=0.261		9/14
	c.466-60A>T	-	intron 4	rs12915043	-		1/14
	c.506A>G	Q65Q	exon 4	-		G=0.0	1/14
FAH	c.915-173delC	-	intron 10	rs3835063	-		6/14

	c.1133C>T	S352S	exon 13	rs1801374	T=0.058		2/14
	c.1038-35A>C	-	intron 12	rs2043691	C=0.08		1/14
	c.1038-223C>T	-	intron 12	rs2043692	T=0.15		1/14
	c.*37A>C	-	3'UTR	rs1049181	-		2/14
	c.*38C>T	-	3'UTR	-	-		2/14
	c.*39C>T	-	3'UTR	rs3210172	-		2/14
	c.*40C>T	-	3'UTR	-	-		2/14
	c.*41C>G	-	3'UTR	-	-		2/14
	c.*50delC	-	3'UTR	-	-		2/14
	c. *93T>C	-	3'UTR	rs1049194	-		2/14
ARNT2	c.312+15C>T	-	intron 3	rs2278708	T=0.058		1/14
	c.361-28insC	-	intron 3	rs5814026	-		3/14
	C574+36C>G	-	intron 4	rs8033507	G=0.017		1/14
	c.892-74C>T	-	intron 7	rs11072922	T=0.175		2/14
	c.1255+21C>T	-	intron 11	rs3924894	T=0.405		8/14
	c.1255+22T>C	-	intron 11	rs3924893	C=0.4		6/14
	c.1279T>A	W399R[e]	exon 12	-		A=0.0	1/14
FAM108C1	c.*692T>C	-	3'UTR	rs1046417	C=0.130		2/14
	c.*1115C>A	-	3'UTR	rs11072940	A=0.104		2/14
KIAA1199	c.-255T>A	-	5'UTR	rs2273886	A=0.295		4/14
	c.1029+26C>G	-	intron 7	-	-		7/14
	c.1346+240C>T	-	intron 9	-	-		1/14
	c.1479+83T>A	-	intron 10	-	-		1/14
	c.2057+57A>G	-	intron 13	rs2271164	G=0.417		5/12
	c.2142T>C	C28C	exon 14	rs2271160	C=0.261		3/14
	c.2462A>C	-	intron 16	rs2271162	C=0.167		3/14
MESDC2	c.*3G>T	-	3'UTR	rs8039607	T=0.121		1/14
	c.*53G>T	-	3'UTR	rs8039384	T=0.142		1/14
	c.1369-20_1369-1416C>T	-	intron 4	-		Del=0.08	1/14
	c.*2153T>C	-	3'UTR	-		T=0.069	1/14
C15ORF26	G496C>G	-	upstream	rs10519305	C=0.017		1/14
	c.505-105insA	-	intron 3	rs2683255	-		3/14
	c.505-87C>G	-	intron 3	-	-		3/14
	c.664-50insA		intron 5	rs11395672	-		7/14
TMC3	c.-13T>C	-	5'UTR	-		C=0.08	1/14
	c.580+55T>C		intron 6	rs12903128	C=0.345		1/14
	c.823-66A>G		intron 7	rs7167034	A=0.153		7/14

Nucleotide variations are presented according to the HUGO nomenclature. Translation start site was used for each gene as the +1 nucleotide. Putative effect on protein: [a] small, polar>polar, charged, positive, aromatic; [b] aliphatic>aliphatic, small; [c] tiny>aromatic; [d] small>positive; [e] aromatic, polar>polar, positive; [f] small>polar.

In searching for the KS gene in the 15q24-25 region, we found 37 SNPs in the mRNA-coding sequences and promoters of 14 genes, including 6 nonsynonymous codon changes. No evidently pathogenic mutations, such as the introduction of stop codons or frame-shift mutations, were detected.

Given the rare occurrence of the disease (approximately 1:40 000), the expected cumulative frequency of all the disease variant(s) in the general population should be below 1%. We also sequenced fragments harboring variants in mRNA-coding sequence in 96 control chromosomes (Table 2). One nonsynonymous amino-acid change (W399R in *ARNT2*), and one silent codon change (A506G in *AWP1*) were not seen in the controls, suggesting that their frequencies might be as low as 1%. However, each of these two variants (in two different genes) was found in heterozygous form in only a single patient, with no other possibly pathogenic complementary allele; these are therefore unlikely to represent the disease-causing mutations. None of the remaining SNPs were particularly rare and their occurrence in patients did not significantly deviate from population frequencies (Table 2).

In conclusion, we did not find any variants that directly fulfilled the criteria of a pathogenic mutation for KS in any of the 21 genes located in the minimal gene-containing region, as indicated by linkage analysis (while assuming the rare occurrence of the disease and the recessive mode of inheritance). The analyzed region may, however, contain pathogenic variants that we failed to find. We did not sequence introns that could contain regulatory elements or harbor gain-of-splicing site mutations. It is also possible that some of the genes have alternative exons not annotated in the current databases and thus not analyzed in our study, and we cannot exclude the possibility that some of the genes harbor heterozygous deletion(s) of one or more exons. We did not detect any extended homozygous regions, in neither the microsatellite data nor in the SNP data, but the relatively low density of heterozygous SNPs could have left some regions noninformative with respect to their purported hemizygosity (see Supplementary table 2).

As the pedigree structures in the cohort we studied were simple, we cannot exclude the possibility that some of the families are false positives, although none of the families was crucial for the significant LOD score. Hence, despite the statistical significance of the LOD score value, it cannot entirely be excluded that the 15q locus is a false positive locus. The current location of the linkage interval is based on two families (114 and 126; Figure 2) that contribute to the left recombination and on one family[18] that contributes to the right recombination. This last family contains only one affected individual and, therefore, contributed very little to the overall LOD score. Under less stringent disease models, the linkage peak would still be significant with a 95% confidence interval matching the minimal gene-containing region. However, the linkage curve would be positive over a larger region with more than 200 genes. Hence, we cannot exclude that the KS gene is located closer to the 15q telomere. Given that this larger region does not contain obvious candidate genes, more families will be needed to establish the exact KS region.

Supplementary information

https://www.dropbox.com/s/ad90beg26f1n1jb/suppl_data.doc

Supplementary table 1. Clinical findings in 7 families selected for sequencing.

Supplementary table 2. List of variations found in the seven KS patients.

Supplementary figure 1. A representative EM picture showing lack of outer and inner dynein arms.

References

1. Geremek M, Witt M. Primary ciliary dyskinesia: genes, candidate genes and chromosomal regions. *J Appl Genet* 2004; 3:347-61.

2. Zariwala MA, Knowles MR, Omran H. Genetic Defects in Ciliary Structure and Function. *Annu Rev Physiol* 2007; 69:423-50.
3. Van's Gravesande KS, Omran H. Primary ciliary dyskinesia: clinical presentation, diagnosis and genetics. *Ann Med* 2005; 37:439-49.
4. Fliegauf M, Olbrich H, Horvath J *et al*. Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. *Am J Respir Crit Care Med*. 2005; 171:1343-9.
5. Krawczynski MR, Witt M. PCD and RP: X-linked inheritance of both disorders? *Pediatr Pulmonol* 2004; 38:88-9.
6. Moore A, Escudier E, Roger G *et al*. RPGR is mutated in patients with a complex X linked phenotype combining primary ciliary dyskinesia and retinitis pigmentosa. *J Med Genet* 2006; 43:326-33.
7. Guichard C, Harricane M-C, Lafitte J-J *et al*. Axonemal dynein intermediate-chain gene (DNAI1) mutations result in situs inversus and primary ciliary dyskinesia (Kartagener syndrome). *Am J Hum Genet* 2001; 68:1030-1035.
8. Pennarun G, Escudier E, Chapelin *et al.* Loss-of-function mutations in a human gene related to *Chlamydomonas reinhardtii* dynein IC78 result in primary ciliary dyskinesia. *Am J Hum Genet* 1999; 65:1508-1519.
9. Olbrich H, Häffner K, Kispert A *et al.* Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. *Nat Genet* 2002; 30:143-144.
10. Hornef N, Olbrich H, Horvath J. DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. *Am J Respir Crit Care Med* 2006; 174:120-6.
11. Zariwala MA, Leigh MW, Ceppa F. Mutations of DNAI1 in primary ciliary dyskinesia: evidence of founder effect in a common mutation. *Am J Respir Crit Care Med* 2006; 174:858-66.
12. Bartoloni L, Blouin JL, Pan Y *et al.* Mutations in the DNAH11 (axonemal heavy chain dynein type 11) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia. *Proc Natl Acad Sci*; 99:10282-6.

13. Duriez B, Duquesnoy P, Escudier E *et al.* A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia. *Proc Natl Acad Sci*; 104:3336-41.
14. Pazour GJ, Agrin N, Leszyk J, Witman GB. Proteomic analysis of a eukaryotic cilium. *J Cell Biol* 2005; 170:103-13.
15. Blouin JL, Meeks M, Radhakrishna U *et al.* Primary ciliary dyskinesia: a genome-wide linkage analysis reveals extensive locus heterogeneity. *Eur J Hum Genet* 2000; 8:109-18.
16. Meeks M, Walne A, Spiden S *et al.* A locus for primary ciliary dyskinesia maps to chromosome 19q. *J Med Genet* 2000; 37:241-244.
17. Jeganathan D, Chodhari R, Meeks M *et al.* Loci for primary ciliary dyskinesia map to chromosome 16p12.1-12.2 and 15q13.1-15.1 in Faroe Islands and Israeli Druze genetic isolates. *J Med Genet* 2004; 41:233-40.
18. Geremek M, Zietkiewicz E, Diehl SR *et al*. Linkage analysis localises a Kartagener syndrome gene to a 3.5 cM region on chromosome 15q24-25. *J Med Genet* 2006; 43(1):e1.
19. Karadag B, James AJ, Gultekin E *et al.* Nasal and lower airway level of nitric oxide in children with primary ciliary dyskinesia. *Eur Respir J* 1999; 13:1402-1405.
20. O'Connell JR, Weeks DE. PedCheck: A program for identifying genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; 63:259-266.
21. Sobel E and Lange K. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker sharing statistics. *Am J Hum Genet* 1996; 58:1323-1337.
22. Steve Rozen and Helen J. Skaletsky. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, NJ, 2000, pp 365-386.

Part II

Genomic approach

Chapter 4

Gene expression studies in cells from primary ciliary dyskinesia patients identify 208 potential ciliary genes

Maciej Geremek, Marcel Bruinenberg, Ewa Ziętkiewicz, Andrzej Pogorzelski, Michał Witt and Cisca Wijmenga

Hum Genet 2011;129: 283-93

Abstract

Cilia are small cellular projections that either act as sensors (primary cilia) or propel fluid over the epithelia of various organs (motile cilia). The organellum has gained much attention lately because of its involvement in a group of human diseases called ciliopathies. Primary ciliary dyskinesia (PCD) is an autosomal recessive ciliopathy caused by mutations in cilia motility genes. The disease is characterized by recurrent respiratory tract infections due to the lack of an efficient mucociliary clearance. We performed whole-genome gene expression profiling in bronchial biopsies from PCD patients. We used the quality threshold clustering algorithm to identify groups of genes that revealed highly correlated RNA expression patterns in the biopsies. The largest cluster contained 372 genes and was significantly enriched for genes related to cilia. The database and literature search showed that 164 genes in this cluster were known cilia genes, strongly indicating that the remaining 208 genes were likely to be new cilia genes. The tissue expression pattern of the 208 new cilia genes and the 164 known genes was consistent with the presence of motile cilia in a given tissue. The analysis of the upstream promotor sequences revealed evidence for RFX transcription factors binding site motif in both subgroups. Based on the correlated expression patterns in PCD-affected tissues, we identified 208 genes that we predict to be involved in cilia biology. Our predictions are based directly on the human material and not on model organisms. This list of genes provides candidate genes for PCD and other ciliopathies.

Introduction

Cilia are built on a scaffold of nine peripheral microtubule doublets. In motile cilia, the nine peripheral microtubule doublets are accompanied by two central microtubules (the 9+2 structure). Radial spokes connect the peripheral and central microtubules, while peripheral doublets are connected with each other by nexin

links. Outer and inner dynein arms are anchored to the peripheral microtubule doublet and produce the force needed for the movement of cilia. Synchronized beating of the cilia generates the flow of mucus and cerebrospinal fluid in the respiratory tract and in the brain, respectively, whereas in the fallopian tubes cilia help in moving the ovum toward the uterus. Flagella, built like motile cilia with a similar scheme of 9+2 microtubules, provide motility to unicellular organisms or cells, such as spermatozoa. Primary cilia—which lack the central pair of microtubules (the 9+0 structure), dynein arms and radial spokes—act as sensory organelles, displaying receptor molecules and sensing chemical and mechanical stimuli.[1] However, it is important to note that motile cilia perform a variety of mechanosensory or chemosensory functions and primary cilia located, for example, on the embryonic node are motile.[2]

Nevertheless, our understanding of the molecular composition of cilia is far from complete. Several approaches have been taken to characterize the cilia genome and proteome. Database search for tissue-specific expression pattern limited to ciliated tissues was used to predict 99 cilia related murine genes.[3] Liquid chromatography-mass spectrometry of human epithelial cilia identified 164 axonemal proteins.[4] A mass spectrometry study of flagella of the unicellular alga *Chlamydomonas reinhardtii* identified 360 proteins very likely to be involved in cilia formation, and 292 that are probably involved.[5] In addition, when comparing the genomes of non-ciliated organisms to ciliated organisms predicted 688 and 183 cilia-related genes in *Chlamydomonas reinhardtii*[6] and *Drosophila melanogaster*,[7] respectively. The presence of X-box promoter elements, which are targets of cilia-related transcription factors, has led to the identification of additional ciliary genes in *Caenorhabditis elegans*.[8,9] The results of these different studies have been assembled in online databases: the Ciliome Database[10] and the Cilia Proteome database,[11] which also contains basal bodies proteins. As far as human motile cilia are concerned, the database is probably far from complete since only a single high-throughput proteomics study was performed on mammalian cilia.[4] Most of what we know about cilia comes from work on the unicellular biflagellate alga

Chlamydomonas reinhardtiii. Although the evolutionary conservation of ciliary proteins is remarkably high, it can be assumed that mammalian cilia are more complex. Moreover, the genes related to the presence of multiple cilia in one epithelial cell, the genes associated with the coordinated beating of multiple cilia, as well as the genes coding for receptor molecules responsible for communication with the environment of the human body, are likely to be missing from the database.

