THE ROLE OF SPAG1 IN THE CYTOPLASMIC ASSEMBLY OF AXONEMAL DYNEIN ARMS IN
HUMAN AIRWAY EPITHELIA

Amanda Jo Smith

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial
fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Cell Biology
and Physiology in the School of Medicine.

Chapel Hill
2022

Approved by:

Lawrence E. Ostrowski

Richard E. Cheney

Robert Tarran

Maimoona A. Zariwala

Wanda K. O'Neal

ABSTRACT

Amanda Jo Smith: The Role of SPAG1 in the Cytoplasmic Assembly of
Axonemal Dynein Arms in Human Airway Epithelia
(Under the direction of Lawrence E. Ostrowski)

Motile cilia are microtubule-based hair-like organelles that extend from the apical surface of cells

and are essential for the propulsion of fluid or cells across the respiratory, ependymal, oviduct, and nodal

epithelium. Dysfunctional motile cilia, due to genetic variants in proteins associated with the structure,

regulation, and biogenesis of motile cilia, result in primary ciliary dyskinesia (PCD), a genetically

heterogeneous, rare disorder characterized by chronic oto-sino-pulmonary disease, infertility, and

laterality defects. The most common ciliary defect seen in PCD is absence of the multiprotein motor

complexes known as axonemal dynein arms. While the composition and regulation of dynein arms have

been extensively studied, how these large, complex structures are assembled remains poorly understood.

Mutations in the gene sperm-associated antigen 1 (*SPAG1*) cause PCD with defects in both the inner and

outer dynein arms, resulting in completely immotile cilia. Due to the observed dynein defects and the

cytoplasmic location of SPAG1, it has been proposed to be a dynein axonemal assembly factor (DNAAF)

that facilitates the assembly of dynein arms in the cytoplasm before their transport into the cilium. To

further elucidate SPAG1's role in dynein assembly, we examined its expression, interactions, and ciliary

defects caused by its absence in human airway epithelia. A previously uncharacterized 60-kDa isoform of

SPAG1 was discovered. The full-length and 94-kDa isoforms of SPAG1 expression corresponded with

ciliogenesis, while the 60-kDa isoform was constitutively expressed. Immunoprecipitations showed that

SPAG1 interacts with multiple DNAAFs, dynein chains, and canonical components of the R2TP

complex, and we identified protein interaction domains between SPAG1 and another DNAAF, DNAAF2.

Examination of PCD subjects with an atypical ultrastructural defect for *SPAG1* variants demonstrated that

the 60-kDa SPAG1 can partially compensate for the absence of full-length SPAG1 to assemble a reduced number of outer dynein arms. In *SPAG1* mutants, dynein heavy chain (DHC) protein levels and their interaction with dynein intermediate chains (DICs) were reduced. In summary, our data demonstrate that SPAG1 is necessary for axonemal dynein arm assembly by scaffolding R2TP-like complexes composed of several DNAAFs that facilitate the folding of the DHCs and/or binding of the DHCs to the DIC complex.

To the individuals and families affected by Primary Ciliary Dyskinesia.
You all are the reason behind my motivation, passion, and purpose.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my graduate advisor, Dr. Larry Ostrowski. It has certainly been an honor to work in your lab and to be your first Ph.D. student. I feel like it was a learning experience for both us and one for which I am incredibly grateful. Thank you for providing guidance, endless encouragement, and for sharing your expertise. You taught me to be confident in my research and confident in myself as a scientist, but also provided opportunities for me to improve as a researcher. Words cannot express my gratitude for your patience and understanding, especially throughout the pandemic. I am forever grateful for how you have been a mentor to me and for what you do to advance research for PCD. I would also like to thank the current and former members of the Ostrowski lab, especially Dr. Ximena Bustamante-Marin, Weining Yin, and Dr. Patrick Sears. Thank you all for the time and effort that you have helped me with learning new methods or techniques, troubleshooting experiments, or providing feedback and encouragement on my work. You all have helped shape me into the scientist that I am today.

I am also grateful for my thesis committee, Dr. Richard Cheney, Dr. Maimoona Zariwala, Dr. Robert Tarran, and Dr. Wanda O'Neal, for their advice, feedback, and support throughout my graduate career. I would like to especially thank my committee chair, Dr. Richard Cheney, for quickly answering my questions and providing advice for navigating through the logistics of graduate school. I would also like to thank Dr. Robert Tarran and Dr. Maimoona Zariwala, not only for providing equipment, reagents, advice, and feedback as colleagues, but also for serving as mentors and allowing me to work in their labs when I was an undergraduate student. The experiences that I had participating in research in your labs as an undergraduate confirmed my goals of pursuing my Ph.D. and have helped navigate me to this point in my career.

I would also like to extend my sincere thanks to the Department of Cell Biology and Physiology and the Marsico Lung Institute. I would like to acknowledge the many members of these departments that have provided feedback and assistance after I have presented my work at seminars and for providing funding for traveling to conferences. I give a special thanks to members of the Marsico Lung Institute, especially those in the Tissue Procurement and Cell Culture Core and the Molecular Biology Core, for providing reagents, training of techniques, suggestions, and ideas to accomplish the aims of this dissertation. I would like to express my deepest appreciation to members of the PCD Research Group, or the "cilia group." I have learned so much from each of you over the years, and this endeavor would not have been possible without you.

I cannot express how grateful I am to my family and friends. Thank you to all of my friends over the years for your support, your laughter, and your love. Thank you for reminding me it's okay to have fun outside of the lab sometimes, for partaking in some of my pastimes and hobbies with me, for sharing in some of my favorite memories, and for being in my life. To my family, I will always be forever grateful for your unwavering support, your encouragement to pursue my dreams, and your help every step of the way. I am so fortunate to have you all as my family. Words cannot express the love and gratitude that I feel for you all. I would also like to thank my dog, my Siberian Husky Sora who I received as a gift at the start of graduate school, as I am incredibly thankful for and have enjoyed the cuddles, laughter, and emotional support from him over the years.

Lastly, I would like to express my deepest appreciation and gratitude to my fiancé and love, Cecil. You have been my biggest supporter and fan since day one. Your words of encouragement, constant support, and unconditional love is what has helped me achieve this dream of mine, especially during times when things became difficult. Even the simple things, like bringing me Starbucks and dinner during late nights in the lab, meant so much to me. I truly could not have accomplished this major achievement or be where I am at in life without you. I love you so much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AGC	Automatic gain control
ALI	Air/liquid interphase
AMPK	Adenosine monophosphate-activated protein kinase
AP-MS	Affinity purification-mass spectrometry
ASL	Airway surface liquid
ATP	Adenosine triphosphate
BEGM	Bronchial epithelial growth medium
BSA	Bovine serum albumin
CA	Central apparatus
CaCC	Ca^{2+}-activated chloride channel
cAMP	Cyclic adenosine monophosphate
CBF	Ciliary beat frequency
CCT	Chaperonin containing TCP1
CFM	Conditioned fibroblast medium
CFTR	Cystic fibrosis transmembrane conductance regulator
CGM	Complete growth medium
cGMP	Cyclic guanosine monophosphate
CRC	Conditionally reprogrammed cells
CRISPR	Clustered regularly interspaced short palindromic repeat
CS	CHORD-containing proteins and SGT1
DAVID	Database for Annotation, Visualization and Integrated Discovery
ddPCR	Droplet digital polymerase chain reaction
DDT	Dithiothreitol
DHC	Dynein heavy chain
DIC	Dynein intermediate chain

DIC/LC	Dynein intermediate chains/light chains complex
DLC	Dynein light chain
DLIC	Dynein light intermediate chain
DMEM	Dulbecco's modified eagle medium
DNAAF	Dynein axonemal assembly factor
DynAP	Dynein axonemal particle
EDTA	Ethylenediaminetetraacetic acid
ENaC	Epithelial Na^+ channel
ERK1/2	Extracellular signal☐regulated protein kinase 1/2
FBS	Fetal bovine serum
FDR	False discovery rate
FEV_1	Forced expiratory volume in the first second
FRET	Fluorescence Resonance Energy Transfer
Gβ1	G protein β 1 subunit
GFP	Green fluorescent protein
gRNA	Guide RNA
GSP	Gene-specific reverse primer
GTP	Guanosine triphosphate
HAEC	Human airway epithelial cell
HBEC	Human bronchial epithelial cell
HCD	Higher energy C-trap dissociation
HNEC	Human nasal epithelial cell
HSP	Heat shock protein
HSVMA	High-speed video microscopy analysis
IDA	Inner dynein arm
IFT	Intraflagellar transport

JMEM	Joklik minimum essential medium
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LRRC	Leucine-rich repeat containing
LUMIER	Luminescence-based mammalian interactome
MCC	Mucociliary clearance
MTD	Microtubule doublet
NCBI	National Center for Biotechnology Information
N-DRC	Nexin-dynein regulatory complex
nNO	Nasal nitric oxide
ODA	Outer dynein arm
PBS	Phosphate-buffered saline
PCD	Primary ciliary dyskinesia
PCL	Periciliary layer
PCP	Planar cell polarity
PDAC	Pancreatic ductal adenocarcinomas
PDGFR	Platelet-derived growth factor receptor
PFA	Paraformaldehyde
PIH1	Protein interacting with HSP90 1
PIH1D1	PIH1 domain-containing protein 1
PIH1D2	PIH1 domain-containing protein 2
PKC	Protein kinase C
PRM	Parallel reaction monitoring
R2SD	RUVBL1-RUVBL2-SPAG1-DNAAF2
R2SP	RUVBL1-RUVBL2-SPAG1-PIH1D2
R2TP	Rvbl1-Rvb2-Tah1-Pih1
RACE	Rapid amplification of cDNA ends

RGMC	Reduced generation of multiple motile cilia
ROCK	Rho-associated kinase
RPAP3	RNA polymerase II associated protein 3
RS	Radial spoke
RT	Room temperature
RT-PCR	Reverse-transcription polymerase chain reaction
RUVBL1	RuvB-like 1
RUVBL2	RuvB-like 2
SD	Standard deviation
sgRNA	Single guide RNA
SHH	Sonic hedgehog
SILAC	Stable isotope labeling by amino acids in cell culture
SPAG1	Sperm-associated antigen 1
TBS	Tris-buffered saline
TBST	TBS with 0.1% Tween 20
TE	Tris-EDTA
TEM	Transmission electron micrographs
TPR	Tetratricopeptide repeat
UPM	Universal Primer Mix
UTP	Uridine triphosphate
VUS	Variant of uncertain significance
WDR92	WD repeat-containing protein 92
WES	Whole-exome sequencing
WT	Wild type

CHAPTER 1: MOTILE CILIA AND PRIMARY CILIARY DYSKINESIA

1.1 Cilia

Cilia are microtubule-based organelles that protrude from the cell surface of almost all cell types in humans. They are highly conserved, complex, hair-like structures that are present in most eukaryotes and are composed of over 650 proteins (Fliegauf et al., 2007). These structures can either be found as a single cilium (a monocilium) on most cell types or up to 200 – 300 cilia (multicilia) on the apical surface of specialized cells in the respiratory airways, the ependyma, the oviducts in females, and the efferent ducts in males (Wallmeier et al., 2020). All cilia possess the basic structural components of the outer microtubule doublets and the ciliary membrane; however, there are variations in the ciliary ultrastructure based on the cilia type. Cilia are usually categorized based on the dichotomy of being immotile, also known as primary or sensory cilia, or being motile (Fliegauf et al., 2007).

Primary, or sensory, cilia act as antennas that sense environmental cues and extracellular signals. The majority of cell types have a primary cilium, which is present as a monocilium that has a $9 + 0$ axonemal ultrastructure that lacks the central pair of singlet microtubules and motor proteins found in motile cilia. The primary cilium is known for being assembled and reabsorbed dependent on the cell cycle phase, where it is only present on a cell in a quiescent state, i.e., interphase of the cell cycle and will be reabsorbed prior to cell division (Fliegauf et al., 2007). Primary cilia function by detecting extracellular signals through ion channels, receptor proteins, and signaling molecules in the membrane of the primary cilium and transforming these cues into signaling cascades that are propagated to the rest of the cell. Primary cilia facilitate development, differentiation, and renal function, as well as the olfactory, auditory, and visual senses, through processes that include signaling pathways, chemosensation, and mechanosensation of fluid flow across cells. These signaling pathways can include the sonic hedgehog (SHH), planar cell polarity (PCP), platelet-derived growth factor receptor (PDGFR), and calcium

signaling pathways (Fliegauf et al., 2007; Mitchison and Valente, 2017). While there has been some sensory function identified in motile cilia, specifically detecting bitter taste, sex hormones, shear stress, and fluid flow (Jain et al., 2012), the main function of motile cilia is to propel extracellular fluid or cells, specifically gametes, across the epithelium (Wallmeier et al., 2020). In this work, the main focus is on motile cilia and its dysfunction, and thus, the focus will be on this specific cilia type from hereafter.

1.1.1 Multiciliogenesis

Motile cilia are only assembled when epithelial cells reach a terminally differentiated state, where this multiciliated cell fate is largely determined by the inhibition of NOTCH1 signaling. In human airways, NOTCH1 is inhibited by the microRNAs R-34 and miR-449, and this inhibition results in the stimulation of geminin coiled-coil domain-containing protein 1 (*GEMC1*) and multicilin (*MCIDAS*) expression. GEMC1 and MCIDAS are the main regulators of ciliogenesis and are essential for determining multiciliated cell fate. Once GEMC1 and MCIDAS are activated, they form a complex with transcription factors E2F4 or E2F5. This complex drives the activation of *P73* and other transcription factors that regulate ciliogenesis, including forkhead box J1 (*FOXJ1*), *RFX2*, *RFX3*, and *MYB*, which further commits progenitor cells to differentiate into multiciliated cells (Spassky and Meunier, 2017). The induction of these multiciliogenesis genes leads to expression of multiple proteins associated with assembling motile cilia, which will eventually lead to massive centriole amplification, basal body docking, and ciliogenesis (Meunier and Azimzadeh, 2017).

To acquire 200 to 300 cilia per cell during multiciliated cell differentiation, the structures that will eventually nucleate motile cilia, known as basal bodies, are derived from massive centriole amplification that occurs through two pathways. The first pathway is the CEP63-dependent canonical pathway that produces additional daughter centrioles from parental centrioles, which forms about 10% of the centrioles produced during multiciliogenesis. The second pathway is where centrioles are produced from spherical structures composed of DEUP1 aggregates termed deuterosomes, which form about 90% of the centrioles produced in multiciliogenesis (Spassky and Meunier, 2017; Legendre et al., 2021). Once fully formed, centrioles will detach from centrosomes and deuterosomes and travel to the apical membrane via

cytoplasmic actin stress fibers of epithelial cells. These centrioles then dock, a process dependent upon cortical actin, using distal appendages bound to the plasma membrane, where they subsequently nucleate motile cilia and are now known as basal bodies. To produce an efficient fluid flow in a single direction across epithelium, motile cilia have to be oriented in a similar plane, which is achieved by the positioning and orientation of the basal bodies. Basal body orientation is predisposed based on the distribution of polarized PCP proteins, but is then further refined to the correct orientation through processes involving hydrodynamic forces produced by fluid flow and the PCP pathway (Meunier and Azimzadeh, 2017; Spassky and Meunier, 2017).

Elongation and extension of the ciliary axoneme and membrane occurs by addition of tubulin dimers at the plus-ends of axonemal microtubules and addition of motile cilia ultrastructures. This process is mediated by the ciliary gate, or transition zone, of the cilia that regulates the proteins entering and exiting the cilium, as well as a mechanism known as intraflagellar transport (IFT). Structures found in the axoneme, specifically the motor proteins called dynein arms, are assembled in the cell body before their recruitment and transport into the ciliary axoneme by IFT (Avasthi and Marshall, 2012). IFT is the active, bi-directional movement of multiprotein complexes, called IFT trains, that bind to cargo (tubulin, dyneins, and other structures found in motile cilia) and motor proteins (kinesin-II and dynein-2) that is required for the anterograde and retrograde transport of cargo in and out of the ciliary axoneme. Finally, the length and stability of the ciliary axoneme is regulated by tubulin turnover at the plus-ends of the axoneme mediated by IFT and post-translational modifications, specifically acetylation and detyrosination, of tubulin (Avasthi and Marshall, 2012).

1.1.2 Functions of Motile Cilia

Multiciliated cells are located throughout the respiratory tract. The respiratory tract in humans stretches from the nose and nasal passages to the alveoli located within the lungs. It is divided between the upper respiratory tract, which includes the nasal passages, the paranasal sinuses, the oral cavity, and the pharynx, and the lower respiratory tract, which includes the trachea and the lungs. The lower respiratory tract is further divided into two sections known as the conducting zone, which includes the

trachea, the main bronchi, and the bronchioles in the lungs, and the respiratory zone, which includes the respiratory bronchioles, alveolar ducts, and alveoli in the lungs. The trachea and the conducting zone of the lungs are composed of epithelium supported by a collagenous lamina propria that contains elastin, smooth muscle that is arranged in a spiral pattern, and cartilage plates that become less prominent traveling from the trachea to the terminal bronchioles (Bustamante-Marin and Ostrowski, 2017). The airway epithelium is composed of many different specialized cell types that change prevalence dependent on the location within the respiratory tract. Within the conducting zone of the lower respiratory tract, there are basal stem cells, club cells, multiciliated cells, and goblet cells, along with mucus cells and serous cells found in submucosal glands. Ionocytes, neuroendocrine, and brush or tuft cells are rare cell types found dispersed in airway epithelium (Bustamante-Marin and Ostrowski, 2017; Lengendre et al., 2021). Within the respiratory zone of the lower respiratory tract, the small airways (<2 mm in diameter) have a simple cuboidal epithelium with an increase in club cells and decrease in goblet and multiciliated cells. Within the alveoli, type 1 alveolar cells participate in gas exchange while type 2 alveolar cells are involved in pulmonary surfactant secretion (Bustamante-Marin and Ostrowski, 2017).

Focusing on the different cell types found in the respiratory tract, basal stem cells are KRT5[+] and TP63[+] cells that are attached to the basal lamina and are progenitors of all other airway cell types. Club cells are anti-inflammatory luminal cells that facilitate xenobiotic detoxification and can differentiate into multiciliated cells through transdifferentiation (Legendre et al., 2021). Goblet, mucous, and serous cells are luminal secretory cells that produce and secrete gel-forming mucins, which are high molecular weight, heavily glycosylated proteins, and/or water (Bustamante-Marin and Ostrowski, 2017). Neuroendocrine cells act as airway sensors that emit neurotransmitters in response to stimuli including allergens, tobacco smoke, pathogens, hypoxia, and mechanical stretching (Xu et al., 2020). Brush cells are chemosensory epithelial cells that express taste receptors and are physically characterized by an apical tuft of microvilli (Krasteva et al., 2011). Ionocytes are a recently discovered, rare cell type with high expression of the cystic fibrosis transmembrane conductance regulator (CFTR) chloride channel (Plasschaert et al., 2018). Finally, multiciliated cells are elongated columnar cells that are approximately 20 μm-high and 6 μm-

wide with limited contact with the basement membrane and 200 to 300 motile cilia extending from their apical surface (Legendre et al., 2021). Multiciliated cells make up approximately 60 – 70% of the cell population of the trachea and bronchi, while goblet cells constitute 20% of cells (Bustamante-Marin and Ostrowski, 2017).

Between the apical surface of airway epithelium and the airway lumen is a thin layer of fluid referred to as airway surface liquid (ASL). ASL consists of two distinct sub-layers: an overlaying mucus layer and a low viscosity periciliary liquid layer (PCL). The mucus layer is mainly composed of water (97%), mucin glycoproteins (1%), salt (1%), and other proteins (1%), and is a protective layer that captures particles and pathogens that enter into the airways. The PCL is a watery, polyanionic layer that contains membrane-associated mucins with a depth of approximately 7 μm, the height of cilia (Bustamante-Marin and Ostrowski, 2017). The PCL provides lubrication and an optimal environment for the beating of cilia and transport of the mucus layer, as well as a barrier between particles and pathogens from entering airway epithelial cells (Button et al., 2012; Bustamante-Marin and Ostrowski, 2017). The hydration of both layers of the ASL is dependent on the export and import of Cl^- through the CFTR and Ca^{+2}-activated chloride channels (CaCC) and Na^+ through the epithelial Na^+ channel (ENaC), respectively. Without proper hydration of the ASL, the mucus layer collapses onto the layer of cilia, suppressing their ability to beat and thus inhibiting transport of the mucus layer (Bustamante-Marin and Ostrowski, 2017).

Respiratory motile cilia beat in a coordinated fashion to generate metachronal waves that travel across the epithelium, and in contact with the mucus layer through their ciliary tips, propel mucus towards the nose and mouth (Bustamante-Marin and Ostrowski, 2017). This process is termed mucociliary clearance (MCC) or mucociliary transport. MCC is an innate airway defense mechanism that allows for inhaled particles and pathogens that are trapped in the mucus layer to be removed from the airways, where they are eventually expelled at the nose and mouth or at the nasopharynx by swallowing. Motile cilia in normal epithelium beat at a frequency that ranges between 10 and 20 Hz, which attributes to a mucociliary clearance rate of about 5 mm/min (Legendre et al., 2021). Factors that can influence the

mucociliary clearance rate include the ciliary beat frequency (CBF) and waveform, the number of functional cilia, the length of the cilia, and the coordination and orientation of the ciliary beating. The CBF, and in turn the mucociliary clearance rate, can be stimulated by an intracellular increase of cAMP, cGMP, and Ca^{2+} or an extracellular increase of ATP or UTP. CBF is also temperature and pH dependent (Bustamante-Marin and Ostrowski, 2017). Aside from the upper and lower respiratory tracts, MCC also occurs in the middle ear and eustachian tube that connects the middle ear with the nasal-sinus cavity. Motile cilia are essential for the vital defense mechanism of MCC to maintain sinus, middle ear, and lung health and function (Spassky and Meunier, 2017).

In addition to the respiratory tract, motile cilia are also present on the ependymal cells that line the brain ventricular system, which is composed of multiple cavities in the brain and a central canal that runs the length of the spinal cord (Ringers et al., 2020). The brain ventricles and the central canal are both filled with cerebrospinal fluid, which functions to deliver nutrients and metabolic signals to neurons, clear out toxic waste from the brain, and also serves as a fluid cushion for the brain (Spassky and Meunier, 2017). Cerebrospinal fluid is produced by epithelial cells called choroid plexuses that secrete proteins and transfer water and ions from blood capillaries into the lumen of the ependyma. The motile cilia found on ependymal cells produce the flow and facilitate circulation of the cerebrospinal fluid throughout the ventricular system to nourish and maintain homeostasis of the brain, as well as enable the production of neurons (Ringers et al., 2020). In humans, there are two major levels of cerebrospinal fluid flow: the macrofluidic, bulk flow among the ventricles that arise from pressure gradients, fluid exchanges with interstitial fluids, and changes in arterial pressure; and the microfluidic, near-wall flow produced by the motility of cilia on ependymal cells (Spassky and Meunier, 2017; Ringers et al., 2020). This microfluidic flow of cerebrospinal fluid driven by motile cilia is suggested to regulate the distribution of specific signaling molecules to distinct target areas of the brain. However, in smaller organisms, the motile cilia of the ependymal cells may be the main drivers of bulk cerebrospinal fluid flow, as hydrocephalus develops when ependymal motile cilia are dysfunctional in mice (Spassky and Meunier, 2017; Ringers et al., 2020). The ependymal cells of the central canal differ by the number of cilia from those found in the

brain, where ependymal cells in the central canal have fewer cilia. Two groups of ependymal cells in the central canal exist in humans where the number of cilia corresponds with the size of the central canal: multiciliated cells with 20 – 30 cilia are located laterally and mono- and bi-ciliated cells are positioned ventrally and dorsally. Ciliary motility in the central canal leads to circulation of the cerebrospinal fluid along the length of the spinal cord (Ringers et al., 2020). Studies in zebrafish have suggested this cerebrospinal fluid circulation in the central canal is important for straightening and maintenance of the body axis in spine development. Additionally, scoliosis can develop in humans with impaired cerebrospinal fluid circulation, which suggests that there is a conserved role for motile cilia-driven cerebrospinal fluid flow to mediate spine development in mammals (Ringers et al., 2020).