Defects in both primary and motile cilia have been associated with a group of pleiotropic, clinically overlapping, human diseases called ciliopathies, such as primary ciliary dyskinesia (PCD), polycystic liver and kidney disease, retinitis pigmentosa, nephronophthisis, Bardet–Biedl syndrome, Alstrom syndrome, and Meckel–Gruber syndrome.[12, 13] PCD is a rare, genetically heterogeneous disorder caused by mutations in genes encoding proteins important for cilia motility.[14-22] The disease is characterized by recurrent respiratory tract infections, bronchiectasis and infertility. Pulmonary symptoms occur due to the lack of an efficient mucociliary clearance caused by kinetic dysfunction of respiratory cilia. Male infertility is caused by immotility of flagella in spermatozoids. *Situs inversus*, a mirror reversal of thoracic organs positioning, is present in approximately half of the PCD patients because of immotility of primary cilia of the embryonic node. Nodal cilia are motile and generate leftward flow of extraembryonic fluid in the nodal pit. The flow has been identified as the initial left–right symmetry breaking event during embryogenesis. According to one of the hypotheses, the flow generates movement of nodal vesicular parcels to the left edge of the node. The parcels contain signaling molecules that activate signaling pathways. In case of immotility of nodal cilia, as in PCD, left–right determination is randomized causing *situs inversus* in approximately half of the patients.[23]

We observed a down-regulation of the expression of dynein genes and other ciliary genes in PCD patients as compared to the controls (unpublished observations). More importantly, we noticed that more genes followed the pattern of dyneins expression in PCD patients, suggesting a functional relationship between the

co-regulated genes. We postulated that gene expression patterns in PCD patients could be used to identify mammalian cilia genes. For this study, we performed gene expression analysis on bronchial biopsies from PCD patients to identify groups of highly correlated cilia-related genes. In contrast to most of previous studies our predictions are based directly on the human material and not on model organisms. We report a cluster of 164 genes previously linked to cilia and 208 new genes that we predict will be involved in cilia-related processes.

Materials and methods

Bronchial biopsies

We collected bronchial biopsies of six clinically diagnosed PCD patients and nine control subjects that were referred to the hospital for unrelated reasons. All patients had a saccharine test and light microscopy imaging of cilia motility characteristic for PCD. An electron microscopy evaluation of ciliary defects was performed, however, in two specimens an insufficient number of epithelial cells was recovered for inspection (Supplementary table 1). The concentration of nitric oxide (NO) in the nasal cavity was measured with a chemiluminescence analyzer, with a threshold value of 200 ppb for diagnosing PCD.[24] Patient #1 had a concentration of nitric oxide in the nasal cavity little below the threshold and no electron microscopy imaging of cilia. Therefore, a second measurement of nitric oxide was performed confirming low concentration. He had a typical course of the disease with a resection of middle lobe due to necrosis and typical for PCD picture in bronchoscopic examination and nasal cavity examination. The specimens from non-PCD controls were obtained through the same protocol; these individuals were referred to the Institute of Tuberculosis and Lung Diseases in Rabka, Poland for regular check-ups, with no symptoms of acute disease, with no bronchoscopic signs of the disturbance of mucociliary transport and with normal ciliary beating in the light microscopy. The study was ethically approved by the institutional review board. Informed consent was obtained from all the subjects.

Gene expression

Anti-sense RNA was synthesized, amplified and purified using the Ambion Illumina TotalPrep Amplification Kit (Ambion, USA) following the manufacturer's protocol. Complementary RNA was hybridized to Illumina HumanRef-12 Whole Genome BeadChips and scanned on the Illumina BeadArray Reader. The gene expression data has been submitted to GEO under accession number GSE11501.

Data preprocessing

The initial steps of data preprocessing and quantile normalization were performed in the Illumina BeadStudio Gene Expression module v3.2. Expression values below 5 were thresholded to 5 and scaled by base 2 logarithm. We limited our data to Ensembl database v52 coding genes. The probes with an expression value variance in the lower 25% of the data were removed. To further limit the computations and to filter out genes not stably expressed in bronchial tissue, we removed probes not detected as present in 4/9 control individuals who had a diagnostic bronchoscopy, but did not display PCD symptoms. The remaining 13,811 probes were subject to clustering.

Clustering

The quality threshold (QT) clustering algorithm was implemented in C according to the algorithm described previously. The QT clustering algorithm performs a computationally extensive search for groups of correlated genes. It looks for largest cluster by iteratively putting every gene in the center of a potential cluster and adding to it genes with the highest Jackknife correlation coefficients in a way that minimizes the cluster diameter *d*, until no further genes can be added without exceeding a predefined *d* value.[25,26] The quality of clusters was ensured by keeping the cluster diameter <0.3.

Database search

For gene annotation enrichment analysis, we used the DAVID (Database for Annotation, Visualization, and Integrated Discovery: http://www.david.abcc.ncifcrf.gov/home.jsp).[27,28] The tool calculates over-representation of specific gene ontology terms with respect to the total number of genes assayed and annotated. DAVID applies a modified Fisher exact test, to establish if the proportion of genes falling into an annotation category significantly differs for a particular group of genes and the background group of genes. Ensembl Gene IDs of the cluster A genes were used as queries and the whole set of genes on the Illumina 12HT chip was used as the background group. The Ciliome Database (http://www.sfu.ca/~leroux/ciliome_home.htm) was queried with Ensembl Gene IDs of cluster A genes and Mouse Ensembl Gene IDs of theirs orthlologues to find known ciliary genes.[10] The Ciliary Proteome Database (http://www.ciliaproteom.com)[11] was queried with UniProt\SwissProt accessions of clusterA genes. Biomart module of the Ensembl database was used for ID\accession conversions. In addition, Pubmed was searched for publications linking individual genes to cilia or flagella.

Tissue-specific expression

The number of EST transcripts for analyzed gene in 30 different tissue types or organ pools was fetched from the UniGene database. Tissue expression enrichment scores were calculated as described in Yu et al.[29,30] In brief, enrichment score $ES_i(g) = o_i(g)/e_i(g)$ is the ratio between observed to expected number of ESTs for gene g in tissue i. The total number of ESTs in UniGene for gene g is $T(g) = \sum_i o_i(g)$. Given the total size of EST libraries in tissue i, s_i, the expected number of ESTs in tissue i for each gene is proportional to $p_i = s_i / \sum_i s_i$. For gene g, if it is expressed equally across all tissues, the expected number of ESTs in tissue i is equal to $e_i = T(g)p_i$. The mean enrichment scores of the analyzed groups of genes in each of the tissues were calculated and plotted for visualization of the tissue-specific expression patterns.

Transcription factor-binding sites

Over-representation of transcription factor binding sites was evaluated in PAINT v3.9[31,32] interfaced with the TRANSFAC Professional v 2009.2 database.[33] The 500 base pairs upstream from transcription starting sites were extracted from Ensembl database v52. A search for transcription factor binding sites was performed by the Match program with matrices deposited in the TRANSFAC database and filter option set to minimize false positives. *p* values were calculated using all human upstream sequences in the PAINT database as control group and false-discovery rate as the multiple testing correction.

Results

Clustering of genes based on the expression profiling of PCD tissue

We performed whole-genome gene expression profiling in bronchial tissue of six PCD patients using Illumina HT-12 bead arrays. From the 48,803 probes on the arrays, 13,811 probes passed the filtering steps and were used for further analysis. We first used the QT (quality threshold) clustering algorithm to identify groups of correlated genes in PCD patients. We identified 12 clusters with more than 100 genes; the clusters ranged in size from 100 (panel L, Figure 1) to 372 genes (panel A, Figure 1). We used DAVID (Database for Annotation, Visualization, and Integrated Discovery)[27,28] to perform functional annotation enrichment analysis on all cluster members. DAVID applies a modified Fisher exact test to establish if the proportion of genes falling into an annotation category significantly differs for a particular group of genes and the background group of genes. The whole set of genes on the Illumina 12HT chip was used as the background group. The terms significantly enriched in cluster A were related to cilia, flagella and microtubules ($p < 0.05$ after Bonferroni's correction) (Table 1). The remaining 11 clusters were not significantly enriched for terms that would have indicated a relation to a specific functional process. We also used DAVID to investigate the tissue expression patterns of the 372 genes in cluster A and found that they are mainly expressed in tissues known to have ciliated epithelium, such as testis, lung and trachea (Table 2).

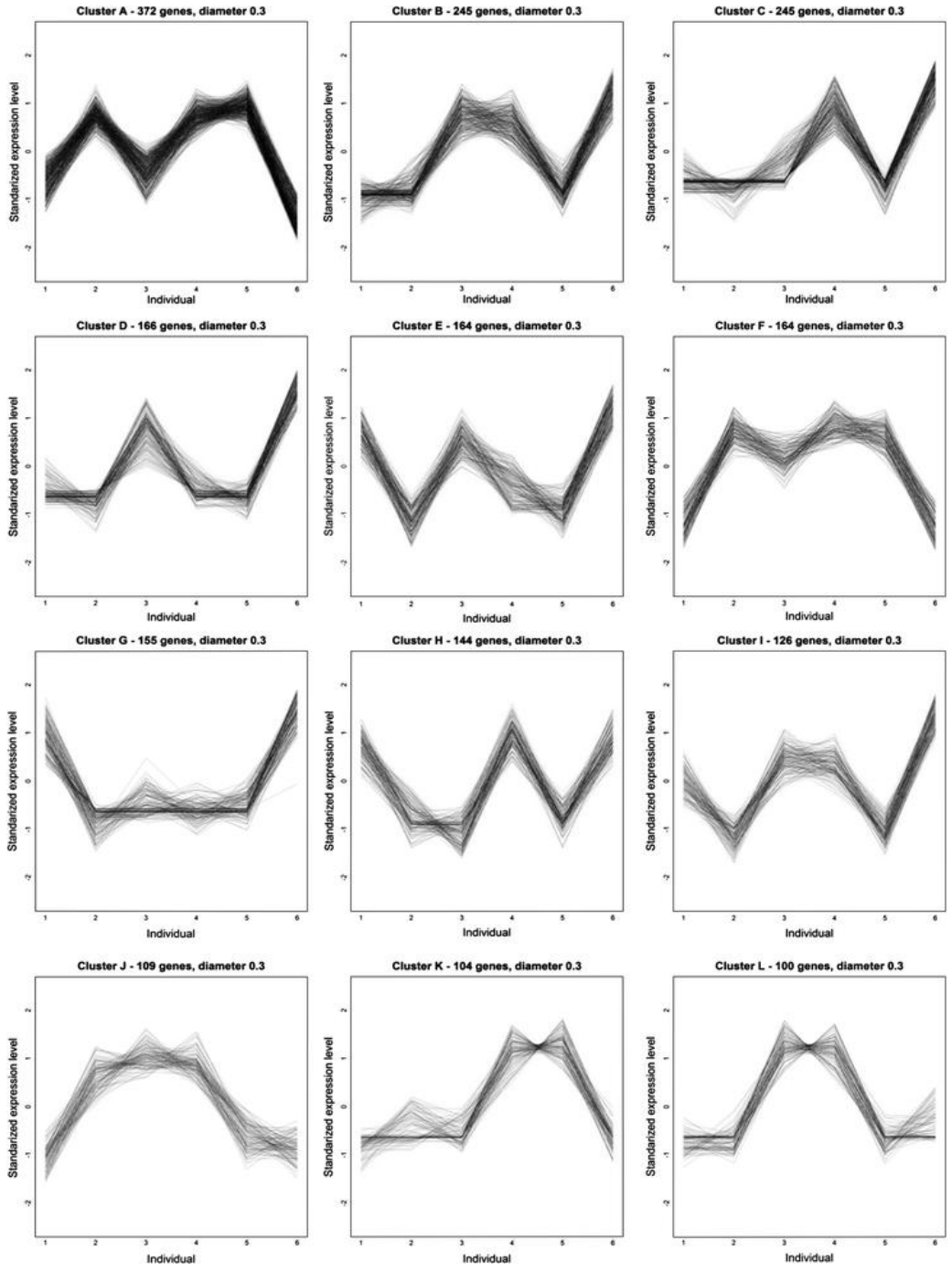

Figure 1. The 12 largest clusters of correlated genes obtained from bronchial tissue of 6 PCD patients. The expression was mean centered and divided by the standard deviation.

Table 1. Distribution of gene ontology (GO) terms in cluster A analyzed by DAVID.