Multiciliated cells with motile cilia are also found in both the female and male reproductive tracts. In females, multiciliated cells are found along the length of epithelium in the oviduct, accounting for nearly 80% of the cells in the fimbriae, 30% in the isthmus, and up to 20% in the uterine cavity at the time of ovulation (Spassky and Meunier, 2017). Motility of cilia in the oviduct facilitates oocyte transport, along with muscle contractions and flow of tubal fluid, towards the uterus, where ciliary tips associate with the oocyte through a crown of glycocalyx (Lyons et al., 2006; Spassky and Meunier, 2017). Ciliary beat frequency increases after ovulation, triggered by follicular fluid, supporting the function of motile cilia in oocyte transport (Spassky and Meunier, 2017). Studies have found close interactions between sperm heads and cilia in human oviducts suggesting motile cilia helps with sperm transport as well (Lyons et al., 2006). It has been suggested that motile cilia present in the uterus are used to transport secretions from glandular and luminal secretory cells; however, their exact function remains unclear (Spassky and Meunier, 2017).

Multiciliated cells are found along the epithelium lining the efferent ducts in the male reproductive tract, where the proportion of ciliated cells progressively increases from the testis to the epididymis. The efferent ducts are known for reabsorbing fluid and ions, as well as stirring seminiferous fluid, to concentrate sperm by twenty-fold before these gametes reach the epididymis for further maturation and storage (Hoque et al., 2022). The seminiferous fluid with spermatozoa is pushed towards

7

the epididymis along a pressure gradient that is created by contractions, seminiferous secretions, and fluid absorption (Spassky and Meunier, 2007). Initially, it was thought that motile cilia in the efferent ducts transport spermatozoa towards the epididymis, similar to ciliary motility attributing to transport of ova or mucus in the oviducts or respiratory tract, respectively. However, novel findings revealed that the motile cilia in the efferent ducts beat with a rotary waveform in opposite directions dependent on their location in the efferent ducts. Thus, motile cilia in the efferent ducts have been suggested to stir and mix seminiferous fluid to prevent sperm aggregation (Spassky and Meunier, 2007; Hoque et al., 2022). Without functional motile cilia in the efferent ducts, seminiferous epithelium degradation and infertility can occur due to fluid retention caused by aggregated sperm resulting in blockages of the efferent ducts (Hoque et al., 2022). Additionally, the sperm flagellum is considered a specialized motile cilium with a similar axonemal structure to that of a respiratory motile cilium. Thus, morphological abnormalities that cause immobility of respiratory motile cilia can also lead to loss of sperm motility, resulting in male infertility (Hoque et al., 2022; Leigh et al., 2019).

Finally, motile cilia are found in a structure termed the node or left-right organizer that forms at the anterior tip of the primitive streak in the developing embryo (Little and Norris, 2021). The node is comprised of two different cell populations: central cells exhibiting a single motile monocilium on their apical surface and peripheral cells that only exhibit an immotile sensory monocilium. The motile nodal cilia are posteriorly tilted and rotate in a clockwise direction, instead of planar beating, with a CBF of approximately 10 Hz, which creates a leftward fluid flow called nodal flow (Little and Norris, 2021). The sensory nodal cilia on the periphery of the node will then sense this nodal flow. Three hypotheses have been proposed to explain how the sensory cilia sense this leftward nodal flow: morphogen, nodal vesicular parcel, or mechanosensation. The morphogen hypothesis suggests that nodal flow transports and causes signaling molecules to increase in abundance on the left side of the node. The nodal vesicular parcel hypothesis suggests that the signaling molecules are packaged in vesicles that break to deliver the contents on the left side of the node. These morphogens will then bind to transmembrane receptors in the sensory cilia membranes on the left side of the node and trigger an influx of calcium, leading to

8

downstream signaling events. In the mechanosensation hypothesis, immotile sensory cilia surrounding the node detect the flow through passive bending and this causes intracellular calcium influx more so on the left side of the node (Wallmeier et al., 2020; Little and Norris, 2021). The calcium signaling then causes asymmetric expression of multiple proteins, including downregulation of CERL2 and upregulation of WNT3, NODAL, LEFTY2, and PITX2, on the left side of the node, which will break symmetry of the developing embryo (Little and Norris, 2021). This symmetry breakage eventually will lead to the left-right asymmetry in the positioning of internal organs in all vertebrates, where the heart, spleen, and pancreas is consistently positioned on the left side of the body and the liver on the right (Fliegauf et al., 2007).

1.1.3 Structure of Motile Cilia

A typical human respiratory motile cilium has an average length of 7 microns and an average diameter of 0.1 - 0.3 microns (Bustamante-Marin and Ostrowski, 2017; Legendre et al., 2021). Cilia structures are divided into five different sub-compartments: the basal body, the transition zone, the axoneme, the ciliary tip, and the ciliary membrane (Figure 1.1). The basal body is a specialized centriole, with a 9 x 3 microtubular structure, surrounded by pericentriolar material with transition fibers at the distal end that dock it to the plasma membrane. As previously discussed, the basal body is used to anchor and nucleate the cilium. The transition zone is the area of the motile cilium where the triplet microtubular structure of the basal body is transformed into the axonemal structure. In the transition zone, transition fibers, or the distal appendages of the basal bodies, connect to the plasma membrane and Y-links connect each axonemal microtubule to the ciliary membrane to act as a ciliary gate controlling the entrance and exit of ciliary proteins into the axoneme. The axoneme is the core structural component of the cilia and is composed of singlet or doublet microtubules with associated proteins. The ciliary tip is the distal most portion of the cilia that contains the microtubule plus-ends and is the location where IFT switches from anterograde transport to retrograde transport. The ciliary membrane surrounds the transition zone, axoneme, and the ciliary lumen where it is continuous with the plasma membrane, and it contains signaling receptors for sensory functions (Fliegauf et al., 2007).

The majority of motile cilia have axonemes that are characterized with a 9 + 2 microtubular arrangement, where nine peripheral microtubule doublets (MTDs) surround two central, singlet microtubules called the central pair (Figure 1.1). The central pair, with its associated protein complexes, is often referred to as the central apparatus (CA). A microtubule doublet is comprised of one complete cylindrical microtubule (A-tubule) bound to an incomplete microtubule (B-tubule). Each MTD is linked to adjacent MTDs through dynein arms and nexin-dynein regulatory complexes (N-DRC). Radial spokes (RSs) connect each MTD to the central apparatus. Along each A-tubule are multiprotein motor complexes, called outer dynein arms (ODAs) or inner dynein arms (IDAs), that provide the force and regulation for cilia motility (Ishikawa, 2017). The dynein arms, RSs, and N-DRC are arranged along the length of the axoneme in 96-nm repeats. Within one 96-nm unit in human axonemes, there are four ODAs with 24-nm spacing, seven distinct IDA subtypes, one N-DRC, and three RSs (Bustamante-Marin and Ostrowski, 2017). In nodal motile cilia, the microtubular ultrastructure is characterized as 9 + 0, where these motile cilia lack the CA and the RSs (Ishikawa, 2017).

Each axonemal substructure is composed of multiple proteins and has different functions in regard to regulating ciliary motility. The RS is morphologically characterized by a T-shape with a stalk, a head, and a neck region. There have been three distinct RS types described within the 96-nm repeat in human respiratory cilia, where RS1 and RS2 have a similar structure while RS3 is physically different (Zhu et al., 2017). The CA is composed of two single microtubules (C1 and C2) that starts near the transition zone of motile cilia and various binding proteins form protrusions from and linkers between the microtubules (Loreng and Smith, 2017). The N-DRC is composed of at least 11 different proteins and has a bilobed structure that connects with an ODA, IDA *g* and *e*, and the base of RS2. The N-DRC includes the inner dynein arms and the nexin links that associate the outer MTDs together and limit the amount of sliding between MTD, which translates the doublet sliding into bending of the ciliary axoneme (Osinka et al., 2019). All of these structures help maintain the structural stability of the 9 + 2 axoneme throughout the cilium's continuous beating. Studies have shown that the CA, RSs, and the N-DRC, can also regulate dynein activity, microtubule sliding, and ciliary waveform through transduction pathways from the CA to

the RS and finally to the dynein arms. These transduction pathways involve second messengers, calcium

signaling, phosphorylation regulated by kinases and phosphatases attached to the axoneme, and physical

interactions between the CA, RS, and IDAs that produce mechanical feedback (Loreng and Smith, 2017;

Zhu et al, 2017; Osinka et al., 2019). The structure, function, and assembly of dynein arms are further

discussed in Chapter 2.
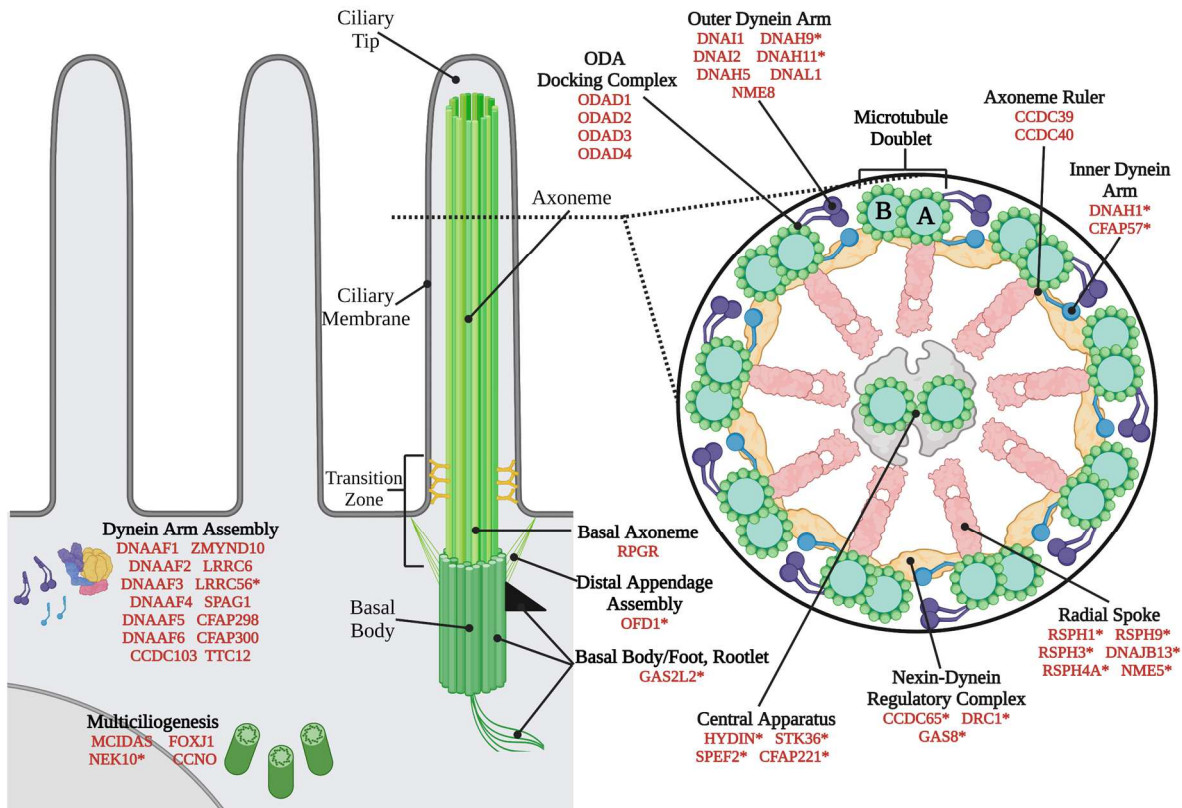
Figure 1.1 Motile cilia structure and associated PCD genes

Schematic showing a multiciliated cell with the five subcompartments of the motile cilium labeled: basal

body, transition zone, axoneme, ciliary membrane, and the ciliary tip. The basal body also includes the

basal foot and rootlet that helps to anchor the motile cilia and orients their planar waveform. The

transition zone includes the distal appendages, also known as transition fibers, and the Y-links that

connects the basal body and ciliary axoneme to the plasma and ciliary membrane, respectively (Fliegauf

et al., 2007). A cross-section of a motile ciliary axoneme reveals the 9 + 2 microtubular arrangement

found in respiratory motile cilia where nine peripheral MTDs surround a central pair of singlet

microtubules. Each MTD is connected to each adjacent doublet by N-DRCs, and each MTD is connected

to the CA by RSs. On each A-tubule of each MTD, ODAs and IDAs are present. Causative PCD genes

that are involved in dynein arm assembly, multiciliogenesis, or functioning as a component of the

structures found in the ciliary axoneme are listed in red text. Genes that are associated with ODAs or

ODA docking complexes result in ODA defects when defective in PCD individuals. Genetic variants in

genes that act as an axoneme ruler result in IDA and microtubular organization defects in PCD

individuals. When dysfunctional, genes that are part of the CA, N-DRC, or RS can result in defects in

these structures or appear normal on TEM in PCD-affected individuals. Mutations in proteins that are

involved in dynein arm assembly or multiciliogenesis results in ODA and IDA defects or oligocilia,

respectively. Genes marked with asterisks (*) can have normal TEMs on cilia cross-sections in PCD

subjects when defective (Leigh et al., 2019; Wallmeier et al., 2020; Legendre et al., 2021; Brennan et al.,

2021). Figure was created with BioRender.com.

1.2 Primary Ciliary Dyskinesia (PCD)

Dysfunction of the structure, regulation, assembly, or biogenesis of motile cilia, leading to

immotile or dyskinetic cilia, results in the rare, genetic disorder known as primary ciliary dyskinesia

(PCD), which is characterized by a clinical phenotype of chronic respiratory disease, infertility, and

laterality defects. While a subject with bronchiectasis and *situs inversus totalis* was first described in 1904

(Siewert, 1904), Manes Kartagener reported patients with a triad of bronchiectasis, chronic sinusitis, and

situs inversus totalis in 1933, which was named Kartagener syndrome (Kartagener, 1933). Once these

patients were identified to have immotile or dyskinetic respiratory cilia and sperm and to indicate the

genetic inheritance of the disease, the name was changed to primary ciliary dyskinesia in 1981 (Leigh et

al., 2019). PCD is considered a rare, orphan disease with an estimated prevalence of 1:7,000 to 1:20,000,

with the prevalence varying between countries and ethnicities. Of note, a high incidence of PCD can be

found in highly consanguineous populations (Wallmeier et al., 2020; Hannah et al., 2022).

Due to impaired mucociliary clearance in the upper and lower respiratory tract, PCD-affected individuals typically have chronic upper and lower respiratory disease. Interestingly, a common clinical characteristic of PCD is having neonatal respiratory distress despite full-term gestation, where this occurs in approximately 80% of PCD patients (Wallmeier et al., 2020). Almost all PCD individuals have a chronic, productive cough that is year-round and initially begins before the age of 6 months. They have recurrent lower respiratory tract infections that are most commonly caused by *Haemophilus influenza*, *Staphylococcus aureus*, *Streptococcus pneumonia* and *Moraxella catarrhalis*, as well as *Psuedomonas aeruginosa* with a prevalence that increases with age. PCD-affected individuals also have reduced lung function that worsens with increasing age, and almost all affected individuals (>80%) develop progressive bronchiectasis by adulthood (Leigh et al., 2019). Bronchiectasis is the irreversible dilation and damage of bronchi due to inflammation, mucus build-up, and deterioration of the elastic connective tissue. In the most severe cases of PCD, bronchiectasis and progressive airway disease can result in end-stage respiratory failure. Due to impairment of mucociliary clearance in the upper respiratory tract, the majority of PCD patients deal with chronic nasal congestion and recurrent sinusitis. In up to 15% of PCD patients, nasal polyps can occur. Recurrent otitis media also occurs in majority of pediatric cases of PCD, which can lead to hearing loss and speech and language delay (Leigh et al., 2019).

In addition to respiratory disease, dysfunction of the motile cilia found on embryonic nodal cells can result in organ laterality defects and congenital heart disease in PCD-affected individuals. These laterality defects can range from *situs inversus totalis*, which is a complete mirror-image reversal of visceral organs, to situs ambiguous, in which there is a partial mispositioning of visceral organs, and heterotaxy syndromes. *Situs inversus totalis* occurs in approximately 50% of PCD patients, while situs ambiguous or heterotaxy occurs in about 6 - 14% of all patients (Leigh et al., 2019; Wallmeier et al., 2020). Situs ambiguous can include the laterality defects known as situs inversus thoracalis, situs inversus abdominalis, left isomerism (polysplenia), and right isomerism (asplenia) (Fliegauf et al., 2007). Congenital heart disease has been observed in 6 – 17% of all individuals with PCD, but is much more common in cases of situs ambiguous (Lengendre et al., 2021). Interestingly, there also appears to be a

higher incidence of scoliosis (5 - 10% of patients) and pectus excavatum (up to 10% of patients) in PCD-affected individuals compared to the general population (Leigh et al., 2019). Dysfunction of the sperm flagella and motile cilia in the efferent ducts in PCD males typically results in infertility; however, some males with PCD can conceive naturally. Dysfunction of the motile cilia in the female oviducts may lead to subfertility and an increased risk of ectopic pregnancy in females with PCD (Leigh et al., 2019; Wallmeier et al., 2020). While the dysfunction of ependymal cilia in the brain ventricles leads to hydrocephalus in multiple PCD model organisms, hydrocephalus is a rare occurrence in humans with PCD (Niziolek et al., 2022).

1.3 PCD Diagnostics and Treatment

PCD is significantly underdiagnosed and can often be misdiagnosed for disorders of similar phenotypes. There is currently not a single diagnostic test with a high specificity and sensitivity for PCD. Therefore, in order to diagnose individuals with PCD, a wide range of tests are employed when a PCD diagnosis is suspected based on clinical history and/or phenotype (Leigh et al., 2019). Ciliary biopsies are usually taken from scraping the inferior nasal turbinate or a brush biopsy of the nasal turbinate or bronchial mucosa during a bronchoscopy to obtain respiratory epithelium. These ciliary biopsies are then analyzed through various methods to examine the ultrastructure and functionality of the motile cilia. Transmission electron micrographs (TEMs) of cilia cross-sections are used to analyze the ultrastructure of the motile cilia and note if there are any defects in the outer dynein arms, the inner dynein arms, the central apparatus, the radial spokes, and/or microtubular organization. Of note, about 30% of PCD-affected individuals have a normal ciliary ultrastructure as examined on TEMs, so the absence of ciliary defects on TEMs cannot be used alone to rule out a PCD diagnosis (Leigh et al., 2019). Immunofluorescence staining of ciliated cells using antibodies targeting ciliary proteins can be used to supplement TEM findings, where mutant proteins or structures can be shown to have altered expression or be mislocalized in a PCD multiciliated cell (Shapiro et al., 2016). Evaluation of ciliary movement by light microscopy and high-speed video microscopy analysis (HSVMA) to examine any reduced ciliary

beat and altered waveform is also performed; however, there are multiple limitations to this method, and it should not be used as a diagnostic test for PCD alone (Leigh et al., 2019).

Several studies have shown that nitric oxide, a small diffusible gas molecule, within the nasal cavity and sinuses is, interestingly, very low in PCD compared to healthy individuals. Thus, measurement of nasal nitric oxide (nNO), performed by using a small nasal probe to analyze nasal gas using a chemiluminescence or hand-held electrochemical device, is commonly used as a diagnostic tool for PCD. Clinicians and experts have established a diagnostic cut-off value of less than 77 nL/min for PCD. However, false positives in nNO measurements can occur due to recent viral infection, acute sinusitis, sinus bleeding, and nNO measurements are not recommended or considered reliable in children under the age of five. Therefore, repeat testing of nNO measurements, at least one month apart, in suspected PCD-affected individuals over the age of five is recommended (Leigh et al., 2019). Other disorders that commonly have low nNO, including cystic fibrosis and primary immunodeficiencies, also need to be ruled out before using nNO as the primary diagnostic test for PCD (Leigh et al., 2019; Brennan et al., 2021). Finally, the most accurate diagnostic test for PCD is genetic testing. With approximately 50 genes currently associated with causing PCD when mutated, several commercial panels have been generated and utilized to test for the majority of these known genes (Wallmeier et al., 2020). However, only 70% of PCD-affected individuals have a genetically-confirmed diagnosis, suggesting that a negative genetic test result should not rule out a PCD diagnosis. Thus, additional PCD-causative genes will most likely be identified through further research to improve PCD diagnostics (Leigh et al., 2019).

There is currently no cure for PCD and therapies that are empirically-tested specifically for PCD are very limited. The majority of therapeutics for PCD are based on and extrapolated from studies in cystic fibrosis and non-cystic fibrosis bronchiectasis. A recent publication from the PCD Foundation provides recommendations for monitoring and treatment of PCD, where management aims to maintain respiratory function and minimize progression through early diagnosis, airway clearance, and treatment of respiratory infections (Shapiro et al., 2016; Leigh et al., 2019). Monitoring the progression of PCD lung disease includes regular sputum cultures, pulmonary function spirometry testing, and imaging.

Monitoring and treatment of otolaryngology symptoms in PCD individuals includes audiology

assessments, hearing aids, placement of pressure equalization tubes, and management of rhinosinusitis.

Management of sinus disease includes nasal steroids, daily nasal saline rinses, and courses of systemic

antibiotics, as well as endoscopic sinus surgery. Airway clearance therapies, which facilitates mucus

clearance, includes high-frequency chest compression using vest therapy, manual chest physiotherapy,

positive pressure expiratory devices, active cycle breathing, autogenic drainage, postural drainage, and

exercise (Leigh et al., 2019). Inhaled medications for PCD can include bronchodilators, inhaled

corticosteroids in patients with asthma, and nebulized hypertonic saline, which enhances airway hydration

and induces cough to increase cough clearance. Systemic and inhaled antibiotics are used to treat

respiratory tract infections, which is directed by identification of infection-causing microbes using sputum

cultures. Lobectomy is only reserved for PCD individuals with severe localized bronchiectasis and

hemoptysis, and lung transplantation is used for individuals who reach end-stage lung disease (Shapiro et

al., 2016; Leigh et al., 2019).

1.4 PCD Genetics

PCD is a heterogenous, genetic disorder that is usually inherited in an autosomal recessive

manner. However, there have been a few genetic variants associated with PCD that are X-linked,

including *PIH1D3*, *RPGR*, and *OFD1*, and de novo autosomal dominant mutations have been recently

discovered in *FOXJ1* in a few cases of PCD (Wallmeier et al., 2020). Since the discovery of the first

PCD-associated gene, *DNAI1*, in 1999, there has been a massive advancement in identifying genes

associated with PCD (Leigh et al., 2019). To date, there has been approximately 50 different causative

genes that have been identified and associated with PCD (Figure 1.1; Table 1.1). These genes are usually

classified based on the location of the protein or the ciliary defect that is present when they are mutated,

including ODA defects, IDA and microtubular disorganization defects, ODA and IDA defects, central

apparatus organization defects, and reduced generation of multiple motile cilia (RGMC) (Leigh et al.,

2019; Wallmeier et al., 2020).