GO term	Number of genes	%	P-Value	Bonferroni corrected p-value
Microtubule-based process	15	3.9	1.10E-06	6.00E-03
Microtubule cytoskeleton	28	7.3	5.50E-12	4.80E-09
Axoneme	9	2.3	2.30E-10	2.00E-07
Microtubule	20	5.2	3.10E-10	2.70E-07
Cilium	10	2.6	6.50E-09	5.60E-06
Dynein complex	9	2.3	2.40E-08	2.10E-05
Cell projection	10	2.6	4.10E-08	3.60E-05
Microtubule associated	13	3.4	8.90E-08	7.70E-05
Cell projection	19	4.9	9.80E-07	8.50E-04
Axonemal dynein	6	1.6	1.10E-06	9.30E-04
Axoneme part	6	1.6	1.50E-06	1.30E-03
Cytoskeletal part	27	7	3.70E-06	3.20E-03
Microtubule motor	11	2.9	4.20E-07	1.20E-03

Characterization of cluster A genes

We queried the publicly available Ciliome Database[10] to check how many members of our cluster A had been previously linked to cilia and found 121 genes to be present in this database. A similar search in the Ciliary Proteome Database[11] resulted in classification of 115 genes as ciliary, 14 of which were not present in the Ciliome Database. A search in PubMed for publications linking cilia or flagella to any of the

individual genes from cluster A led to classification of an additional 29 ciliary genes, resulting in a total of 164 known ciliary genes (i.e. 44% of cluster A; Supplementary table 2). These results strongly suggest that cluster A is a ciliary gene cluster (binominal $p < 1.10^{-8}$ if we assume 10% of human genes to be related to cilia). This would also suggest that the remaining 208 genes are related to cilia and these may represent new cilia genes not previously reported (Supplementary table 3). We observed the highest percentage of shared genes with the part of the Ciliome Database that was built on the results of experimental studies of motile cilia (Table 3). Cluster F (Figure 1) has an expression curve shape very similar to cluster A and 30% of the genes in cluster F also proved to be linked to cilia. However, we limited our analysis to cluster A genes only.

Table 2. Tissue-specific expression enrichment analysis. Presence of a given gene in a tissue 3rd quartile means that 75% of genes have a lower number of transcripts (ESTs) in the given tissue. The count and percentage of genes from cluster A is shown as well as statistical significance of over-representation in cluster A compared to the genomic control.

UNIGENE EST Term	Count	%	P-Value	Bonferroni corrected p-value
Testis 3rd quartile	230	59.9	8.90E-29	6.80E-27
Lung 3rd quartile	202	52.6	9.30E-27	7.10E-25
Uterine tumor 3rd quartile	153	39.8	5.50E-16	4.20E-14
Trachea 3rd quartile	121	31.5	7.90E-13	6.00E-11
Uterus 3rd quartile	135	35.2	1.10E-08	8.60E-07
Brain 3rd quartile	171	44.5	2.20E-08	1.70E-06
Kidney 3rd quartile	146	38	5.00E-08	3.80E-06

To further investigate tissue expression patterns of cluster A genes, we calculated expression enrichment scores for the 30 tissues from the Unigenes dbEST database.[29,30] We performed the calculations both on the full set of cluster A genes and on two subsets of cluster A members: genes previously linked to cilia and the new genes identified in this study as potential ciliary genes. Each of the genes was given an enrichment score based on the ratio between the number of ESTs representing it in a given tissue and the expected number of ESTs for that gene in the tissue. The expected number of ESTs for a given gene was calculated assuming equal expression across all tissues. The mean enrichment scores for all cluster A genes and the subsets were calculated for each tissue to present the tissue-specific expression pattern of the genes (Figure 2).

Table 3. The intersection of cluster A members and the Ciliome database

Type of cilia	Primary (non-motile)				Motile cilia				
Target organism	Drosophila melanogaster		Caenorhabditis elegans		Chlamydomonas reinhardtii			Homo sapiens	Tetrahymena thermophila
Study	Avidor-Reiss [7]	Blacque [8]	Blacque [8]	Efimenko [9]	Li [6]	Pazour [5]	Stolc [19]	Ostrowski [6]	Smith [20]
Method	CG	SAGE	XBOX	XBOX	CG	MS	RNA	MS	MS
Number of genes mapped to human genome	160	681	768	184	309	323	91	128	60
Number of genes in cluster A	31	16	24	11	56	67	38	44	22
% of genes from each study in cluster A	19.3	2.3	3.1	5.9	18.1	20.7	41.7	34.3	36.6

CG comparative genomics; SAGE serial analysis of gene expression; XBOX X-box binding motif; MS mass spectrometry; RNA gene expression during flagella regeneration

The mean enrichment score for cluster A genes and the two subsets was the highest for testis, trachea and lung, which are tissues known to have motile cilia. Brain tissue, where motile cilia are present in the ependyma, kidney and connective tissue having primary cilia, and the eye where rods and cones have a small connecting cilium, were ranked high; 8 out of the 10 highest ranking tissues were the same in the two subgroups of known ciliary genes and genes not previously linked to cilia.

We screened the 500 base pairs (bp) upstream of each gene from cluster A for over-representation of transcription factor binding sites using PAINT v 3.9[31,32] and the TRANSFAC Professional v 2009.2 database[33] (Table 3). The binding site for regulatory factor X (RFX) family of transcription factors was significantly over-represented (false-discovery (FDR) rate adjusted $p < 10^{-6}$) in the upstream sequences of cluster A genes, and in both the known and new ciliary genes in cluster A as well as in the "motile part" of the Ciliome Database (Table 4; Supplementary table 4).

The expression of the cluster A genes in the control samples was stable, but not highly correlated (Figure 3a). The QT clustering in 6 controls returned 13 clusters of more than 100 genes. None of the clusters was significantly enriched in annotation terms related to cilia. The biggest control cluster contained 383 genes (Figure 3b) and was not significantly enriched in any annotation term. Thirteen percent of genes in this cluster have been previously related to cilia, which was not statistically significant. The tissue expression pattern was different from cluster.

A particularly the control cluster genes had much lower expression in testis. A similar pattern could be observed also for other control clusters (Figure 3d). The clustering performed in 6 patients together with 6 controls returned a cluster of 205 genes with similar to cluster A pattern of expression in the patients and 171(83%) of genes present also in cluster A (Figure 3c). Forty-four percent of genes in this cluster have been previously related to cilia.

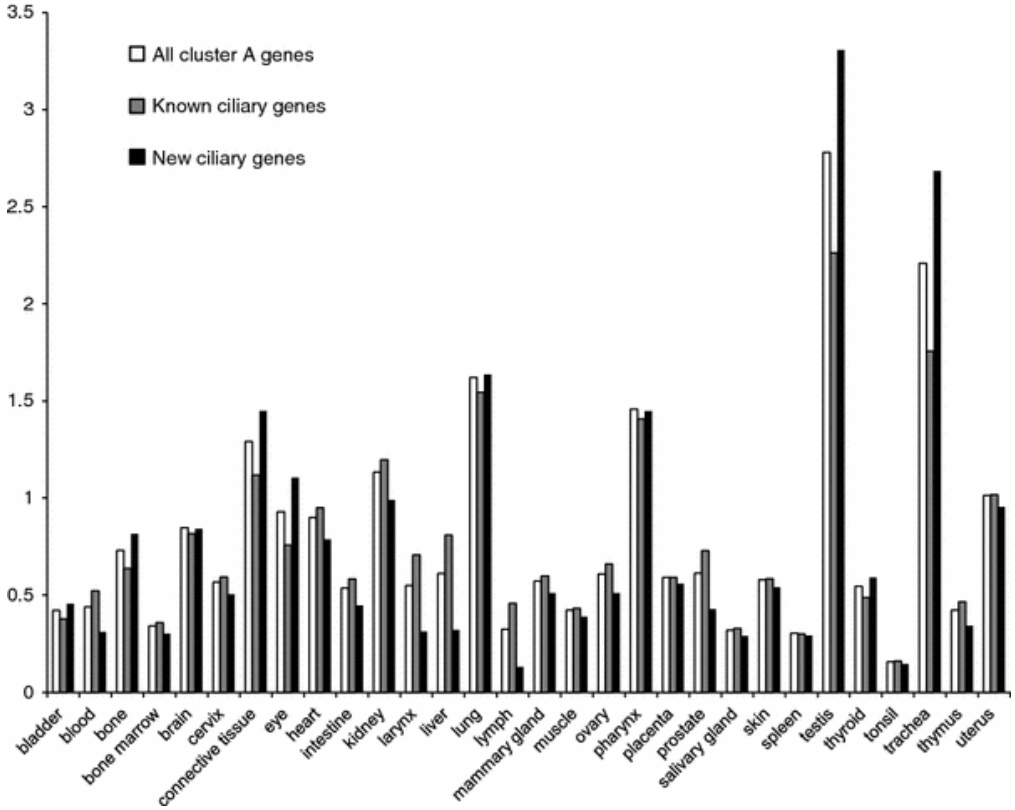

Figure 2. Mean expression enrichment scores of genes from cluster A in 30 different tissues.

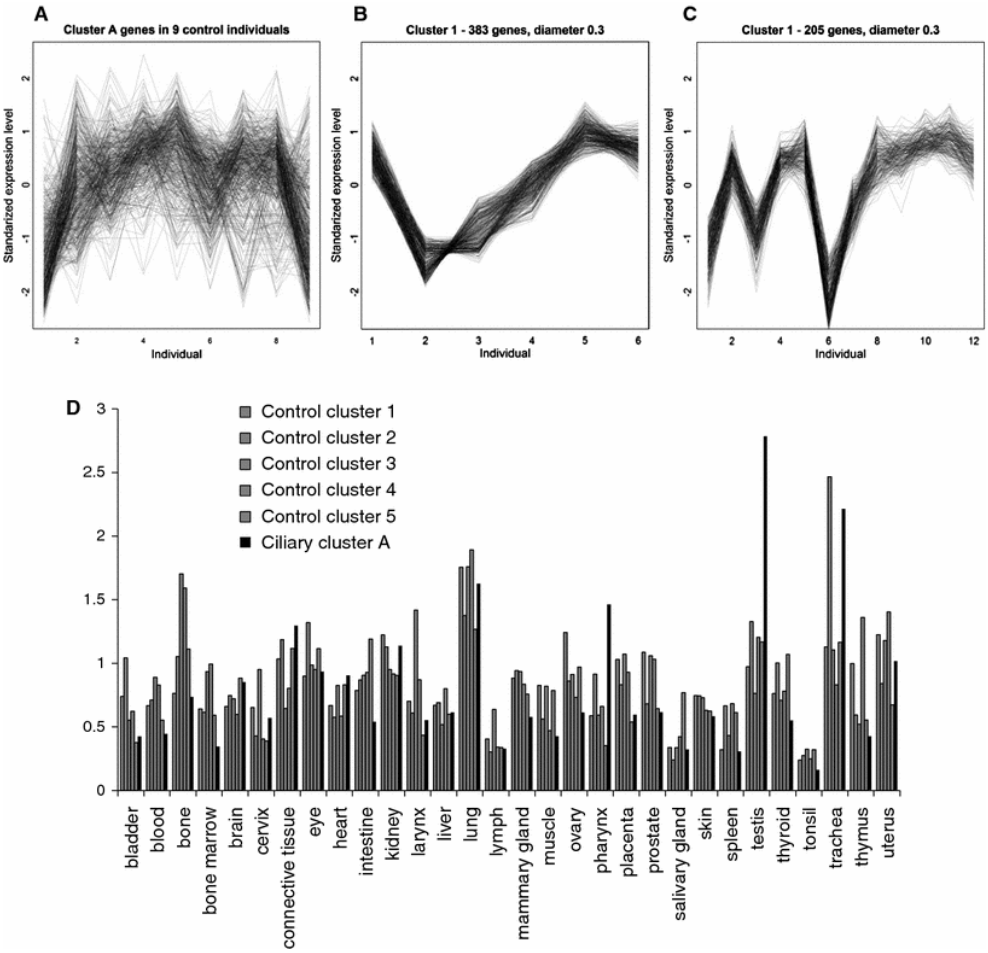

Figure 3. The expression of 372 cluster A genes in 9 non-PCD control individuals (a). The biggest cluster of correlated gene expression in 6 control individuals (b). The biggest cluster in six control individuals and six patients showing a similar curve as cluster A (c). Tissue-specific expression pattern of five biggest control clusters (grey) and the cluster A (black). Cluster A is the only one with high relative expression in testis

Table 4. Over-representation of transcription factors binding sites in the 500 bp upstream sequences of cluster A genes and motile cilia genes from the Ciliome database.

Transcription Factor*	Cluster A (n=372)		New ciliary prediction in cluster A (n=208)		Known ciliary genes in cluster A (n=164)		Motile ciliome (n=648)		Number of sites in the reference group (n=44236)§
	Number of sites	FDR adjusted p-value	Number of sites	FDR adjusted p-value	Number of sites	FDR adjusted p-value	Number of sites	FDR adjusted p-value	
Kid3	336	0.000000	189	0.015281	147	0.025793	572	0.000000	36299
MAZ	81	0.000172	43	0.030229	38	0.011130	113	0.003006	5854
RFX	128	0.000000	70	0.000000	58	0.000000	172	0.000000	8768
RFX1	49	0.000000	23	0.002173	26	0.000000	61	0.000000	1918
Sp1	91	0.000000	45	0.002853	46	0.000000	163	0.000000	5329
v-Myb	107	0.000000	58	0.015938	49	0.005906	204	0.000000	8273
ZF5	118	0.000000	62	0.000000	56	0.000000	216	0.000000	7490
ETF	88	0.000000	45	0.015281	43	0.000318	165	0.000000	5911

*Only the transcription factors that remained significant after multiple testing correction in all groups of genes are shown. §All human promoters sequences in PAINT databases were used as reference set. FDR false-discovery rate.

Discussion

We have identified a group of 208 new genes that are likely to be involved in cilia-related processes. We used gene expression profiling of bronchial tissue from PCD patients to obtain a cluster of 372 genes with highly correlated expression. Forty-four percent of the cluster members have been previously related to cilia based on high-throughput studies or individual experimental studies. The remainder of the cluster members is very likely to be cilia-related genes as they showed tissue-specific expression patterns in accordance with the presence of cilia in the given tissues and they also showed over-presentation of RFX-binding sites in the upstream sequences. The tissue-specific expression pattern consistent with the presence of

cilia was seen for all 372 cluster genes, as well as 164 known ciliary genes and the 208 cluster genes not previously linked to cilia. The expression libraries that we used for analysis of tissue-specific expression pattern were obtained from organs and tissues containing not only ciliated epithelium. Gene expression in different tissues containing ciliated cells has been successfully used to identify cilia-related genes.[3,34] High expression in axoneme-containing tissues, especially in testis, was unique for cluster A and not observed in control clusters. Bronchial biopsies are a compound source of RNA containing not exclusively ciliated epithelial cells.[35] However, the gene expression pattern of the cluster A was extracted from patients and, therefore, is related to a disease that affects the function of ciliated cells.

We found a statistically significant presence of the RFX transcription factors binding sites in the 500-bp upstream sequences of the all cluster A genes, known ciliary genes in the cluster A, and those not previously linked to cilia. In addition, we found RFX-binding sites in the upstream sequences of genes related to motile cilia and deposited in the Ciliome Database. The X-box-binding RFX transcription factors contain a highly homologous DNA-binding domain and share the same DNA-binding sites.[36] RFX3, present among the 372 genes in the cluster A, is known to play a crucial role in the biogenesis of motile cilia.[37]

Based on the sequence homology with *Chlamydomonas reinhardtiii* proteins, several human orthologs have been assigned to outer and inner dynein arms.[38] The largest cluster in our analysis (cluster A) included genes encoding for proteins predicted to be part of outer dynein arms, like *DNAH9*, *DNAI1* or *DNAI2*. Inner dynein arms were represented in this cluster by *DNAH7*, *DNAH3*, *DNAH2,* and *WDR78*. A number of known outer and inner dynein arm genes showed extremely low levels of expression and did not pass the filtering steps (*DNAH14*, *DNAH17*, *TXNDC3* and *DNAH8*). *DNAH6* and *DNAH1* were not classified in our cluster A although they followed very similar expression patterns. *DNAH5* and *DNAH11* followed a different pattern of expression. Cluster A contained two genes known to be involved in building radial spoke complexes (*RSHL3* and *RSPH10B)*,

four intraflagellar transport proteins (*IFT140, IFT172, IFT57, IFT74*) and three genes coding for cytoplasmic dyneins (*DYNC2LI1, DYNLL1, DYNLRB2*), which play a role in the retrograde transport in the flagella in *Chlamydomonas*. Thirty-eight genes from cluster A have been shown to be mutated in human diseases and have entries in the OMIM database, including genes mutated in ciliopathies, such as Bardet–Biedl syndrome (*BBS4, BBS5, BBS11*), Meckel–Gruber syndrome (MKS1), nephronophtisis (*NPHP1, NPHP2, NPHP3*) and PCD (*DNAI1, DNAI2, RSPH4A, RSPH9, LRRC50*). Cluster A also contains two limb–girdle muscular dystrophy genes (*FUKUTIN, TRIM32*), one of which is also mutated in Bardet–Biedl syndrome (*TRIM32*), and three genes associated with retinitis pigmentosa (*PDE6B*, *Pronin* and *RPGR*). *RPGR* is known to cause a complex phenotype with both retinitis pigmentosa and PCD in some patients.[39]

The ciliary proteome has been studied extensively using different high-throughput methods, including mass spectrometry proteomic studies and comparative genomics. All studies, but one[4] were performed in lower organisms and required a Blast search to identify possible human orthologs. Our ciliary set is based on the human transcriptome data, and this is one of our study's strengths. Our dataset, however, might contain false-positive genes that matched the ciliary pattern by chance or are related uniquely to a pathology present in PCD. We based our prediction exclusively on PCD patients' cells, what also limits our dataset to that part of a ciliary puzzle altered by this particular disease process.

PCD is a highly heterogeneous disorder, for which nine genes are known to be mutated. Yet, in the majority of patients, the causative genes remain undiscovered. Our gene set could be considered a list of PCD candidate genes. Because one of the PCD loci is on chromosome 19q[40,41] and we found three genes (*CACNG6, C19ORF51* and *AURKC*) from our cluster that map to this region, they might well be considered disease candidate genes. A 1-bp deletion in the aurora kinase c gene (*AURKC*) has been associated with male infertility characterized by

large-headed, multi-flagellar, polyploid spermatozoa, what makes this gene an interesting candidate for further investigation in PCD patients.[42]

The expression of cluster A genes was not highly correlated in the control samples. We observed a disease-specific pattern of expression for 372 functionally related genes. This phenomenon occurs in response to the mutation of a single gene, but most probably not the same gene in each patient studied, since PCD is a genetically heterogeneous disease. It indicates that information from the cytoplasm of the PCD epithelium, where cilia are misassembled[43], is transferred to the nucleus and transformed on the genomic level into a regulatory signal altering expression of ciliary genes. Our results show how a monogenic disease affecting multi-protein complexes can be used to study more complex genomic networks.

In conclusion, we identified a group of 208 new cilia-related genes. This list of genes provides candidate genes for PCD and other ciliopathies.

Acknowledgements

The cooperation of all the Polish families who participated in this study was invaluable. We thank Jackie Senior for editing the manuscript, Ewa Rutkiewicz for laboratory assistance, and the staff of the Genomics Facility, UMCG, for scientific and technical work. MG was supported by the International PhD Program of Utrecht University, The Netherlands coordinated by the International Institute of Molecular and Cell Biology, Warsaw, Poland.

Supplementary information

https://www.dropbox.com/s/ad90beg26f1n1jb/suppl_data.doc

Supplementary table 1. Electron microscopy, NO measurements and situs status in the PCD patients studied

Supplementary table 2. Known ciliary genes in the cluster A

Supplementary table 3. Cluster A genes previously not linked to cilia

Supplementary table 4. Transcription factors significantly over-represented in the 500bp upstream sequences of analyzed groups of genes. The matrix ID and factor name are from the Transfac database

References

1. Satir P, Christensen ST. Structure and function of mammalian cilia. *Histochem Cell Biol* 2008; 129:687-93.
2. Bloodgood RA. Sensory reception is an attribute of both primary cilia and motile cilia. J *Cell Sci* 2010; 123:505-9.
3. McClintock TS, Glasser CE, Bose SC, Bergman DA. Tissue expression patterns identify mouse cilia genes. *Physiol Genomics* 2008; 32:198-206.
4. Ostrowski LE, Blackburn K, Radde KM *et al*. A proteomic analysis of human cilia: identification of novel components. *Mol Cell Proteomics* 2002; 1:451-65.
5. Pazour GJ, Agrin N, Leszyk J, Witman GB. Proteomic analysis of a eukaryotic cilium. J *Cell Biol* 2010; 170:103-13.
6. Li JB, Gerdes JM, Haycraft CJ *et al*. Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* 2004; 117:541-52.
7. Avidor-Reiss T, Maer AM, Koundakjian E *et al*. Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell* 2004; 117:527-39.
8. Blacque OE, Perens EA, Boroevich KA *et al*. Functional genomics of the cilium, a sensory organelle. *Curr Biol* 2005; 15:935-41.

9. Efimenko E, Bubb K, Mak HY *et al*. Analysis of xbx genes in *C. elegans*. *Development* 2005; 132:1923-34.
10. Inglis PN, Boroevich KA, Leroux MR. Piecing together a ciliome. *Trends Genet* 2006; 22:491-500.
11. Gherman A, Davis EE, Katsanis N. The ciliary proteome database: an integrated community resource for the genetic and functional dissection of cilia. *Nat Genet* 2006; 38:961-2.
12. Badano JL, Mitsuma N, Beales PL, Katsanis N. The ciliopathies: an emerging class of human genetic disorders. *Annu Rev Genomics Hum Genet* 2006; 7:125-48.
13. Marshall WF. The cell biological basis of ciliary disease. *J Cell Biol* 2008; 180:17-21.
14. Geremek M, Witt M. Primary ciliary dyskinesia: genes, candidate genes and chromosomal regions. *J Appl Genet* 2004; 45:347-61.
15. Pennarun G, Escudier E, Chapelin C *et al*. Loss-of-function mutations in a human gene related to Chlamydomonas reinhardtii dynein IC78 result in primary ciliary dyskinesia. *Am J Hum Genet* 1999; 65:1508-1519.
16. Olbrich H, Haffner K, Kispert A *et al*. Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. *Nature Genet* 2002; 30:143-144.
17. Bartoloni L, Blouin JL, Pan Y *et al*. Mutations in the DNAH11 (axonemal heavy chain dynein type 11) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia. *Proc Natl Acad Sci* 2002; 99:10282–10286.
18. Loges NT, Olbrich H, Fenske L *et al*. DNAI2 mutations cause primary ciliary dyskinesia with defects in the outer dynein arm. *Am J Hum Genet* 2008; 83:547-558.
19. Omran H, Kobayashi D, Olbrich H *et al*. Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. *Nature* 2008; 456:611-616.

20. Castleman V H, Romio L, Chodhari R *et al*. Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with central-microtubular-pair abnormalities. Am J Hum Genet 2009; 84:197-209.
21. Duquesnoy P, Escudier E, Vincensini L *et al*. Loss-of-function mutations in the human ortholog of Chlamydomonas reinhardtii ODA7 disrupt dynein arm assembly and cause primary ciliary dyskinesia. *Am J Hum Genet* 2009; 85:890-896.
22. Loges NT, Olbrich H, Becker-Heck A *et al*. Deletions and point mutations of LRRC50 cause primary ciliary dyskinesia due to dynein arm defects. *Am J Hum Genet* 2009; 85:883-889.
23. Nobutaka H, Yosuke T, Yasushi O. Left–Right Determination: Involvement of Molecular Motor KIF3, Cilia, and Nodal Flow. *Cold Spring Harb Perspect Biol*. 2009; 1:a000802.
24. Karadag B, James AJ, Gultekin E. Nasal and lower airway level of nitric oxide in children with primary ciliary dyskinesia. *Eur Respir J* 1999; 13:1402-1405.
25. Heyer LJ, Kruglyak S, Yooseph S. Exploring expression data: identification and analysis of coexpressed genes. *Genome Res* 1999; 9:1106-15.
26. Coppe A, Ferrari F, Bisognin A *et al*. Motif discovery in promoters of genes co-localized and co-expressed during myeloid cells differentiation. *Nucleic Acids Res* 2009; 37:533-49.
27. Huang da W, Sherman BT, Tan Q *et al*. (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 2007; 8:R183.
28. Kouwenhoven EN, van Heeringen SJ, Tena JJ *et al*. Genome-wide profiling of p63 DNA-binding sites identifies an element that regulates gene expression during limb development in the 7q21 SHFM1 locus. *PLoS Genet* 2010; 6:e1001065.

29. Yu X, Lin J, Zack DJ, Qian J (2006) Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. *Nucleic Acids Res* 2006; 34:4925-36.
30. Colecchia F, Kottwitz D, Wagner M *et al*. Tissue-specific regulatory network extractor (TS-REX): a database and software resource for the tissue and cell type-specific investigation of transcription factor-gene networks. *Nucleic Acids Res* 2009; 37:e82.
31. Vadigepalli R, Chakravarthula P, Zak DE *et al*. PAINT: a promoter analysis and interaction network generation tool for gene regulatory network identification. *OMICS* 2003; 7:235-52.
32. Riehle KJ, Campbell JS, McMahan RS *et al*. Regulation of liver regeneration and hepatocarcinogenesis by suppressor of cytokine signaling. *J Exp Med* 2008; 205:91-103.
33. Wingender E, Dietze P, Karas H, Knüppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* 1996; 24:238-241.
34. Kubo A, Yuba-Kubo A, Tsukita S *et al*. Sentan: a novel specific component of the apical structure of vertebrate motile cilia. *Mol Biol Cell* 2008; 19:5338-46.
35. Regamey N, Hilliard TN, Saglani S *et al*. Quality, size, and composition of pediatric endobronchial biopsies in cystic fibrosis. *Chest* 2007; 131:1710-7.
36. Iwama A, Pan J, Zhang P *et al*. Dimeric RFX proteins contribute to the activity and lineage specificity of the interleukin-5 receptor alpha promoter through activation and repression domains. *Mol Cell Biol* 1999; 19:3940-50.
37. El Zein L, Ait-Lounis A, Morlé L *et al*. RFX3 governs growth and beating efficiency of motile cilia in mouse and controls the expression of genes involved in human ciliopathies. *J Cell Sci* 2009; 122:3180-9.
38. Pazour GJ, Agrin N, Walker BL, Witman GB. Identification of predicted human outer dynein arm genes: candidates for primary ciliary dyskinesia genes. *J Med Genet* 2005; 43:62-73.

39. Moore A, Escudier E, Roger G *et al*. RPGR is mutated in patients with a complex X linked phenotype combining primary ciliary dyskinesia and retinitis pigmentosa. *J Med Genet* 2006; 43:326-33
40. Meeks M, Walne A, Spiden S *et al*. A locus for primary ciliary dyskinesia maps to chromosome 19q. *J Med Genet* 2000; 37:241-4.
41. Blouin JL, Meeks M, Radhakrishna U *et al*. Primary ciliary dyskinesia: a genome-wide linkage analysis reveals extensive locus heterogeneity. *Eur J Hum Genet* 2000; 8:109-18.
42. Dieterich K, Soto Rifo R, Faure AK *et al*. Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. *Nat Genet* 2007; 39:661-5.
43. Fliegauf M, Olbrich H, Horvath J *et al*. Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. *Am J Respir Crit Care Med* 2005; 171:1343-9.

Chapter 5

Ciliary genes are down-regulated in bronchial tissue of primary ciliary dyskinesia patients

Maciej Geremek, Marcel Bruinenberg, Lude Franke, Ewa Ziętkiewicz, Andrzej Pogorzelski, Cisca Wijmenga, Michał Witt

Submitted for publication

Abstract

Primary ciliary dyskinesia (PCD) is a rare, genetically heterogeneous disease characterized by recurrent respiratory tract infections, sinusitis, bronchiectasis and male infertility. The pulmonary phenotype in PCD is caused by the impaired motility of cilia in the respiratory epithelium, due to ultrastructural defects of these organelles. We hypothesized that defects of multi-protein ciliary complexes should be reflected by gene expression changes in the respiratory epithelium. We have previously found that large group of genes functionally related to cilia share highly correlated expression pattern in PCD bronchial tissue. Here we performed an explorative analysis of differential gene expression on the bronchial tissue from six PCD patients and nine non-PCD controls, using Illumina HumanRef-12 Whole Genome BeadChips. We observed 1,323 genes with at least two-fold difference in the mean expression level between the two groups (t-test p-value <0.05). Annotation analysis showed that the genes down-regulated in PCD biopsies (601) were significantly enriched for terms related to cilia, whereas the up-regulated genes (722) were significantly enriched for terms related to mitosis and ribosomes. We assembled a list of human genes predicted to encode orthologs of flagellar proteins from *Chlamydomonas reinhardtii*: outer dynein arms, inner dynein arms, radial spokes, and intraflagellar transport proteins. A significant down-regulation of the expression of genes from all the four groups was observed in PCD, compared to non-PCD biopsies. Our data suggest that the down-regulation of the ciliome genes plays an important role in the molecular pathomechanism of PCD.