Table 1.1 Genetics of Primary Ciliary Dyskinesia[1]

Gene Name	Aliases	TEM Defects	Ciliary Beat Defects	Laterality Defects	Nasal Nitric Oxide[2]
Outer Dynein Arm and Docking Complex					
DNAI1		ODA	Reduced, minimal movements	Yes	Low
DNAI2		ODA	Reduced, minimal movements	Yes	NR
DNAH5		ODA	Immotile; Stiff	Yes	Low
DNAH9		Subtle ODA	Hypokinetic; Reduced distal bending	Yes	Low
DNAH11		Normal	Hyperkinetic	Yes	Low
DNAL1		ODA	Immotile; Weak	Yes	Low
NME8	*TXNDC3*	ODA	Normal	Yes	NR
ODAD1	*CCDC114*	ODA	Immotile; Flickering	Yes	Low
ODAD2	*ARMC4*	ODA	Flickering	Yes	Low
ODAD3	*CCDC151*	ODA	Immotile	Yes	NR
ODAD4	*TTC25*	ODA	Immotile; Flickering	Yes	Low
Axoneme Ruler					
CCDC39		IDA + MTD	Immotile	Yes	Low
CCDC40		IDA + MTD	Immotile; Stiff	Yes	Low
Inner Dynein Arm					
DNAH1		Normal	NR	Yes	NR
CFAP57	*WDR65*	Normal	Reduced; Symmetrical waveform	NR	Low
Radial Spokes					
RSPH1		Normal; CA	Reduced bending angle	NR	Low; Normal
RSPH3		Normal; CA	Reduced bending angle	NR	Low
RSPH4A		Normal; CA	Rotational pattern; Stiff	NR	Low; Normal
RSPH9		Normal; CA	Rotational pattern; Stiff	NR	Low; Normal
DNAJB13		Normal; CA	Reduced amplitude	NR	Low
NME5		Normal; CA	NR	NR	NR
Central Apparatus					
HYDIN		Normal	Immotile; Rigid; Rotational	NR	Low; Normal
STK36		Normal; CA	Stiff	NR	Normal
SPEF2		Normal	Stiff; Rotational pattern	NR	Low
CFAP221	*PCDP1*	Normal	Rotational pattern	NR	Normal
Nexin-Dynein Regulatory Complex					
CCDC65	*DRC2*	Normal	Stiff; Hyperkinetic	NR	Low
CCDC164	*DRC1*	Normal	Stiff	NR	Low
GAS8	*DRC4*	Normal	Normal; Subtle beat abnormality	NR	Low; Normal
Dynein Arm Assembly					
DNAAF1	*LRRC50*	ODA + IDA	Immotile	Yes	NR
DNAAF2	*KTU*	ODA + IDA	Immotile	Yes	Low
DNAAF3		ODA + IDA	Immotile	Yes	Low
DNAAF4	*DYX1C1*	ODA + IDA	Immotile	Yes	Low
DNAAF5	*HEATR2*	ODA + IDA	Immotile	Yes	Low
DNAAF6	*PIH1D3*	ODA + IDA	Immotile	Yes	Low
ZMYND10	*DNAAF7*	ODA + IDA	Immotile	Yes	Low
LRRC6	*DNAAF11*	ODA + IDA	Immotile	Yes	Low
LRRC56	*DNAAF12*	Normal	Immotile; Stiff; Twitching	Yes	Low
SPAG1	*DNAAF13*	ODA + IDA	Immotile	Yes	Low
CFAP298	*DNAAF16*	ODA + IDA	Immotile	Yes	Low
CFAP300	*DNAAF17*	ODA + IDA	Immotile	Yes	Low
CCDC103		ODA; ODA + IDA	Immotile; Normal	Yes	Low; Normal
TTC12		IDA	Normal; Immotile; Reduced beating angle	NR	Low; Normal

Multiciliogenesis and Orientation				
MCIDAS	Oligocilia	Immotile	NR	Low
CCNO	Oligocilia	Hypokinetic	NR	Low
FOXJ1	Oligocilia	Normal	Yes	Normal
NEK10	Normal	Normal; Shortened cilia	NR	Normal
GAS2L2	Normal	Hyperkinetic; Uncoordinated	NR	Low; Normal
RPGR	Oligocilia	Normal; Uncoordinated	NR	Normal
OFD1	Normal	Normal; Immotile; Stiff	Yes	Low

Abbreviations: TEM, transmission electron micrographs; ODA, outer dynein arm defect; IDA, inner dynein arm defect; IDA + MTD, inner dynein arm and microtubular disorganization defect; CA, central apparatus defect; ODA + IDA, outer and inner dynein arm defect; NR, not reported in literature.
[1]Table is modified from Brennan et al., 2021 and Zariwala et al., 2007.
[2]Nasal nitric oxide is considered low if <77 nL/min (Leigh et al., 2013).

1.4.1 ODA Defects

Absence or truncation of ODAs alone on cilia TEMs are usually due to mutations or defects in structural components of the ODA and/or the ODA docking complex. Genetic variants have been identified in components of the ODA, including *DNAI1*, *DNAI2*, *DNAH5*, *DNAH9*, *DNAH11 DNAL1*, and *NME8* (*TXNDC3*), that result in PCD. More than 30% of PCD individuals have mutations in two genes that cause ODA defects, *DNAI1* and *DNAH5* (Leigh et al., 2019). Genetic variants have also been identified in components of the ODA docking complex, including *ODAD1* (*CCDC114*), *ODAD2* (*ARMC4*), *ODAD3* (*CCDC151*), and *ODAD4* (*TTC25*), in PCD-affected individuals. Cilia with defects in the ODA or ODA docking complex are mostly immotile or have a reduced ciliary beat frequency. One dynein heavy chain, *DNAH11*, however, actually results in normal ciliary ultrastructure on TEMs, due to the limitations of this method to detect the absence of the proximal type 1 ODA. Due to the motility defect being limited to the proximal cilia, genetic variants in *DNAH11* leads to a hyperkinetic cilia (Legendre et al., 2021).

1.4.2 IDA + MTD Defects

Another major ultrastructure defect seen in PCD ciliary axonemes is absent or truncated IDA and microtubular disorganization. Of note, an isolated IDA defect on TEMs of ciliary axonemes is not a typical ultrastructure defect and is frequently a secondary change in cilia due to damage or infection of ciliated cells. Unless an IDA alone defect occurs on repeated biopsies, this should not be considered

diagnostic (Leigh et al, 2019). The IDA and microtubular disorganization defects are caused by genetic variants in the coiled-coil domain containing proteins, *CCDC39* and *CCDC40*. These proteins act as a "ruler," designating the attachment sites of the IDAs and N-DRCs in 96-nm repeats along the length of the ciliary axoneme. Mutations in these two proteins leads to defects in both the N-DRC and IDA and disruption of the 96-nm repeat along the ciliary axoneme, which results in a reduced amount of beating cilia that are rigid with a severely reduced beat amplitude (Leigh et al., 2019; Legendre et al., 2021). These causative genes account for 12% of all PCD cases, and these PCD-affected individuals typically have greater lung disease based on spirometry and CT scans, as well as poorer nutritional status compared to other PCD patients with other ultrastructural defects (Leigh et al., 2019). While genetic variants have been identified in an inner dynein arm heavy chain, *DNAH1*, in PCD-affected individuals, TEMs were not used to examine their cilia, but it is expected that these patients would have a near-normal ultrastructure (Brennan et al., 2021). Additionally, *CFAP57* has been identified as a causative gene for PCD, where it localizes to the axoneme and facilities with the assembly of IDA; however, these patients have what appears to be normal TEMs with reduced ciliary beat frequency and an altered ciliary waveform (Bustamante-Marin et al., 2020).

1.4.3 CA/RS/N-DRC Defects

Defects in structural proteins of the CA, the RS, and the N-DRC can lead to either CA defects (missing central pair, central pair off-center, transposition of outer doublet into central position), microtubular disorganization, and/or subtle defects in ciliary ultrastructure that appear normal using TEMs (Leigh et al, 2019; Brennan et al., 2021). Components of the central apparatus that have been identified as causative PCD genes are *HYDIN*, *STK36*, *SPEF2*, and *CFAP221* (*PCDP1*). Components of the radial spoke head, *RSPH1*, *RSPH4A*, and *RSPH9*, and components of the radial spoke stalk, *DNAJB13*, *NME5*, and *RSPH3*, have also been identified as causative genes of PCD (Legendre et al., 2021). These genes when dysfunctional have been suggested to lead to subtle abnormalities in the ciliary beat frequency or waveform of motile ciliary, specifically being stiff with a rotational pattern or a reduced bending angle (Brennan et al., 2021). Interestingly, due to the lack of the central pair apparatus in nodal

cilia, PCD-affected individuals with mutations in the central pair complex and radial spokes do not have situs abnormalities. Additionally, milder respiratory disease with normal levels of nitric oxide have also been reported in some PCD patients with *RSPH1* mutations (Leigh et al., 2019). Genetic components of the N-DRC that have been identified as PCD causative genes are *CCDC65*, *DRC1* (*CCDC164*), and *GAS8*. These PCD-affected individuals have near normal TEMs with absent N-DRC links or misaligned outer MTDs in some TEMs of cilia cross-sections and subtle ciliary beat abnormalities that range from normal to stiff or hyperkinetic (Brennan et al., 2021; Legendre et al., 2021).

1.4.4 Reduced Generation of Multiple Motile Cilia (RGMC) and Orientation Defects

Mutations in genes that regulate ciliogenesis in multiciliated cells have also been identified in PCD-affected individuals. MCIDAS and CCNO are essential for massive centriole amplification through the deuterosome-dependent pathway in multiciliogenesis. Thus, genetic variants in *MCIDAS* and *CCNO* result in a significant reduction in basal bodies, and thus cilia per cell. FOXJ1, another multiciliogenesis transcription factor, is vital for basal body docking at the apical membrane of multiciliated cells. Therefore, mutations in *FOXJ1* lead to a reduction of correctly assembled motile cilia, but not due to a disruption in centriole amplification. The normal number of basal bodies is present but mislocalized in the cytoplasm of *FOXJ1*-deficient cells, suggesting this ciliary defect is due to the inhibition of basal body docking. TEMs of cilia samples from these PCD patients show very few cilia, and this defect is referred to as ciliary aplasia or reduced generation of multiple motile cilia (RGMC) (Wallmeier et al., 2020). Another ciliogenesis gene that has been implicated in PCD is *NEK10*, which encodes a ciliated cell-specific kinase that is involved in ciliogenesis by regulation of the motile ciliary proteome. Studies analyzing the defects in these *NEK10*-deficient PCD subjects suggest that the motile cilia are shorter, but the number of cilia, their ultrastructure, and ciliary beat is normal (Wallmeier et al., 2020; Legendre et al., 2021).

Additionally, there are proteins that localize at the base of the cilium, either at the transition zone, the basal body, or are centrosomal, that have been implicated in causing PCD when dysfunctional. *GAS2L2* localizes to the basal body, the basal feet, rootlets, and actin filaments in motile multiciliated

cells. Dysfunctional *GAS2L2* has been shown to cause defects in ciliary orientation, with a normal TEM ultrastructure and hyperkinetic motile cilia, that results in PCD (Bustamante et al., 2019). Two genes that are known for being causative genes for non-motile ciliopathies have also been implicated in causing PCD: *RPGR* and *OFD1*. *RPGR* is an X-linked gene that codes for a protein that localizes to the retinal rod photoreceptor or the transition zone at the base of respiratory cilia and causes retinitis pigmentosa when dysfunctional. *RPGR* genetic variants have also been found in PCD-affected individuals with retinitis pigmentosa, due to the fact that it plays a role in maintaining cilia stability and regulating cilia orientation outside of its photoreceptor function. PCD patients with *RPGR* genetic variants can have respiratory symptoms with near normal TEMs and a cilia orientation defect (Leigh et al., 2019; Wallmeier et al., 2020). Additionally, genetic variants in *OFD1* leads to a X-linked syndromic ciliopathy, called oral-facial-digital type 1 syndrome. OFD1 is a centrosomal protein that is vital for SHH and canonical Wnt signaling pathways in multiple organ systems. Patients with OFD1 have respiratory symptoms consistent with PCD, normal ciliary ultrastructure on TEMs, and have a lower number of cilia, shorter cilia, mislocalized basal bodies, and dyskinetic ciliary beating with disturbed cilia orientation (Leigh et al., 2019; Wallmeier et al., 2020).

1.4.4 ODA + IDA Defects

Finally, another common ciliary ultrastructural defect found in PCD is the absence or truncation of both the ODAs and the IDAs. The first gene associated with this defect was identified in 2008 to be *DNAAF2* (*KTU*) (Omran et al., 2008). Since then, several causative PCD genes have been identified to cause this ODA and IDA defect, including *DNAAF1* (*LRRC50*) (Loges et al., 200), *DNAAF3* (Mitchison et al., 2012), *DNAAF4 (DYX1C1)* (Tarkar et al., 2013), *DNAAF5 (HEATR2)* Horani et al., 2012), *DNAAF6* (*PIH1D3*) (Olcese et al., 2017), *DNAAF7* (*ZMYND10)* (Moore et al., 2013), *DNAAF11* (*LRRC6*) (Kott et al., 2012), *DNAAF12* (*LRRC56*) (Bonnefoy et al., 2018), *DNAAF13* (*SPAG1*) (Knowles et al., 2013), *DNAAF16* (*CFAP298/C21orf59*) (Austin-Tse et al., 2013), and *DNAAF17* (*CFAP300/C11orf70*) (Höben et al., 2018; Braschi et al., 2022). Mutations in these proteins lead to completely immotile cilia. Due to defects in both dynein arms and the localization of these proteins only

in the cytoplasm of multiciliated cells, instead of the ciliary axoneme, these proteins are hypothesized to be involved in the assembly of the dynein arms in the cytoplasm before their transport into the ciliary axoneme by IFT (Legendre et al., 2021). Additionally, mutations in *CCDC103 (*putative *DNAAF19)* have been shown to either cause an ODA and IDA defect or an ODA alone defect. *CCDC103* is thought to be an axonemal protein that acts as an attachment factor of dynein arms and is independent of the dynein arms and their docking complexes (Panizzi et al., 2012). Interestingly, another PCD-causative gene, *TTC12*, leads to both ODA and IDA defects in sperm flagella, but only single headed IDA defects in respiratory cilia when dysfunctional, suggesting that dynein arm assembly mechanisms could be different between motile cilia and sperm flagella (Thomas et al., 2020).