Introduction

Cilia are small cellular projectiles, extending from the cell surface, with which they share the cell membrane. The ciliary axoneme is built on a scaffold of nine peripheral microtubule doublets, associated with many multi-protein complexes, including outer and inner dynein arms (ODA and IDA), nexin links and radial spokes. Cilia act either as sensors (primary cilia) or propel fluid over the epithelia of various organs (motile cilia)[1]; their dysfunction is the underlying cause of many systemic diseases.

Primary ciliary dyskinesia (PCD) is a rare genetic disease characterized by recurrent respiratory tract infections, bronchiectasis and infertility. Pulmonary symptoms occur because of the lack of an efficient mucociliary clearance, caused by the kinetic dysfunction of the respiratory cilia. Male infertility is caused by the dysmotility of flagella in spermatozoids. *Situs inversus*, a mirror reversal of body organ positioning, is present in approximately half of the PCD patients because of the immotility of primary cilia of the embryonic node.[2,3] The diagnosis of PCD is based on the clinical criteria, the delayed saccharin test,[4] and low levels of nasal nitric oxide.[5] These symptoms are supported by the imaging studies of the respiratory cilia. Tissue specimens obtained through bronchoscopy are evaluated in the light microscope for ciliary beating and with the electron microscope for defects of the ciliary ultrastructure. Among a wide spectrum of various ultrastructural ciliary lesions documented in PCD patients[6], the lack of ODA and/or IDA is the most frequently observed defect.

PCD is an autosomal recessive trait with extensive genetic heterogeneity. To date, eleven genes have been found to be mutated in PCD: *DNAH5*,[7] *DNAH11*,[8] *DNAI1*,[9] *DNAI2*,[10] *RSPH9*,[11] *RSPH4A*,[11] *TXNDC3*,[12] *Ktu/PF13*,[13] *LRRC50*,[14,15] *CCDC39*[16] and *CCDC40*[17]. Products of most of these genes have conserved orthologs in *Chlamydomonas reinhardtii* and have a well-defined ultrastructural

localization in the axoneme. Collectively, all these genes explain only about one third of PCD cases.

In search for a better understanding of the human cilia genome and proteome, we have performed gene expression profiling in PCD bronchial biopsies. We have previously found that large group of ciliary genes have highly correlated gene expression in 6 PCD bronchial biopsies. We applied cluster analysis to extract the pattern of expression of these ciliary genes and used that pattern to identify new genes involved in human cilia biology. Herein we report an explorative study of differential gene expression in PCD. We performed gene expression profiling in bronchial biopsies from the six PCD patients and nine non-PCD controls. We report that the significant proportion of the genes that are down-regulated in PCD encode elements of the axonemal ultrastructure.

Material and methods

Bronchial biopsies and subjects

Material for this study, obtained from six unrelated PCD patients and nine unrelated control individuals, has been described in detail in Geremek et al.[18] Briefly, clinical evaluation of the PCD patients was performed at the Institute of Tuberculosis and Lung Diseases in Rabka, Poland, by the experienced physician (AP). The primary bronchopulmonary symptoms in the patients were: sinusitis, nasal polyps, bronchiectasis, and recurrent infections of the upper respiratory tract. PCD status was also indicated by the positive results of routine diagnostic tests: the delayed saccharine test and lack of the ciliary motility in light microscopy imaging. In three patients (#1, 3, 4, 5) PCD diagnosis was confirmed by the low level of NO in the nasal cavity (chemiluminescence analyzer, the threshold value of 200 ppb)[5] (Supplementary table 1). The second patient (#2) had ODA defect and the sixth patient (#6) had IDA defect, but no NO data were available for both individuals. The

mutational status was known for one patient (patient #4), who was a compound heterozygote for two mutations in *DNAH5*. Non-PCD controls were individuals referred to the Institute of Tuberculosis and Lung Diseases in Rabka for different reasons unrelated to PCD; they presented no symptoms of acute pulmonary disease, no bronchoscopic signs of the disturbance of mucociliary transport, and had normal ciliary beating in the light microscopy.

Biopsies were obtained during a diagnostic bronchoscopy, through the same protocol in PCD and non-PCD individuals. Specimens were stored in RNAlater buffer and used for RNA isolation. The informed consent was obtained from the participating patients. The study was approved by the appropriate local Ethics Committees.

Gene expression analysis

The quality and concentration of RNA were determined by using the 2100 Bioanalyzer (Agilent, Amstelveen, The Netherlands) with the Agilent RNA 6000 Nano Kit. Anti-sense RNA was synthesized, amplified and purified using the Ambion Illumina TotalPrep Amplification Kit (Ambion, USA) following the manufacturer's protocol. Complementary RNA was hybridized to Illumina HumanRef-12 Whole Genome BeadChips and scanned on the Illumina BeadArray Reader. Data was handled through the Illumina BeadStudio Gene Expression module v3.2.[19]

Data-analysis

Initial steps of data preprocessing, quantile normalization, background subtraction and quality control were performed using BeadStudio software. The data were exported to GeneSpring 9. Expression values with a threshold value below 5 were raised to 5, and the data were scaled by a base 2 logarithm.

Fold-change analysis for all the probes was performed using GeneSpring 9 software. Genes that had at least two-fold expression change (with an uncorrected t-

test p-value<0.05) were subject to gene enrichment annotation analysis using DAVID (Database for Annotation, Visualization, and Integrated Discovery).[21] DAVID applies a modified Fisher exact test, to establish if the proportion of genes falling into an annotation category differs for a particular group of genes and the background group of genes (in this case, the whole set of genes on the Illumina 12HT chip); the p-values were corrected for multiple testing using the FDR method. The resulting annotation terms were then clustered into functional groups.

For the search of human orthologues of *Chlamydomonas reinhardtii* flagellar genes, the data were processed using the R package. The Ciliome database (www.ciliome.com)[22] and PubMed were used as a reference for known ciliary genes.

The mean expression values in the patients and controls were calculated, and a Wilcoxon Mann-Whitney test was used to obtain individual p-values. The unweighted Z-method of combining probabilities from independent tests was used to assess if the distribution of p-values deviated from normal.[23] The Z-transform test converts one-tailed p-values, from each of the independent tests, into standard normal deviates; the sum of these standard normal deviates, divided by the square root of the number of tests, has a standard normal distribution if the common null hypothesis is true.

Results

Whole genome analysis (fold change and gene enrichment)

We identified 1,323 genes (1,396 probes), which displayed at least 2-fold expression level difference in PCD compared to non-PCD bronchial tissue (t-test p-value <0.05); 722 of these genes (755 probes) were over-expressed, whereas 601 genes (641 probes) showed a reduced expression in PCD (Supplementary table 2). To investigate the biological meaning of these differentially expressed genes, we performed annotation enrichment analysis, separately for the up-regulated and

down-regulated genes, using DAVID software.[21] Annotations of genes with the reduced expression in PCD patients were significantly enriched for terms related to microtubules, motility and cilia (Table 1). In the group of up-regulated genes, four high-scoring annotation clusters were found: first related to mitotic spindle, cell cycle and mitosis, second related to ribosomes, ribonucleoprotein and translation, third related to cellular organelles, and last related to ribosomes (Table 2). To investigate how many of the down- and up-regulated genes were functionally linked to cilia, the two groups (genes from Tables 1 and Table 2) were used to query the Ciliome database[20] and PubMed. This search revealed that 105 (22%) of the down-regulated and 46 (12%) of the up-regulated genes were related to cilia. Under a theoretical assumption that approximately 10% of the human genes are related to cilia (the Ciliome database contains 2024 entries), the proportion of ciliary genes found in the down-regulated group was significantly higher than in the whole set of genes on the Illumina 12HT chip (binominal p-value 0.0003). The list of up-regulated genes was not significantly enriched for genes related to cilia (p-value 0.296967).

Table 1. Gene annotation analysis of the 601 genes down-regulated in PCD (two-fold change, and $p<0.05$)

Functional Group 1 Geometric p-value: 2.9868321894195106E-6				
GO Category	Term	Number of genes with the annotation term	P-value	FDR*
cellular component	axoneme part	10	1.64E-12	2.54E-09
cellular component	Axoneme	11	4.47E-12	6.93E-09
cellular component	dynein complex	13	7.23E-12	1.12E-08
cellular component	microtubule cytoskeleton	35	3.66E-11	5.67E-08

cellular component	axonemal dynein complex	9	4.82E-11	7.47E-08
cellular component	cell projection part	14	1.56E-10	2.42E-07
cellular component	microtubule motor activity	15	5.00E-09	8.96E-06
cellular component	microtubule associated complex	17	8.30E-09	1.29E-05
cellular component	cell projection	29	1.24E-08	1.92E-05
biological process	microtubule-based movement	16	2.42E-08	4.63E-05
biological process	microtubule-based process	22	2.99E-08	5.72E-05
biological process	cytoskeleton-dependent intracellular transport	17	4.07E-08	7.79E-05
molecular function	motor activity	18	1.60E-07	2.87E-04
cellular component	Microtubule	21	3.06E-07	4.75E-04
biological process	ciliary or flagellar motility	6	1.08E-06	0.002070472
cellular component	Cytoskeleton	50	7.87E-06	0.012198098

*FDR false discovery rate.

Table 2. Gene annotation analysis of the-722 genes up-regulated in PCD (two-fold change and $p<0.05$)

Functional Group 1 Geometric p-value: 5.19460677949439E-8				
GO category	Term	Number of genes with the annotation term	P-value	FDR*
biological process	mitotic cell cycle	28	1.04E-09	1.98E-06
biological process	cell cycle	49	2.15E-09	4.12E-06
biological process	M phase	26	2.48E-09	4.75E-06

biological process	cell cycle process	44	3.03E-09	5.80E-06
biological process	cell cycle phase	28	1.07E-08	2.04E-05
biological process	Mitosis	22	1.88E-08	3.59E-05
biological process	M phase of mitotic cell cycle	22	2.20E-08	4.21E-05
biological process	cell division	22	5.10E-08	9.75E-05

Functional Group 2

Geometric p-value: 7.2875016098167925E-6

GO category	Term	Number of genes with the annotation term	P-value	FDR
cellular component	intracellular non-membrane-bound organelle	78	8.84E-10	1.37E-06
cellular component	non-membrane-bound organelle	78	8.84E-10	1.37E-06

Functional Group 3

Geometric p-value: 6.297400435535407E-5

GO category	Term	Number of genes with the annotation term	P-value	FDR
molecular function	structural constituent of ribosome	24	1.21E-09	2.16E-06
cellular component	Ribosome	24	7.87E-09	1.22E-05
cellular	ribonucleoprotein	31	4.63E-07	7.17E-04

component	complex			
cellular component	ribosomal subunit	14	3.29E-06	0.005103915
cellular component	cytosolic ribosome (sensu Eukaryota)	11	4.55E-06	0.007050104
molecular function	structural molecule activity	37	4.99E-06	0.008947295
cellular component	large ribosomal subunit	10	1.46E-05	0.022608839

Functional Group 4

Geometric p-value: 1.291005170513454E-4

GO category	Term	Number of genes with the annotation term	P-value	FDR
biological process	DNA replication	19	6.98E-07	0.001335559
biological process	response to DNA damage stimulus	20	4.90E-05	0.043594568

*FDR false discovery rate

Known ciliary genes

To search for evidence of differential expression of genes related to specific axonemal structure elements, we assembled a list of proteins predicted to be a part of the known ultrastructural ciliary components. For this purpose, we searched the literature for data linking individual proteins with particular ultrastructural components of flagella of *Chlamydomonas reinhardtii*. The selected proteins from the alga were blasted against the human proteome, and the best scoring human

proteins, with the significance score higher than the arbitrary threshold of 1e-15, were assigned as orthologues. We also used data from Pazour et al.[24], who applied phylogenetic analysis to assign human heavy dynein chains to ultrastructural components of flagella.

The list of 47 human genes encoding proteins predicted to be a part of the known ultrastructural ciliary components included: 13 ODA genes, 14 IDA genes, 8 radial spoke genes and 12 intraflagellar transport genes (Table 3). These 47 genes were represented by 61 probes on the Illumina 12HT chip used here to characterize the level of expression in bronchial tissues from PCD patients and non-PCD controls. The mean expression levels in PCD and non-PCD bronchial tissues, and the corresponding Wilcoxon Mann-Whitney test p-values are presented in Table 3. Only 3 of the 61 probes showed the higher expression level in the PCD group. The remaining 58 probes, representing 45 ciliary genes, indicated the lower expression level in PCD tissue. Significant differences between the levels of down-regulation in PCD and non-PCD groups were also observed when the p-values were calculated, using an unweighted Z-method,[23] in four subgroups of genes encoding ultrastructural ciliary components: ODA, IDA, radial spoke and intraflagellar transport proteins (Table 3).

Table 3. Differential gene expression of genes encoding outer dynein arms, inner dynein arms, radial spokes and intraflagellar transport proteins in the group of PCD cases compared to the non-PCD controls

Gene	Probe	PCD (Log2 of the expression)	Non-PCD (Log2 of the expression)	Direction	Wilcoxon Mann-Whitney p-value	Unweighted Z-method combined p-value
Outer dynein arms						**5.94E-07**
DNAH9	6400255	7.687516	9.826032	Down	0.002398	

DNAH5	2350554	2.840433	5.349519	Down	0.010312	
DNAI1	2810300	8.674844	9.844953	Down	0.033167	
DNAI2	5960685	7.177193	8.689865	Down	0.002398	
DNAL1	2810500	6.708057	6.29279	Up	0.909491	
DNAL1	6100682	3.770269	5.508203	Down	0.005994	
TCTEX1D1	7560300	7.027727	8.381239	Down	0.043956	
CALM3	2750184	10.55856	10.60304	Down	0.431968	
DYNLRB2	5340725	10.71992	11.83949	Down	0.136064	
DNAL4	270678	10.26707	10.75581	Down	0.172674	
DYNLL1	6220086	12.17424	12.25021	Down	0.344456	
DYNLL2	3400551	6.986979	6.429329	Up	0.992199	
CCDC114	7210482	6.548517	8.35885	Down	0.005994	
DNAH11	7330360	3.002923	5.663427	Down	0.002598	
Inner dynein arms						**2.59E-13**
DNAH7	6100661	6.514025	8.269907	Down	0.003796	
DNAH1	4900612	7.099972	8.823093	Down	0.002398	
DNHD2	1710253	8.860691	10.45194	Down	0.017982	
DNAH10	2000377	2.428852	3.237659	Down	0.014854	
DNAH3	2680059	5.769333	8.094914	Down	0.005605	
DNAH2	6860689	8.60512	10.73228	Down	0.001399	
DNHL1	7150161	3.006248	3.17991	Down	0.381969	
WDR63	5560730	6.41027	8.011633	Down	0.017982	
TCTEX1D2	1580021	8.612464	8.696081	Down	0.568032	
TCTEX1D2	7150433	11.46435	11.89049	Down	0.431968	
DNALI1	6200309	9.602157	11.071	Down	0.002398	

DYNLT1	6350634	11.13857	11.51753	Down	0.136064	
CETN2	2120707	12.38314	13.28572	Down	0.08747	
WDR78	2360386	6.378204	8.079985	Down	0.012787	
WDR78	2650754	3.57718	5.360144	Down	0.019236	
WDR78	3190731	4.836388	5.143086	Down	0.612188	
DNAH11	7330360	3.002923	5.663427	Down	0.002598	
Radial spokes						**4.16E-06**
RSPH3	2650678	7.413316	7.836832	Down	0.227972	
RSHL3	2630519	9.918507	10.80253	Down	0.163836	
RSHL3	4850241	9.109246	9.932796	Down	0.111888	
NME5	5290463	9.938878	10.88896	Down	0.264336	
RSPH1	6020451	10.40699	11.50298	Down	0.072328	
C6ORF206	840433	8.570299	9.296023	Down	0.163836	
RSPH10B	4060360	6.716065	8.128109	Down	0.02956	
RSPH10B	5810324	7.983313	9.466374	Down	0.008791	
PPIL6	3120259	6.766707	8.011536	Down	0.111888	
DNAJB13	1780309	4.103136	5.749771	Down	0.033167	
Intraflagellar transport complex						**1.69E-10**
DYNC1H1	3450600	8.218894	8.928265	Down	0.033167	
IFT57	5700224	8.122991	9.303622	Down	0.033167	
IFT57	5700224	8.122991	9.303622	Down	0.033167	
IFT52	520367	6.271189	6.359964	Down	0.523515	
IFT52	520367	6.271189	6.359964	Down	0.523515	
IFT52	3420270	6.503832	6.703533	Down	0.136064	
IFT81	1090491	4.485731	5.681726	Down	0.072328	

IFT81	7040706	3.850651	4.345121	Down	0.157369
IFT88	3610537	5.57095	5.331608	Up	0.615962
IFT88	4490253	2.321928	4.275251	Down	0.001754
IFT88	5670288	7.041648	8.121763	Down	0.012787
IFT122	1570687	6.441277	7.374693	Down	0.012787
IFT122	4210463	7.780974	8.710176	Down	0.005994
IFT122	5390554	3.13301	3.147856	Down	0.5
IFT140	6960376	6.501864	7.674088	Down	0.017982
DYNC2LI1	5720609	7.462295	7.494275	Down	0.344456
DYNC2LI1	6520543	6.808812	7.84292	Down	0.024775
IFT172	130070	8.232366	9.177806	Down	0.017982
IFT57	5700224	8.122991	9.303622	Down	0.033167
KIFAP3	1940482	8.670183	9.185607	Down	0.072328

Discussion

The gene expression profiling using the whole-genome Illumina panel, performed in bronchial biopsies from six PCD patients and nine non-PCD controls, revealed 1,396 probes (representing 1,323 genes) with at least two-fold expression level difference between the PCD and non-PCD groups (t-test p-value < 0.05). A combination of the two-fold change with the significance measured by an uncorrected t-test, which takes into consideration the variance between the samples, has been shown to be a reliable and reproducible measure of differential expression.[25]

Functional annotation of the differentially expressed genes revealed that the down-regulated genes were related to microtubules, motility and cilia, and the up-regulated genes to translation, ribosomes, cell cycle and mitosis.

To further characterize the genes expression changes in the PCD, we compared expression levels of a number of cilia-related genes in the PCD and non-PCD tissue. We assembled the list of 47 human orthologues of the *Chlamydomonas reinhardtii* flagellar proteins. The list was divided into three subgroups related to the ultrastructural components of motile cilia (outer dynein arms, inner dynein arms and radial spokes), plus one functional subgroup (intraflagellar transport proteins). In the analysis of bronchial biopsies from PCD patients, we demonstrated the reduced expression of the genes in all the four groups. Given the lack of linear relation between RNA and protein, we cannot make more specific conclusions about the possible correlation of the RNA level and the ultrastructural ciliary defect. Large multi-protein complexes of the dynein arms are pre-assembled in the cytoplasm, before their intraflagellar transport to the ciliary axonemes. Immunohisto-chemical staining of the ciliated epithelium from PCD patients with the identified *DNAH5* mutations has shown that not only the mutated DNAH5, but also DNAH9 protein was missing from the axonemes, indicating that in patients with mutations in one of the ciliary genes, the delivery of other, non-mutated members of the ODA complexes to the cilium is impaired.[26] Our data suggest that a regulatory system may exist, which – in the presence of a mutated ciliary gene – down-regulates the expression of the entire ciliary genome. Such a regulation would play an important role in the molecular pathomechanism of PCD.

Of the genes known to be mutated in PCD[7-17], six (*DNAH5*, *DNAH11*, *DNAI1*, *DNAI2*, *RSPH4A*, *LRCC50*) were down-regulated in the PCD patients examined in our study. *RSPH9* and *Ktu/PF13* were equally expressed in the PCD and non-PCD samples, *TXNDC3* and *CCDC40* were not detected on the Illumina array and *CCDC39* was not present on the array. The disease genes in several genomic regions reported to be linked to PCD (for the review see ref. 27), remain

not identified, mainly because of the noise in the linkage data, and because of difficulties in the functional prioritization of the candidate genes. In this context, observation of a down-regulation of the particular gene's expression in PCD might provide clues for selecting candidate genes from the linkage peaks.

To summarize, in our explorative study of gene expression in bronchial biopsies from PCD patients we found evidence of a significant down-regulation of the large group of cilia-related genes, predicted to encode elements of outer dynein arms, inner dynein arms, radial spokes, and intraflagellar transport proteins. These data suggest the existence of a regulatory system, capable of down-regulating the expression of the ciliome, which would play an important role in the molecular pathomechanism of PCD.

Acknowledgements

The cooperation of all the Polish families who participated in this study was invaluable. We thank Jackie Senior for editing the manuscript and Ewa Rutkiewicz for laboratory assistance. We thank the staff of the Genomics Facility, UMCG, for scientific and technical work. MG was supported by the International PhD Program of the University of Utrecht. This work was covered in part by the grant NN401 095537 from the Ministry of Science and Higher Education, Poland.

Supplementary information

https://www.dropbox.com/s/ad90beg26f1n1jb/suppl_data.doc

Supplementary table 1. Electron microscopy, NO measurements and situs status in the PCD patients studied

Supplementary table 2. Differentially expressed (> two-fold change $p<0.05$) genes used for annotation enrichment analysis.

References

1. Satir P, Christensen ST. Structure and function of mammalian cilia. *Histochem Cell Biol* 2008; 129:687-93.
2. Marshall WF. The cell biological basis of ciliary disease. *J Cell Biol* 2008; 180:17-21.
3. Geremek M, Witt M. Primary ciliary dyskinesia: genes, candidate genes and chromosomal regions. *J Appl Genet* 2004; 45:347-61.
4. Canciani M, Barlocco EG, Mastella G, de Santi MM, Gardi C, Lungarella G. The saccharin method for testing mucociliary function in patients suspected of having primary ciliary dyskinesia. *Pediatr Pulmonol* 1988; 5:210-4.
5. Karadag B, James AJ, Gultekin E. Nasal and lower airway level of nitric oxide in children with primary ciliary dyskinesia. *Eur Respir J* 1999; 13:1402-1405.
6. Zariwala MA, Knowles MR, Omran H. Genetic defects in ciliary structure and function. *Annu Rev Physiol* 2007; 69:423-50.
7. Olbrich H, Häffner K, Kispert A *et al*. Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. *Nat Genet* 2002; 30(2):143-4.
8. Bartoloni L, Blouin JL, Pan Y *et al*. Mutations in the DNAH11 (axonemal heavy chain dynein type 11) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia. *Proc Natl Acad Sci* 2002; 99:10282-6.
9. Pennarun G, Escudier E, Chapelin C *et al*. Loss-of-function mutations in a human gene related to Chlamydomonas reinhardtii dynein IC78 result in primary ciliary dyskinesia. *Am J Hum Genet* 1999; 65:1508-1519.
10. Loges NT, Olbrich H, Fenske L *et al*. DNAI2 mutations cause primary ciliary dyskinesia with defects in the outer dynein arm. *Am J Hum Genet* 2008; 83:547-558.
11. Castleman V H, Romio L, Chodhari R *et al*. Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with

central-microtubular-pair abnormalities. Am J Hum Genet 2009; 84:197-209.

12. Duriez B, Duquesnoy P, Escudier E *et al.* () A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia. *Proc Natl Acad Sci* 2007; 104:3336-41.
13. Omran H, Kobayashi D, Olbrich H *et al*. Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. *Nature* 2008; 456:611-616.
14. Duquesnoy P, Escudier E, Vincensini L *et al*. Loss-of-function mutations in the human ortholog of Chlamydomonas reinhardtii ODA7 disrupt dynein arm assembly and cause primary ciliary dyskinesia. *Am J Hum Genet* 2009; 85:890-896.
15. Loges NT, Olbrich H, Becker-Heck A *et al*. Deletions and point mutations of LRRC50 cause primary ciliary dyskinesia due to dynein arm defects. *Am J Hum Genet* 2009; 85:883-889.
16. Merveille AC, Davis EE, Becker-Heck A *et al*. CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. *Nat Genet* 2011; 43:72-8.
17. Becker-Heck A, Zohn IE, Okabe N *et al*. The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. *Nat Genet* 2011; 43:79-84.
18. Geremek M, Bruinenberg M, Ziętkiewicz E *et al.* Gene expression studies in cells from primary ciliary dyskinesia patients identify 208 potential ciliary genes. *Hum Genet* 2011; 129:283-93.
19. Niedoszytko M, Bruinenberg M, de Monchy J *et al*. Gene expression analysis in predicting the effectiveness of insect venom immunotherapy. *J Allergy Clin Immunol* 2010; 125:1092-7.
20. Hackstadt AJ, Hess AM. Filtering for increased power for microarray data analysis.
    BMC Bioinfo 2009; 10:11.

21. Huang da W, Sherman BT, Tan Q *et al*. (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 2007; 8:R183.
22. Inglis PN, Boroevich KA, Leroux MR. Piecing together a ciliome. *Trends Genet* 2006; 22:491-500.
23. Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol* 2005; 18:1368-73.
24. Pazour GJ, Agrin N, Walker BL, Witman GB. Identification of predicted human outer dynein arm genes: candidates for primary ciliary dyskinesia genes. *J Med Genet* 2005; 43:62-73.
25. Shi L, Jones WD, Jensen RV *et al*. The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. *BMC Bioinfo* 2008; 9:S10.
26. Fliegauf M, Olbrich H, Horvath J *et al*. Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. *Am J Respir Crit Care Med* 2005; 171:1343-9.
27. Geremek M, Witt M. Primary ciliary dyskinesia: genes, candidate genes and chromosomal regions. *J Appl Genet* 2004; 45:347-61.

Part III

Discussion

Chapter 6

Genaral discussion

Primary ciliary dyskinesia (PCD) is a rare genetic disease characterized by recurrent respiratory tract infections, bronchiectasis and infertility. Pulmonary symptoms occur in PCD patients because of the lack of an efficient mucociliary clearance caused by immotility of respiratory cilia. Male infertility is caused by immotility of flagella in spermatozoids. The disease is genetically heterogeneous - mutations in many different genes have a similar functional effect on the motility of cilia and cause identical or similar phenotypes.

We undertook two different strategies in an attempt to elucidate the pathogenesis of PCD. Firstly, we applied a classical genetic approach to look for new PCD loci and genes using linkage analysis and candidate gene sequencing. We identified a genomic region on chromosome 15q containing a gene mutated in a subtype of PCD called Kartagener syndrome, which is characterized by the presence of *situs inversus* in approximately half of the affecteds in a pedigree. The Kartagener syndrome gene on chromosome 15q is still unknown. Secondly, we used a functional genomics approach to find new disease-associated pathways and functionally related genes in a genome-wide gene expression experiment. We identified 210 new ciliary genes through clustering analysis of the gene expression data. The results of our pilot gene expression analysis suggest that down-regulation of ciliary genes might be a general phenomenon in PCD and that, in the future, expression of selected genes might aid the diagnosis of the disease.