CHAPTER 2: SPAG1'S POTENTIAL ROLE IN DYNEIN ARM ASSEMBLY AND OTHER FUNCTIONS

2.1 Dynein Arms

Dynein arms are the large multiprotein motor complexes associated with the MTDs along the length of axonemes of motile cilia. Dyneins are approximately 1 - 2 MDa, and in humans, they are composed of ~520 kDa dynein heavy chains (DHCs), ~70 to 140 kDa dynein intermediate chains (DICs), ~53 to 59 kDa dynein light-intermediate chains (DLICs) and several ~10 to 30 kDa dynein light chains (DLCs) (King, 2016; Braschi et al., 2022). There are two main types of dynein arms, inner dynein arms (IDAs) and outer dynein arms (ODAs), dependent on their location on the A-tubules of each MTD. Within each main type, there are subtypes based on the number of DHCs and various dynein chain proteins that comprise the dynein arm. There are two types of ODAs found in respiratory motile cilia, the proximal type 1 ODA and the distal type 2 ODA, that are "double-headed" with two DHCs. There are seven distinct subtypes of inner dynein arms with one double-headed dynein *Il/f* and six single-headed dyneins *a*, *b*, *c*, *d*, *e*, and *g*. As previously stated, within each 96-nm repeat of the motile cilia axoneme, there are four ODAs and one set of inner dynein arms, which includes, in order of localization along the axoneme, dynein *f/I1*, *a*, *b*, *c*, *e*, *g*, and *d* (Figure 2.1). The ODA are known for providing the force for microtubule sliding, and thus regulating CBF, and the IDA are known for mediating bend formation and waveform (King, 2016; Bustamante-Marin and Ostrowski, 2017). The structure and function of dynein arms and how they allow cilia to move is described further below.
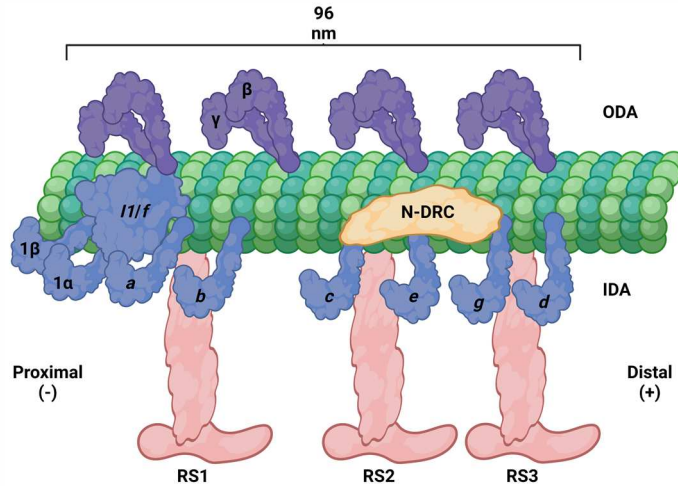
Figure 2.1 96-nm axoneme repeat consists of four ODAs and one set of IDAs

Schematic showing a longitudinal 96-nm repeat of the A-tubule in a motile ciliary axoneme that contains

four outer dynein arms (ODAs) and a single set of inner dynein arms (IDAs), which includes the double-

headed dynein *I1/f* and single-headed dyneins *a*, *b*, *c*, *e*, *g*, and *d*. Each 96-nm repeat also contains a

nexin-dynein regulatory complex (N-DRC) and three radial spokes (RSs) in human respiratory cilia.

Figure was created with BioRender.com.

2.1.1 Structure of Dynein Arms

Dynein arms are composed of one or two DHCs, a few DICs or DLICs and several DLCs, where

the dynein heavy chains are the motor unit. The structure of an ~450 kDa DHC includes four distinct

domains: the amino-terminal tail that binds to cargo, the stalk region that associates with the microtubule,

a linker region, and the AAA+ ATPase head domain (Figure 2.2). The motor head domain is composed of

a hexameric ring of distinct AAA+ domains numbered 1 – 6 (King, 2016; Bustamante-Marin and

Ostrowski, 2017). AAA1 through AAA4 subunits bind ATP, where AAA1 is the main site for ATP

hydrolysis. The microtubule-binding domain of the dynein is located at the end of coiled-coil stalks that

originate from AAA4 or AAA5, where this stalk is called the buttress. The linker region, which is an N-

terminal α-helical bundle, stretches across the AAA ring and facilitates motor activity. The N-terminal tail

associates with and provides a platform for binding of the DICs, the DLICs, and the DLCs (King, 2016;

Canty et al., 2021). In humans, there are 15 genes that encode for dynein heavy chains, where 13 of them specifically code for heavy chains of the axonemal dyneins (Braschi et al., 2022).
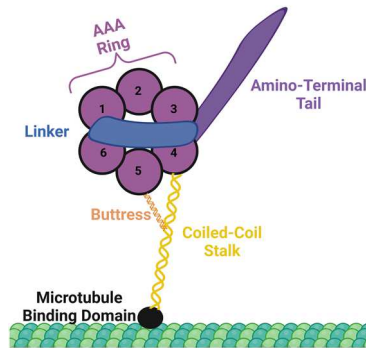
Figure 2.2 Dynein heavy chain structure

Schematic demonstrating the structure of the dynein heavy chain. The protein contains an AAA+ ATPase ring that is the motor unit, a linker region that spans across the AAA+ ring, a coiled-coil stalk and buttress that binds the dynein to microtubules through its microtubule binding domain, and an amino-terminal tail that binds to any cargo. Figure created with BioRender.com.

One group of dynein arms, which makes up the monomeric IDAs, consist of one DHC that is associated with actin and either the light chain DNALI1 or centrin. Dyneins *a*, *c*, and *d* are composed of actin and DNALI1, while dyneins *b*, *e*, and *g* are composed of actin and centrin. (Figure 2.3C). TTC29 and ZMYND12 are two additional subunits of dynein *d*. Each of these single-headed dynein arms have distinct DHC proteins, where the exact correspondence between the human and *Chlamydomonas* orthologues is unclear (King, 2016; Braschi et al., 2022). It is estimated that DNAH7 is present in IDA *a*, DNAH3 is present in IDA *b*, DNAH12 is present in IDA *c*, DNAH1 is present in IDA *d*, and DNAH6 is present in IDA *g* (Yamamoto et al., 2020). There are also minor dyneins in the ciliary axoneme of *Chlamydomonas* located at the ciliary base which suggests they are involved in bend initiation (King, 2016).

The second group of dynein arms are double-headed, consisting of two DHC motors, and includes both the ODAs and the IDA named dynein *I1* or *f* (Figure 2.3A, 2.3B). The two DHC motors in a double-headed dynein arm interact together through their amino-terminal tails; however, these two DHCs are distinct from each other. The DHCs also use their amino-terminal tails to bind to another subcomplex, termed the dynein intermediate chain/light chain (DIC/LC) complex. This DIC/LC complex is comprised of two or three dynein intermediate chains and several dynein light chains of three conserved classes: LC8-type family (DYNLL1, DYNLL2, DNAL4), Tctex1-type family (DYNLT1, DYNLT2, DYNLT2B, and DYNLT3), and the roadblock-type family (DYNLRB1 and DYNLRB2) (King, 2016; Braschi et al., 2022). In looking at the specific protein composition of each double-headed dynein arm, the proximal type 1 ODAs are composed of two DHCs, DNAH5 and DNAH11, which are orthologous to the *Chlamydomonas* γ-HC and β-HC, respectively. In the distal type 2 ODAs, DNAH11 is replaced with DNAH9. All ODAs contain the two DICs, DNAI1 and DNAI2, and several different DLCs (DNAL1, DNAL4, DYNLL1, DYNLL2, DYNLRB1, DYNLRB2, DYNLT1/TCTEX1, DYNLT2, and/or NME8/DNAI8/TXNDC3) (Braschi et al., 2022). In human sperm flagella, there is reported to be a type 3 ODA that is comprised of DNAH8 and DNAH17, instead of the main respiratory dynein heavy chains DNAH5, DNAH9, or DNAH11 (Whitfield et al., 2019). The double-headed IDA *I1/f* is comprised of two heavy chains (DNAH2 and DNAH10), three intermediate chains (DNAI3, DNAI4, and DNAI7) and several light chains (DYNLL1, DYNLL2, DYNLRB1, DYNLRB2, DYNLT3, DYNLT2B), compared to the single-headed IDAs that are associated with only actin, centrin, and/or DNALI1 (Canty et al., 2021; Braschi et al., 2022). The ODAs also require the ODA docking complex for their attachment to MTDs, and this complex has been suggested to be facilitate the regulation of ODA activity. The ODA docking complex is composed of five protein subunits in mammals: *ODAD1* (*CCDC114*), *ODAD2* (*ARMC4*), *ODAD3* (*CCDC151*), *ODAD4* (*TTC25*), and *ODAD5* (*CLXN*) (Braschi et al., 2022). The different dynein arms, their protein compositions, and *Chlamydomonas* orthologues are listed in Table 2.1 (Figure 2.3).
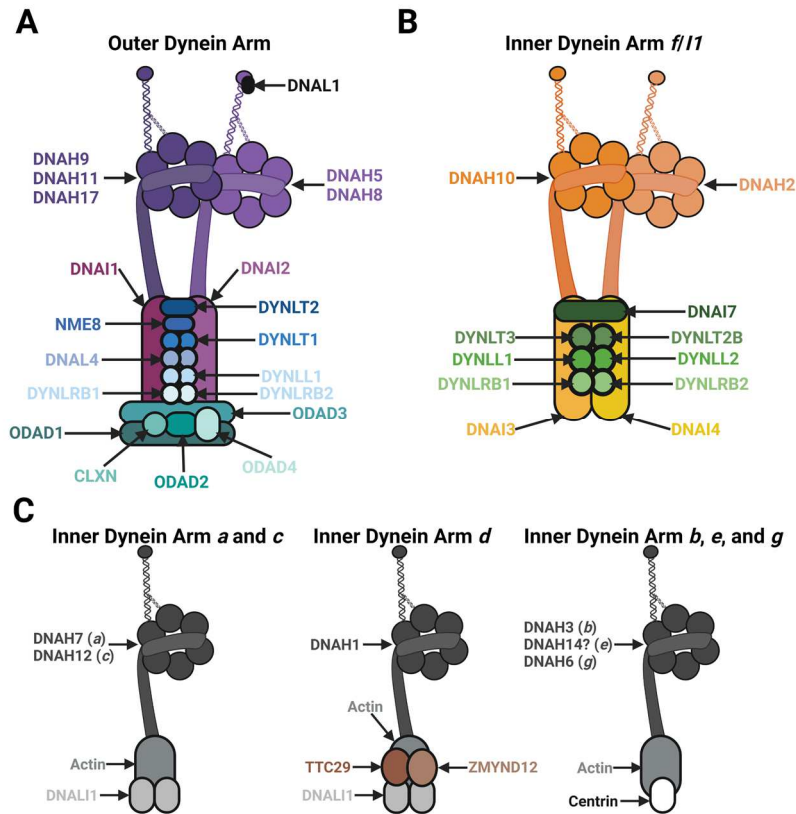
Figure 2.3 Protein composition of the axonemal dynein arms

(A) Schematic of the double-headed ODA and its protein composition with its dynein heavy chains (purple), dynein intermediate chains (magenta), dynein light chains (blue), and the docking complex (teal) shown. **(B)** Schematic of the double-headed IDA *I1/f* and its protein composition with its dynein heavy chains (orange), dynein intermediate chains (yellow), and dynein light chains (green) shown. **(C)** Schematic of the monomeric IDAs *a*, *b*, *c*, *d*, *e*, and *g*. IDAs *a*, *c*, and *d* are composed of a dynein heavy chain (dark grey), actin (grey), and DNALI1 (light grey), with IDA *d* containing additional proteins TTC29 and ZMYND12 (brown). IDAs *b*, *e*, and *g* are composed of a dynein heavy chain (dark grey), actin (grey), and centrin (white). Figure was made from information provided in Braschi et al., 2022. Figure created with BioRender.com.