The genes identified by the genomic approach represents a list of PCD candidate genes and should be enriched in ciliary genes related to the motility. A combined analysis of the chromosome 15q linkage region and results of the genomic approach is presented in figure 1. Several candidate genes can be pinpointed based on the clustering analysis and gene expression results. Five genes located in the linkage region or its surroundings are present in the ciliary cluster described in the chapter 4. Mutations were not found in *C15ORF26*. Interestingly, *FAM154B* located in close vicinity of the minimal linkage region described in chapter 3 is present in the ciliary cluster. Seven genes located in the linkage region or its surroundings were

downregulated in PCD bronchial biopsies. *CCDC33* was downregulated in patient biopsies and is present in the ciliary cluster. Moreover it encodes a protein that contains a coiled-coil domain similarly to two PCD genes: *CCDC39* and *CCDC40*.

Several significant developments in the understanding of cilia proteomic composition, genetic control of cilia and ciliogenesis have occurred during the last few years. This progress was partially followed by medical genetic research in the field of PCD. Although several new disease genes have been identified, most of the cases still remain unexplained on the molecular level. The new PCD genes are mutated in only a few families or even in a single family. The main obstacle to finding more frequently mutated disease genes, if they exist, seems to be the disease's extensive genetic heterogeneity in the absence of phenotypical diversity. This problem will be discussed further in the context of new technological developments in genomics. At the end of this chapter, the possible clinical relevance of medical genetics research for PCD patients will be discussed.

Clinical homogeneity

Accurate clinical work-up is a key issue in medical genetic research. It is, in a way, the first stage of linkage analysis. Assuming that the clinical picture differs in patients with mutations in different genes, careful phenotypical stratification of patients can reduce the genetic heterogeneity of the cohort under study. Compared to the other ciliopathies, the clinical picture of PCD is less diverse. The patients present with otitis media with effusion (95% of patients) and upper respiratory tract symptoms, like rhinosinustitis (100% of patients) .[1, 2] Impaired mucociliary clearance in the lower respiratory tract results in recurrent bronchitis and pneumonia. Occurrence of bronchiectasis reflects the disease progression. The severity of the respiratory symptoms differs between affected individuals, however, it is probably related to other non-genetic (environmental) factors or immunity.

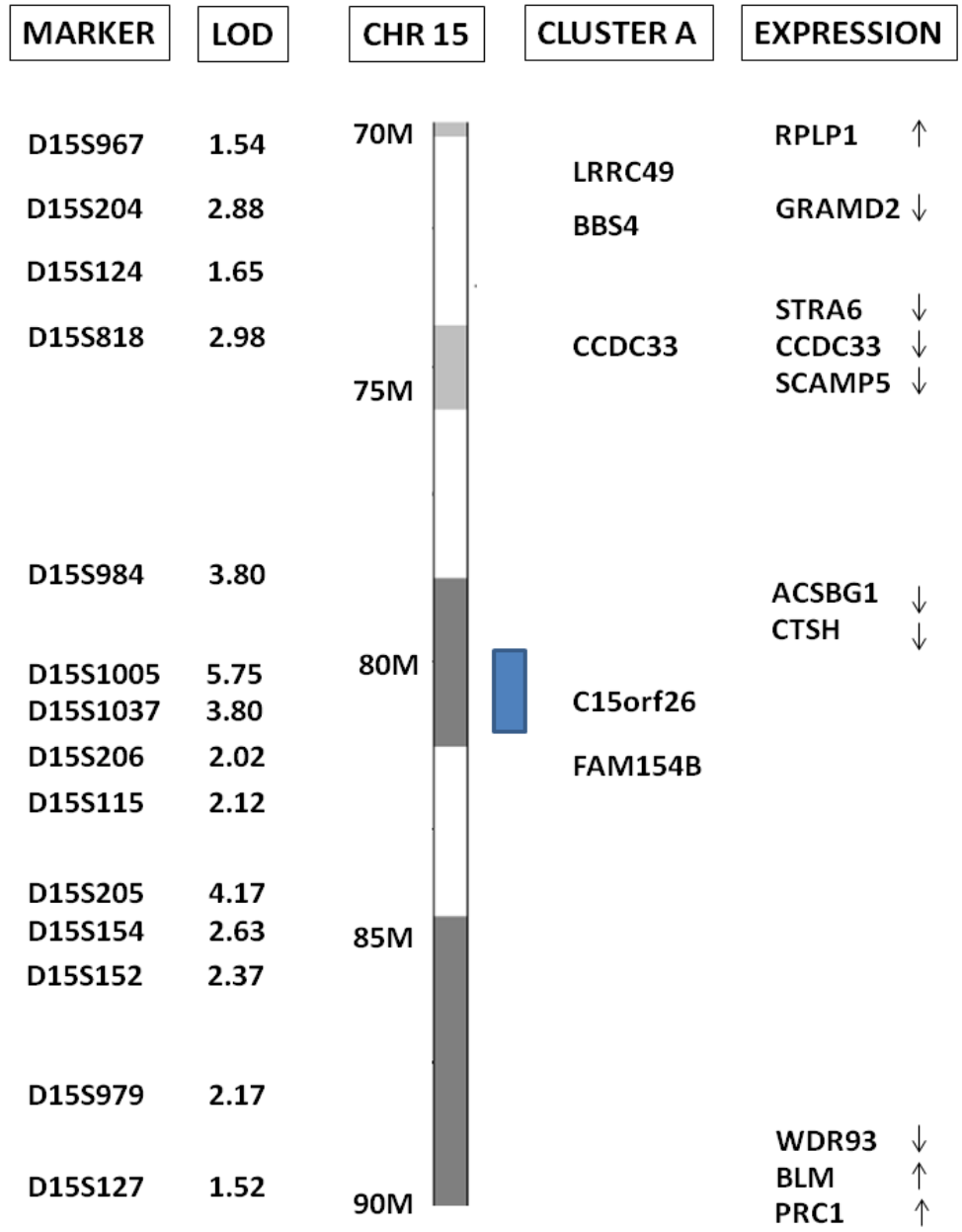

Figure 1. Intersection of the genetic and the genomic approach. On the left a selection of chromosome 15 microsatellite markers and corresponding LOD scores are presented (see chapter 1). The blue bar represents the region analyzed by sequencing (see chapter 2). Genes present in the ciliary cluster A (see chapter 3) and genes that were up- or downregulated in the patient biopsies are presented on the right (see chapter 4).

Situs inversus totalis occurs in approximately 50% of PCD patients, while 25% of patients with *situs inversus* have PCD. Therefore, the occurrence of respiratory symptoms together with abnormal visceral localization is a strong diagnostic indication. Mutations in *DNAI1* and *DNAH5*, the two PCD genes that were characterized in large cohorts of patients, cause *situs* randomization.[3] However, all the patients with mutations in *RSPH9* and *RSPH4A* described so far have normal visceral localization.[4] This suggests that for some PCD genes or mutations, cilia on the embryonic node are not affected despite the dyskinesia of respiratory cilia. Therefore, classification of families based on the presence of *situs inversus* might homogenize the studied cohort.

The electron microscopy imaging of cilia is a routine way of classifying PCD families phenotypically. Mutations in genes coding for proteins comprising the outer and inner dynein arms cause defects in their respective ultrastructural elements. Electron microscopic (EM) imaging of cilia is the most widely used and most successful way of phenotypically stratifying patients for genetic studies, but it requires an invasive diagnostic protocol and is therefore not always available. Moreover, patients with mutations in the same gene do not always have the same ultrastructural defects and electron microscopy in patients with *DNAH11* mutations shows normal ciliary ultrastructure .[5] Immunostaining experiments with antibodies targeting dynein proteins have revealed the existence of at least two distinct outer dynein arms types in normal ciliated airway epithelium. This indicates that ultrastructural components of cilia seen in electron microscopy, like the outer dynein arms, are molecularly heterogeneous and, therefore, defects described in EM imaging might, in the future, be divided into more categories allowing for more accurate phenotyping.

Females with PCD have reduced fertility and increased risk of ectopic pregnancy due to abnormal ciliary motility in the Fallopian tubes. Male infertility in PCD is caused by impaired motility of the sperm. However, some PCD patients are fertile and, according to one study, half of PCD patients have motile spermatozoa.[6]

The analysis of semen in adult PCD males seems to be a promising tool for the phenotypical stratification of male patients. Some patients with outer dynein arms defects in respiratory cilia have normal dynein arms in the sperm tail. This might indicate that the genetic control of ciliogenesis is different in the epithelium to that in spermatoza. Ultrastructural and molecular analysis of ciliary and sperm axonemes, together with semen motility analysis, might allow groups of patients with related molecular defects to be identified in the future.

Genetic heterogeneity

The relatively low diversity in the clinical picture of PCD is, however, accompanied by an extensive genetic heterogeneity. Mutations in twelve genes have been found in PCD patients. Except for *DNAH5* and *DNAI1*, which are mutated in approximately 20% and 10% of patients, respectively, the remaining genes have been implicated in very few families. Two strategies in cloning novel PCD associated genes were successful: genetic linkage in large consanguineous families and a candidate gene approach based on functional studies of cilium in model organisms. The limitation of these successful study designs is that the importance of the given gene had to be established afterwards by sequencing in a larger cohort. Linkage analysis in multiple pedigrees, if successful, can point to a potentially important gene. However, gene mapping in heterogeneous diseases such as PCD is a difficult task since the data contains noise coming not only from false-positive signals but also from other true disease loci.

Different aspects of ciliopathies can be understood in the context of cilia biology. Primary and motile cilia can be seen as two, not entirely independent, types of one organellum. Imaging, proteomic and gene expression studies indicate that genetic control, proteomic composition, ultrastructural and cellular organization, and the mechanism of ciliogenesis is at least partially shared by both types of cilia.[7, 8] Mutations of ciliary genes depend on the functional role played by the gene product affecting primary cilia, motile cilia or both of them. This explains the clinical overlap among the ciliopathies, i.e. the presence of *situs inversus* or respiratory

symptoms in primary cilia-associated diseases. Primary cilia can be found on many different cell types, which explains the clinical complexity of most of the ciliopathies. The PCD phenotype is less diverse because the disease is caused by dysfunction of only part of the ciliome related to motility.

Large, multi-protein complexes of the dynein arms are pre-assembled in the cytoplasm before they get transported via intraflagellar transport to the ciliary axoneme. The delivery of outer dynein arms complex to the cilium is impaired in PCD respiratory epithelium.[9] The presence in the cytoplasm of misassembled dynein arms could trigger the down-regulation of the ciliary genes expression. Therefore, the mechanism of dynein arm defects could be at least partly independent of the specific mutated gene. This would explain why genetic heterogeneity of PCD is not accompanied by heterogeneity in the phenotype.

The impact of technological progress and translational research

Although Mendelian disorders have a simple genetic disease model, several features make medical genetic analyses in this field complex. Many phenotypes of Mendelian diseases are extremely genetically heterogeneous, as in the case of deafness or retinitis pigmentosa that are caused by mutations in more than 100 genes. Mutations in the same gene may lead to phenotypes that are inherited as dominant or recessive traits, as in the case of *LMNA* lesions causing autosomal dominant or recessive forms of Emery-Dreifuss muscular dystrophy.[10] Digenic and triallelic inheritance has been described as a bridge between Mendelian and complex disorders. Pedigrees with Bardet-Biedl syndrome (BBS) have been identified in which three mutations at two different BBS loci segregate with expression of the disease traditionally classified as autosomal recessive ciliopathy.[11]

We need to identify specific DNA changes in order to definitively diagnose diseases, and provide counseling and reliable prognostic information. Reduced sequencing costs and the development of next-generation sequencing is now revolutionizing medical genetic research.[12, 13] So far, large families were needed

for linkage analysis and homozygosity mapping in order to find and fine-map genomic regions prior to sequencing of candidate genes. Next-generation sequencing makes large sequencing projects possible and will enable us to discover mutated genes in the linkage regions. It might also enable direct identification of mutated genes by exome or genome sequencing. The pilot results from the 1000 Genomes Project confirm the existence of thousands of low frequency and rare variants previously not described in both coding and non-coding DNA.[14] But analysis of large amounts of sequencing data in the context of disease-causing variants is challenging. Some successes have been booked: for example, sequencing of a 12.5 Mb region linked to autosomal recessive cerebellar ataxia, together with bioinformatic prioritization of variants based on evolutionary conservation, has identified mutations in a new ataxia gene, *ANO10*.[15] Whole genome sequencing in an individual suffering from Charcot-Marie-Tooth neuropathy revealed disease-causing mutations in the *SH3TC2* gene.[16] The initial sequence analysis revealed 9069 non-synonymous codon changes, including 121 nonsense mutations and 21 described as causing Mendelian disease. The disease gene was found only after limiting the genes in focus to those previously implicated in neuropathic or related conditions. This illustrates the enormous power of next-generation technology, but also shows that its application is still limited by our incomplete knowledge of the variation in the human genome. Before this limitation is overcome, we will still need to select candidate genes. In ciliopathies, long lists of ciliome genes, based on high-throughput proteomic and gene expression experiments, are already available. In the case of diseases with extensive genetic heterogeneity, such as PCD, which is at least partially characterized by private mutations present in only one or a few families, next-generation sequencing seems to be the method of choice for finding new genes, not only in unsolved linkage regions but also throughout the whole genome.

One of the main challenges for medical genetics research is to translate the scientific progress into a knowledge-based clinical approach. PCD remains an underdiagnosed condition. The high genetic heterogeneity of PCD and the low frequency of mutations in the known PCD genes limit the application of DNA

diagnostics in clinical work-up. Currently, there is no gene-oriented clinical approach, so that the patient benefits only modestly from information about the specific gene mutation he/she carries. Electron microscopy of cilia obtained from the nasal cavity, together with NO measurements and functional testing of cilia motility, might provide enough information to establish the diagnosis of PCD. In uncertain cases, additional tests like cilia immunostaining or gene expression might aid the diagnostic process. However, it is possible that, in the future, in an era of personalized medicine, the genomic sequence of the patient will be available for mutation search.