Table 2.1 Protein Composition of the Outer Dynein Arms and the Inner Dynein Arms[1]

Symbol	Name	Aliases	*Chlamydomonas* Ortholog
Outer Dynein Arm Subunits			
DNAI1	Dynein axonemal intermediate chain 1	DIC1	*DIC1* (IC1)
DNAI2	Dynein axonemal intermediate chain 2	DIC2	*DIC2* (IC2)
DNAH5	Dynein axonemal heavy chain 5		*DHC15* (ɣ HC)
DNAH9	Dynein axonemal heavy chain 9		*DHC14* (β HC)
DNAH11	Dynein axonemal heavy chain 11		*DHC14* (β HC)
DNAH8	Dynein axonemal heavy chain 8		*DHC15* (ɣ HC)
DNAH17	Dynein axonemal heavy chain 17		*DHC14* (β HC)
DNAL1	Dynein axonemal light chain 1		*DLU1* (LC1)
DNAL4	Dynein axonemal light chain 4		*DLL3* (LC10)
DYNLL1	Dynein light chain LC8-type 1	DNCL1; LC8	*DLL1* (LC8)
DYNLL2	Dynein light chain LC8-type 2	DNCL1B	*DLL1* (LC8)
DYNLRB1	Dynein light chain roadblock-type 1	DNCL2A; ROBLD1	*DLR1* (LC7a)
DYNLRB2	Dynein light chain roadblock-type 2	DNCL2B; ROBLD2	*DLR2* (LC7b)
DYNLT1	Dynein light chain Tctex-type 1	TCTEL1; TCTEX1	*DLT1* (LC9); *DLT3* (Tctex1)
DYNLT2	Dynein light chain Tctex-type 2	TCTE3; TCTEX2	*DLT2* (LC2)
NME8	NME/NM23 family member 8	DNAI8; TXNDC3	*DLX1*?
ODAD1	Outer dynein arm docking complex subunit 1	CCDC114	*DCC2* (ODA1) and *DCC3* (ODA5)
ODAD2	Outer dynein arm docking complex subunit 2	ARMC4	
ODAD3	Outer dynein arm docking complex subunit 3	CCDC151	*DCC1* (ODA-DC1, ODA3) and *ODA10* (ODA10)
ODAD4	Outer dynein arm docking complex subunit 4	TTC25	
ODAD5	Outer dynein arm docking complex subunit 5	CLXN; EFCAB1	
Inner Dynein Arm I1/f Subunits			
DNAI3	Dynein axonemal intermediate chain 3	WDR63; DIC3	*DIC3* (IC140)
DNAI4	Dynein axonemal intermediate chain 4	WDR78; DIC4	*DIC4* (IC138)
DNAI7	Dynein axonemal intermediate chain 7	CFAP94; CASC1	*DII6* (FAP94; IC97)
DNAH2	Dynein axonemal heavy chain 2		*DHC10* (1β HC)
DNAH10	Dynein axonemal heavy chain 10		*DHC1* (1α HC)
DYNLL1	Dynein light chain LC8-type 1	DNCL1; LC8	*DLL1* (LC8)
DYNLL2	Dynein light chain LC8-type 2	DNCL1B	*DLL1* (LC8)
DYNLRB1	Dynein light chain roadblock-type 1	DNCL2A; ROBLD1	*DLR1* (LC7a)
DYNLRB2	Dynein light chain roadblock-type 1	DNCL2B; ROBLD2	*DLR2* (LC7b)
DYNLT2B	Dynein light chain Tctex-type 2B	TCTEX1D2	DLT4 (Tctex2b)
DYNLT3	Dynein light chain Tctex-type 3	TCTEX1L	*DLT1* (Tctex1)

Monomeric Inner Dynein Arm Subunits			
DNALI1 (IDA *a*, *c*, *d*)	Dynein axonemal light intermediate chain 1	P28	*DII1* (p28)
ACTA1	Actin alpha 1, skeletal muscle	ACTA; NEM3	
CETN1 (IDA *b*, *e*, *g*)	Centrin 1	CEN1; CETN	*DLE2*
CETN2 (IDA *b*, *e*, *g*)	Centrin 2	CEN2; CALT	*DLE2*
CETN3 (IDA *b*, *e*, *g*)	Centrin 3	CEN3	*DLE2*
ZMYND12 (IDA *d*)	Zinc finger MYND-type containing 12		*DII2* (p38)
TTC29 (IDA *d*)	Tetratricopeptide repeat domain 29		*DII3* (p44)
DNAH1 (IDA *d*)	Dynein axonemal heavy chain 1		*DHC2* (DHC2)
DNAH3 (IDA *b*)	Dynein axonemal heavy chain 3		*DHC4* (DHC4)
			DHC5 (DHC5)
			DHC6 (DHC6)
			DHC8 (DHC8)
			DHC9 (DHC9)
			DHC11 (DHC11)
DNAH6 (IDA *g*)	Dynein axonemal heavy chain 6		*DHC2* (DHC2)
DNAH7 (IDA *a*)	Dynein axonemal heavy chain 7		*DHC4* (DHC4)
			DHC5 (DHC5)
			DHC6 (DHC6)
			DHC8 (DHC8)
			DHC9 (DHC9)
			DHC11 (DHC11)
DNAH12 (IDA *c*)	Dynein axonemal heavy chain 12		*DHC4* (DHC4)
			DHC5 (DHC5)
			DHC6 (DHC6)
			DHC8 (DHC8)
			DHC9 (DHC9)
			DHC11 (DHC11)
DNAH14 (IDA *e*?)	Dynein axonemal heavy chain 14		*DHC4* (DHC4)
			DHC5 (DHC5)
			DHC6 (DHC6)
			DHC8 (DHC8)
			DHC9 (DHC9)
			DHC11 (DHC11)

[1]Information provided and organized from Yamamoto et al., 2020 and Braschi et al., 2022.

2.1.2 Function of Dynein Arms

Dyneins are motor proteins that travel down microtubules toward the minus-ends. Dyneins accomplish this through cycles of conformational changes, caused by ATP hydrolysis in the AAA+ ring of dynein, that disengages the dynein from the microtubule and changes the conformation of the linker region to generate a step along the microtubule. This process is known as the mechanochemical cycle where chemical energy is repeatedly converted to mechanical work. In state 1 of this process, AAA1 is

nucleotide-free (apo state), the linker region is in the straight conformation, and dynein is associated with the microtubule. In state 2 and 3, ATP binds to AAA1 and the dynein detaches from the microtubule, allowing the linker to move across the AAA+ ring to another conformational state. In state 4, ATP undergoes hydrolysis, the linker region transforms to a bent conformation, and this shifts the microtubule binding domain further down the microtubule towards the minus end. In state 5, the dynein rebinds to the microtubule through its microtubule binding domain. In state 6, the inorganic phosphate (Pi) is released from AAA1 and the linker changes to the straight conformation which produces the powerstroke that allows the dynein to pull its cargo in the forward direction. After ADP release, the dynein returns to the initial nucleotide-free state to prepare to repeat the cycle (Canty et al., 2021).

Cilia motility is generated by the inner and outer dynein arms, that are permanently associated with the A-tubule of one MTD through their DHC N-terminal tails, stepping along the B-tubule of an adjacent MTD through their microtubule binding domains in an ATP-dependent manner. These steps along the MTDs create a sliding force that is limited due to the N-DRCs linking the MTDs together, which thus causes this sliding force to be converted into bending of the ciliary axoneme. For an axonemal bend to create and propagate a ciliary wave, sequential activation or deactivation of the dynein arms have to occur typically from the base of the cilia to the ciliary tip (King, 2016). A few models for how activation or inactivation of dynein arms leads to bend formation have been proposed. One is known as the switch point model, where dyneins are active on one side of the axoneme while those on the opposite side are inactive. Recently, the majority of ODAs were found in a pre-powerstroke or active state in portions of cilia that were straight, which suggests that dyneins are sequentially inactivated to initiate and propagate a ciliary wave. This model is referred to as the switch inhibition hypothesis (Antony et al., 2021; Canty et al., 2021). Additionally, there are three proposed models that involve mechanical feedback as a mechanism to regulate the coordinated activation and deactivation of dynein arms for bending of the cilium, which include the geometric-clutch model, the sliding-control model, and the curvature-control model. The geometric clutch model suggests that bending of the axoneme causes MTDs to be pushed further apart or brought closer together that inhibits the dyneins from binding to the microtubules or

increases the ability of dyneins to bind to the microtubules, respectively. The sliding-control model

suggests that nexin links create resistant forces that limit the degree of microtubule sliding. However,

once a certain threshold is reached, this resistance results in the abrupt disassociation of dynein arms from

the microtubules, which allows for microtubule sliding on the opposite side of the axoneme. The

curvature-control model suggests that dyneins can sense the curvature of the MTDs for their regulation,

where they are either compressed on the inner surface of bent microtubules or stretched on the outer

surface of bent microtubules. A combination of these models for the regulation of dynein activity could

occur together to produce a ciliary wave, and thus, may not be mutually exclusive (Canty et al., 2021).

Additionally, two models of mechanosensory mechanisms providing regulation of the activation and

deactivation of dynein arms have been proposed. The first model suggests that the CA and RSs produce

signals that regulate the IDAs, and the other affects the ODAs through the light chain DNAL1 that is

associated with DNAH5 through its microtubule-binding domain (King, 2016). The propagation of

axonemal bending eventually yields a ciliary wave with an effective stroke, which propels fluids in the

forward direction, and a recovery stroke, which returns the cilium to its starting position, within the same

plane (Bustamante-Marin and Ostrowski, 2017).

Both the ODA and IDA are capable of generating force for cilia movement. The presence of

ODAs increase force output by the cilia by fourfold, which suggests they are the main drivers of ciliary

motility and regulate the ciliary beat frequency. The CBF can be regulated by the levels of Ca^{2+} or by

alterations in cellular redox state detected by light chains or intermediate chains, respectively (King,

2016). Meanwhile, the IDAs provide regulation and propagation of signals to create specific cilia

waveforms. Dynein activation or deactivation on individual MTDs is controlled by signals that originate

from the CA, RSs, and the N-DRC and are transmitted to IDAs through physical interactions (King,

2016). IDAs were also revealed to generate torque, which may contribute to the ciliary waveform (Canty

et al., 2021). The phosphorylation status of the DNAI4 intermediate chain of the IDA *I1/f* also controls

the IDAs. Phosphorylation by casein kinase 1 or protein kinase A causes slower microtubule sliding while

dephosphorylation by protein phosphatase 1 and/or protein phosphatase 2A leads to faster gliding (King, 2016).