Functional studies of new PCD genes are needed before the genetic knowledge can be translated into therapeutic trials. Most of the PCD genes encode structural proteins, so that gene therapy holds the promise of successful treatment in the future, if the problem of vector delivery can be solved. As a proof of concept, DNAI1-deficient respiratory epithelial cell culture was transfected with cDNA containing the *DNAI* gene, and the immotile cilia recovered a normal beat and outer dynein arms reappeared.[17] A significant number of mutations in *DNAH5*, so far the most important gene in PCD, are premature stop codon mutations. The aminoglycosides are known to promote nonsense mutation suppression. Thus, antibiotic treatment could circumvent the genetic defect in patients with premature stop codons due to nonsense mutations. The presence of even a small amount of the protein in the cytoplasm could restore the normal regulatory network of the gene expression and remove the blockade of ciliary gene expression in PCD.

Concluding remarks

Despite the significant progress made in PCD research, the field is challenging and many future projects can be envisaged. Although several new disease genes have been identified, most of the cases remain unexplained at the molecular level. PCD seems to be characterized by a large number of private mutations; even genes that have been previously screened in a large number of pedigrees with negative results are sometimes found to be mutated in another study. However, this extensive

heterogeneity is not accompanied by such diversity in the clinical picture. Two possible research directions can be envisaged to solve this problem. Firstly, new diagnostic tools allowing high-resolution phenotyping could provide well-homogenized cohorts of patients for a classical genetic approach. And secondly, developments in genomics, like next-generation sequencing technology, could make large scale analyses of individual genomes possible.

High-resolution phenotyping based on immunostaining of specific ciliary components will allow further division of the currently known ultrastructural ciliary defects and enable genetic studies in more homogenous patient groups. Semen analysis has not yet been performed in a large group of PCD patients. The limited data that has been published so far suggests that semen analysis might be an interesting tool for stratifying the phenotypes of PCD patients.

Next-generation sequencing allows identification of mutations without needing extended pedigrees, which is important in a heterogeneous disease with infertility as a part of its clinical picture. Whole genome sequencing is still limited by the number of variations that have to be analyzed. However, in the case of ciliary genes, a large number of candidates can be selected based on high-throughput proteomic or gene expression data. Gene expression changes of ciliary genes seem to be an important aspect of PCD pathogenesis. Further analysis performed on a larger cohort of patients, and including bronchial biopsies from individuals with a known mutational status and ultrastructural defects, is needed to elucidate the role of global gene expression changes in the pathogenesis of PCD.

Careful examination of a large cohort of PCD patients might point to new diseases rarely associated with PCD and therefore with cilia motility, as in the case of retinitis pigmentosa or mental retardation. Some diseases of unknown or complex etiology might be related to cilia motility. Hydrocephalus is part of a phenotype in many animal models of PCD. Although in man, hydrocephalus is not part of the classical clinical picture of PCD, cilia motility genes might be involved in normotensive hydrocephalus. Cilia motility genes also represent male infertility

candidate genes. Mutations in *DNAH5*, *DNAH11* and *DNAI1* have been described in individuals with isolated asthenozoospermia, a common cause of male infertility.[18]

References

1. Bush A, Chodhari R, Collins N, Copeland F *et al*. Primary ciliary dyskinesia: current state of the art. *Arch Dis Child* 2007; 92:1136-40.
2. Noone PG, Leigh MW, Sannuti A *et al*. Primary ciliary dyskinesia: diagnostic and phenotypic features. Am J *Respir Crit Care Med* 2004; 169:459-67.
3. Leigh MW, Pittman JE, Carson JL *et al*. Clinical and genetic aspects of primary ciliary dyskinesia/Kartagener syndrome. *Genet Med* 2009; 11:473-87.
4. Castleman V H, Romio L, Chodhari R *et al*. Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with central-microtubular-pair abnormalities. *Am J Hum Genet* 2009; 84:197-209.
5. Schwabe GC, Hoffmann K, Loges NT *et al.* Primary ciliary dyskinesia associated with normal axoneme ultrastructure is caused by DNAH11 mutations. *Hum Mutat* 2008; 29:289-98.
6. Munro NC, Currie DC, Lindsay KS *et al*. Fertility in men with primary ciliary dyskinesia presenting with respiratory infection. *Thorax* 1994; 49:684–7.
7. Satir P, Christensen ST. Structure and function of mammalian cilia. *Histochem Cell Biol* 2008; 129:687-93.
8. Bloodgood RA. Sensory reception is an attribute of both primary cilia and motile cilia. *J Cell Sci* 2010; 123:505-9.
9. Fliegauf M, Olbrich H, Horvath J *et al*. Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. *Am J Respir Crit Care Med* 2005; 171:1343-9.

10. Prigogine C, Richard P, Van den Bergh P *et al*. Novel LMNA mutation presenting as severe congenital muscular dystrophy. *Pediatr Neurol* 2010; 43:283-286.
11. Katsanis N, Ansley SJ, Badano JL *et al*. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science* 2001; 293:2256-9.
12. Nikopoulos K, Gilissen C, Hoischen A *et al*. Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy. *Am J Hum Genet* 2010; 86:240-7.
13. Clendenning M, Buchanan DD, Walsh MD *et al*. Mutation deep within an intron of MSH2 causes Lynch syndrome. *Fam Cancer* 2011; 10:297-301.
14. Durbin RM, Abecasis GR, Altshuler DL, Auton A et al. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467:1061-73.
15. Vermeer S, Hoischen A, Meijer RP *et al*. Targeted next-generation sequencing of a 12.5 Mb homozygous region reveals ANO10 mutations in patients with autosomal-recessive cerebellar ataxia. *Am J Hum Genet* 2010; 87:813-9.
16. Roach JC, Glusman G, Smit AF *et al*. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010; 328:636-9.
17. Chhin B, Negre D, Merrot O *et al*. Ciliary beating recovery in deficient human airway epithelial cells after lentivirus ex vivo gene therapy. *PLoS Genet* 2009; 5.
18. Neesen J, Kirschner R, Ochs M *et al*. Disruption of an inner arm dynein heavy chain gene results in asthenozoospermia and reduced ciliary beat frequency. *Hum Mol Genet* 2001; 10:1117-28.

Summary

Motile cilia are small cellular projectiles that by synchronized beating propel fluid over the tissues. Cilia generate the flow of mucus and cerebrospinal fluid in the respiratory tract and in the brain, respectively, whereas in the Fallopian tubes cilia help to move the ovum toward the uterus. Flagella, built on a similar to cilia scheme, provide motility to spermatozoa. The motility of cilia plays an important role in initial left-right symmetry breaking event during embryogenesis. Primary ciliary dyskinesia (PCD) is a rare genetic disorder caused by mutations in genes functionally related to cilia motility. Lack of ciliary beating leads to recurrent respiratory tract infections and bronchiectasis because of defective mucociliary clearance. Kinetic dysfunction of the sperm tails causes infertility. Due to immotility of cilia on the embryonic node, left-right determination is randomized causing *situs inversus* in approximately half of the patients. PCD is inherited, in most of the cases, as autosomal recessive trait. The genetic heterogeneity of the disease is very high. Mutations in eleven genes have been identified in PCD patients so far. Despite this, large majority of cases remain unexplained on the molecular level. High genetic heterogeneity in conjunction with low diversity of the clinical phenotype makes difficult cloning of new genes by means of linkage analysis. Cilia are build of multiprotein complexes controlled by hundreds or even thousands of genes which makes selection of disease candidate genes a difficult task.

In this thesis classical genetic and genomic approaches have been undertaken to study the pathogenesis of PCD. Chapter 2 describes genome wide linkage analysis in 70 PCD families. We hypothesized that genetic background of PCD might be different in families with Kartagener syndrome (KS), a subtype of PCD, that is characterized by *situs inversus* accompanying typical PCD symptoms. Therefore in order to reduce the genetic heterogeneity we divided PCD families in two subgroups : ciliary dysfunction-only (CDO) when all the patients in the family had *situs solitus,* and Kartagener syndrome (KS) when at least one patient in the family had *situs inversus*. We genotyped 462 microsatelite markers spread in the

genome in 52 KS and 18 CDO families. The maximal pairwise LOD score of 3.36 in the KS families indicated linkage of a KS locus to the long arm of chromosome 15. Genotyping of additional markers on chromosome 15q returned maximal pairwise LOD score of 4.34 for a marker localized in the chromosomal region 15q24-25.

As described in chapter 3, in order to identify the KS gene on chromosome 15q we focused on the minimal linkage region defined by the two closest recombination events. By genotyping additional microstatelite markers, the KS candidate region was refined to a 1.8 Mb segment containing 18 known genes. The coding regions of these genes and three neighboring genes were subjected to sequence analysis in 7 KS probands. We found a number of single nucleotide sequence variants, some of which resided in mRNA coding sequences. However, none of the variations alone could explain the occurrence of the disease in these patients.

The genomic approach enables analysis of functional changes of large group of genes occurring in a disease. We hypothesized that defects of multi-protein complexes in PCD cilia should be reflected by gene expression changes in the respiratory epithelium.

In the analyses described in chapter 4 we used the Quality Threshold clustering algorithm to identify groups of genes that revealed highly correlated RNA expression patterns in the PCD biopsies. The largest cluster contained 372 genes and was significantly enriched for genes related to cilia. The database and literature search showed that 150 genes in this cluster were known cilia genes, strongly indicating that the remaining 210 genes were likely to be new cilia genes. The genes identified in our studies represents a list of PCD candidate genes and should be enriched in ciliary genes related to the motility.

Chapter 5 describes an explorative analysis of differential gene expression in the bronchial tissue from six PCD patients and nine non-PCD controls, using Illumina HumanRef-12 Whole Genome BeadChips. Annotation analysis showed that the genes down-regulated in PCD biopsies were significantly enriched for terms

related to cilia. Our data suggest that the down-regulation of the ciliome genes plays an important role in the molecular pathomechanism of PCD.

Samenvatting

Trilharen (of cilia) zijn sterk gedifferentieerde structuren die specifieke cellulaire functies vervullen, zoals voortbeweging, waarneming van omgevingssignalen, of het geleidelijk naar buiten werken van de slijmlaag in de luchtwegen. Positionering van organen in het lichaam is deels ook afhankelijk van de werking van trilharen.

Primaire ciliaire dyskinesie (PCD) is een erfelijke aandoening die gepaard gaat met een verminderde beweeglijkheid van de trilharen. Doordat hierdoor de luchtwegen worden aangetast, ontstaan er blijvende verwijdingen van de luchtwegen (bronchiën), aangeduid als bronchiëctasie en een chronische bijholteontsteking.

Er bestaat ook een stoornis van de beweeglijkheid van de trilharen waarbij de organen in spiegelbeeld gepositioneerd zijn (situs inversus); deze noemt men wel het syndroom van Kartagener.

In dit proefschrift wordt onderzoek gepresenteerd gericht op het ontrafelen van de oorzaak van PCD middels klassieke genetische en genomische studies.

Het klassieke genetische onderzoek maakt gebruik van koppelingsonderzoek (linkage analyse). Dit onderzoek richt zich op het vinden van een ziektegen met behulp van zogenaamde genetische markers waarbij de overerving van het ziektegen dat de ziekte veroorzaakt, eenzelfde overervingspatroon volgt als de DNA marker.

In hoofdstuk 2 wordt koppelingsonderzoek in 70 families met PCD beschreven. In deze studie is de aanname gedaan dat patiënten met het syndroom van Kartagener genetisch gezien andere afwijkingen hebben dan patiënten met PCD. Het hele erfelijke materiaal van 52 families met het syndroom van Kartagener is bekeken met 462 microstalliet markers. Eenzelfde studie is gedaan in 18 families met PCD. Deze studie suggereerde dat genen op chromosoom 15q24-q25 betrokken zijn bij het syndroom van Kartagener (LOD score 4.34).

In hoofdstuk 3 wordt verder ingezoomed op chromosoom 15q24-q25, specifiek in families met het syndroom van Kartagener. Door het testen van extra

genetische markers werd het chromosomale gebied gereduceerd tot 1.8 Mb; dit gebied bevat 18 verschillende genen die ieder voor zich positionele kandidaat genen zijn. Van ieder gen is het coderende deel bekeken op aanwezigheid van ziekteveroorzakende mutaties middels sequentie analyse in 7 onafhanelijke patiënten met het syndroom van Kartagener (afkomstig uit de meest veelbelovende families). De DNA sequentie analyse leverde een aantal variaties op, maar geen enkele variatie kon het voorkomen van het syndroom van Kartagener in de families verklaren.

Met behulp van genoomonderzoek is het mogelijk om functionele verandering van grote groepen genen tegelijkertijd te onderzoeken. Dit onderzoek maakt gebruik van zogenaamde DNA-chips of microarrays. Hiermee kan heel snel en nauwkeurig DNA-materiaal van gezonde mensen en PCD patiënten worden vergeleken.

In hoofdstuk 4 wordt beschreven welke genen tot expressie komen in weefsel van PCD patiënten. In deze studie is de aanname gedaan dat afwijkingen in de trilharen gereflecteerd worden door afwijkingen in de expressie van genen die onderdeel uit maken van de cellulaire trilhaar complexen in het epitheel van de luchtwegen. Met behulp van bioinformatica is een groep van 372 genen geidentificeerd, waarin 150 bekende cilia genen maar ook 210 nieuwe cilia genen. Deze 210 nieuwe cilia genen zijn interessante kandidaten om te testen voor genetische afwijkingen in families.

In hoofdstuk 5 wordt heel specifiek het verschil onderzocht tussen PCD patiënten en gezonde mensen door wederom alle 22.000 genen in het menselijk genoom te onderzoeken. Ook dit onderzoek laat zien dat genen die deel uitmaken van cilia een afwijkend expressie patroon laten zien in PCD patiënten in vergelijking met gezonde controles. Het lijkt er met name op dat de cilia-gerelateerde genen lager tot expressie komen in PCD patiënten.