2.2 Axonemal Dynein Arm Assembly

While the protein composition and the regulation of axonemal dynein arms have been extensively studied, the process in which these multiprotein motor complexes are assembled is poorly understood. Due to the lack of protein production machinery in cilia, protein structures that are present in the ciliary axoneme, which includes dynein arms, have to be translated, folded, and assembled in the cytoplasm. This process, termed axonemal dynein arm assembly, involves various dynein components that are assembled to form distinct, stable subcomplexes (i.e., DIC/LC complex including DNAI1-DNAI2 and light chains for ODAs) that then interact with folded heavy chains and regulatory light chains to produce a complete dynein arm. The axonemal dyneins are then transported from the cytoplasm into the ciliary axoneme and to their proper, pre-assembled docking complex (ODAD1-4 for ODAs and potentially CCDC39/CCDC40 for IDAs) and attachment factor (CCDC103 for ODAs) by IFT machinery (Figure 2.4). Due to their massive size, time to synthesize, and potential for protein aggregation, multiple cytoplasmic factors, termed dynein axonemal assembly factors (DNAAFs), have evolved in multiciliated cells that are required for the production and assembly of dynein arms (King, 2016). These factors, which only localize to the cytoplasm of multiciliated cells, were discovered through identification of causative genes associated with PCD, which cause defects in both the ODAs and the IDAs and completely immotile cilia (Fabczak and Osinka, 2019). Recently, it has been suggested that dynein arm assembly takes place in dynein axonemal particles (DynAPs) generated by liquid-liquid phase separation in the cytoplasm of multiciliated cells, where dynein subunits are steadily present in these organelles while DNAAFs and chaperones fluctuate in and out (Huizar et al., 2018; Lee et al., 2020). However, others argue that dynein assembly occurring in these observed membrane-less cytoplasmic particles might be due to the sheer number of DNAAFs and chaperones needed and the time required to translate large DHC mRNAs and associate them with other dynein chains to create these dynein complexes (King, 2021). While several DNAAFs have been identified, how they interact with each other and the mechanisms in which they

facilitate the assembly of axonemal dyneins is largely unknown (Figure 2.4). Further information on their interactions and what is known for how they contribute to dynein arm assembly is given for each DNAAF below.
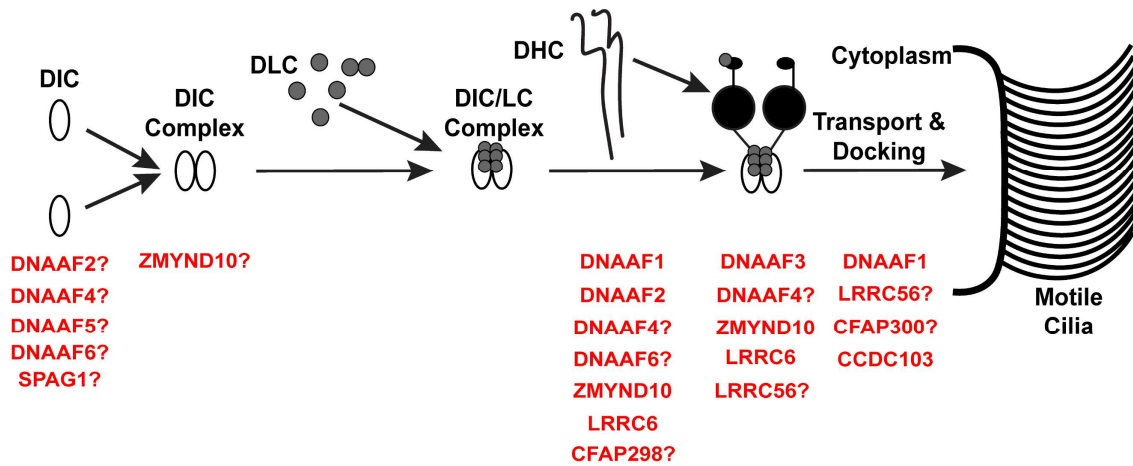
Figure 2.4 Axonemal dynein arm assembly and possible roles of DNAAFs

Schematic showing the process of axonemal dynein arm assembly occurring in the cytoplasm of respiratory multiciliated cells. Dynein intermediate chains (DICs) are folded and stabilized in a complex together (i.e., DNAI1-DNAI2 for ODAs). Dynein light chains (DLCs) are then folded and associated with the DIC complex to form the DIC/LC complex. Finally, dynein heavy chains (DHCs) (i.e., DNAH5, DNAH9, or DNAH11) are folded and bound to the DIC/LC complex to form a functional dynein arm before its transport into the ciliary axoneme and docking to its appropriate docking complex or attachment site. Dynein arm assembly factors (DNAAFs) that are known PCD causative genes in humans are listed in red text underneath the particular step in dynein arm assembly for which there is evidence of them participating in or facilitating that stage of assembly. These different steps can include folding of the DICs, forming the DIC complex, folding the DHCs, associating the DHCs to the DIC/LC complex or stabilizing and maturation of the dynein arm, and/or transport and docking of dynein arms into the ciliary axoneme. Some DNAAFs are repeatedly listed at different stages of dynein assembly due to evidence of the DNAAF participating in multiple steps or lack of evidence to pinpoint a particular stage in either early

or late dynein assembly. Evidence to categorize which stage in dynein assembly these DNAAFs participate in include expression data, DNAAF protein interaction studies, measurement of dynein chain levels, and dynein arm component protein interaction measurements.

2.2.1 Dynein Arm Assembly Factors (DNAAFs)

DNAAF1 (LRRC50) is a 725-amino acid protein with a N-terminal leucine rich repeat-containing (LRRC) domain (a.a. #107 – 242), a coiled-coil domain (a.a. #321 – 341), and C-terminal proline-rich domain (a.a. #391 – 513) (Fabczak and Osinka, 2019) (Figure 2.5). *Chlamydomonas oda7* mutants, which are deficient in the orthologue DNAAF1, have reduced stability of dynein heavy chains and association of dynein heavy chains to dynein intermediate chains that results in impaired ODA assembly and decreased flagellar motility (Fowkes and Mitchell, 1998). *DNAAF1* mutants in mice and *Danio rerio* also have typical phenotypes associated with motile cilia defects, including hydrocephalus, laterality defects, body axis curvature, heart looping defects, or sinusitis, due to disruption of ODA and IDA assembly (Ha et al., 2016; Hartill et al. 2018). In humans, PCD-affected individuals with *DNAAF1* mutations have completely immotile cilia due to the loss of both ODA and IDA, as shown by TEMs and immunofluorescence of their respiratory cilia (Loges et al., 2009). Outside of PCD, *DNAAF1* mutations have been associated with seminomas and testicular cancers, as well as neural tube defects and scoliosis, in zebrafish and humans (Basten et al., 2013; Miao et al., 2016; Wang et al., 2020). Both RUVBL1 and RUVBL2, two AAA+ ATPases that form a hexamer together that are involved in protein folding and multiprotein complex formation, interact with DNAAF1 (Hartill et al., 2018). DNAAF1 has also been suggested to be involved in the folding of the DHCs, due to their decreased abundance and sensitivity to proteases in *Chlamydomonas oda7* mutants (Mitchison et al., 2012). It has also been shown that DNAAF1 interacts with an intraflagellar transport protein, IFT88, suggesting DNAAF1 might also play a role in loading IFT trains with dynein arm cargo or intraflagellar transport itself (Hartill et al., 2018). Thus, DNAAF1 is necessary for motile cilia function and involved in dynein arm assembly and IFT most likely by

facilitating the folding of DHCs through a complex with RUVBL1 and RUVBL2 and loading of cargo onto IFT trains.

DNAAF2 (KTU) was the first causative PCD gene that was discovered to be a DNAAF (Omran et al.,2008). It is an 837-amino acid protein with a N-terminal protein interacting with heat shock protein 90 (PIH1) domain (a.a. #10 – 182) and a CHORD-containing proteins and SGT1 (CS) domain (a.a. # 254 – 349) (Fabczak and Osinka, 2019) (Figure 2.5). In *Chlamydomonas* mutants of the *DNAAF2* orthologue*, pf13*, the flagella are completely immotile due to defects in both ODA and IDA. *Dnaaf2* was identified to be mutated in medaka fish (*Oryzias latipes*) with heart looping defects, body axis curvature, and infertility due to impaired ciliary and sperm motility. In respiratory cilia of PCD-affected individuals with *DNAAF2* genetic variants, there is an absence of ODAs and IDAs, demonstrated by TEMs and immunocytochemistry, that leads to immotile cilia (Omran et al., 2008). In *Danio rerio* mutants deficient in DNAAF2, there was defective ciliary motility in the Kupffer's vesicle due to the loss of ODAs and IDAs *b*, *c*, and *e* (Yamaguchi et al., 2018). Co-immunoprecipitations, GST-pull down assays, and mass spectrometry analysis has identified interactions between DNAAF2 with DNAI2 and HSP70 (Omran et al., 2008). DNAAF2 has also been shown to interact with DNAAF4 (DYX1C1), and these two proteins have been suggested to form a R2TP-like complex that associates with heat shock protein 90 (HSP90) and 70 (HSP70) (Omran et al., 2008; Tarkar et al., 2013; Maurizy et al., 2018). Additionally, DNAAF2 was reported to interact and co-localize in DynAPs with DNAAF5 (HEATR2) and SPAG1 in a preassembly complex in the early steps of airway epithelial cell differentiation, before the expression of other DNAAFs (Horani et al., 2018; Fabczak and Osinka, 2019). In *Chlamydomonas pf13* mutants, the dynein heavy chains demonstrated a significant reduction in their abundance, while the intermediate chains increased, and DHCs association with a pre-assembled DIC/LC complex was disrupted in cytoplasmic extracts (Omran et al., 2008). DNAAF2 is required for ODA and IDA heavy chain folding and stability as shown by protease sensitivity assays in *Chlamydomonas* mutants (Mitchison et al., 2012). Thus, DNAAF2 functions in dynein arm assembly by folding DHCs with the help of other DNAAFs, possibly in distinct complexes with DNAAF4 and heat shock proteins or DNAAF5 and SPAG1.

DNAAF3 is a protein with a total of 590 amino acids and has two domains DUF4470 and

DUF4471 of unknown function (Figure 2.5). Genetic variants causing loss-of-function of DNAAF3 were

first identified in PCD-affected individuals with ODA and IDA defects in immotile respiratory cilia by

Mitchison and colleagues (Mitchison et al., 2012). *DNAAF3* knockdown in zebrafish also causes dynein

arm defects, resulting in defective ciliary motility phenotypes including hydrocephalus and laterality

anomalies. *Chlamydomonas pf22* mutants, which lack the DNAAF3 ortholog, have defects in assembly of

both ODAs and some IDAs resulting in immotile flagella. Based on PF22's cytoplasmic location and the

increased protease sensitivity of DHCs in *pf22* mutant cells, it was hypothesized that DNAAF3 functions

at a later step in axonemal dynein assembly, specifically in stabilization or maturation of the DHCs by

binding to the DIC/LC complex (Mitchison et al., 2012). Additionally, RNAi depletion and CRISPR-

generated *DNAAF3* mutants in *Drosophila* exhibit loss of both ODAs and IDAs that leads to immotile

cilia and defective ciliary phenotypes of auditory deficiency and male infertility. Comparison between

wild-type and *DNAAF3* mutants by proteomics suggest a reduction in DHCs of all dynein subtypes,

corresponding with the hypothesis that DNAAF3 is involved in DHC stabilization or maturation (zur

Lage et al., 2021).

DNAAF4 (DYX1C1) has 420 amino acids with a N-terminal CS domain (a.a. #3 – 87) and a C-

terminal tetratricopeptide repeat (TPR) domain with three TPR motifs (a.a. #290 – 399) (Figure 2.5)

(Fabczak and Osinka, 2019). Mice and zebrafish that have mutations in *DNAAF4* have dysfunctional

motile cilia phenotypes, which include hydrocephalus, laterality defects, kidney cysts, and defective body

axis curvature. These phenotypes are due to cilia with reduced motility caused by the loss of both ODAs

and IDAs. In humans, the loss of DNAAF4 leads to disrupted ODA and IDA assembly, as shown by

TEM and immunofluorescence of ODA and IDA dynein chains, and results in immotile cilia and thus

PCD (Tarkar et al., 2013). Mass spectrometry and co-immunoprecipitation studies of the interactome of

DNAAF4 has suggested that DNAAF4 interacts with HSP70, HSP90, and the chaperonin CCT (Tarkar et

al., 2013). Yeast two-hybrid and co-immunoprecipitation studies demonstrating interactions between

DNAAF4 and either DNAAF2 or PIH1D3 suggest that HSP cochaperone complexes are formed (Omran

et al., 2008; Maurizy et al., 2018; Fabczak and Osinka, 2019). However, it is not clear if this co-chaperone complex formed with DNAAF4, DNAAF2 or PIH1D3, and HSPs facilitates the folding of DICs or DHCs in axonemal dynein assembly. However, DNAAF4's interaction with ZMYND10, which is a DNAAF necessary for the stabilization of DNAI1 and DNAI2, may suggest it is involved in the folding and stabilization of DICs (Cho et al., 2018).

DNAAF5 (HEATR2) is an 855 amino acid protein with ten HEAT repeat domains spanning the length of the protein (Figure 2.5) (Fabczak and Osinka, 2019). PCD-affected individuals with *DNAAF5* mutations have defective ODAs and IDAs and completely immotile cilia, and this phenotype has been recapitulated in *Chlamydomonas* and *Drosophila* mutant models (Horani et al., 2012; Diggle et al., 2014). Using immunoprecipitation studies, DNAAF5 was shown to interact with DNAI2, but not HSP70 or HSP90 (Diggle et al., 2014). It has been demonstrated that DNAAF5 is expressed during early airway epithelia differentiation, and DNAAF5 also interacts with SPAG1 directly and DNAAF2 indirectly in a complex that is involved in axonemal dynein assembly at an earlier step than other DNAAFs. DNAAF5 has also been shown to co-localize in DynAPs with DNAAF2 and SPAG1 (Horani et al., 2018; Huizar et al., 2018; Lee et al., 2020). Mutations in *DNAAF5* do not lead to dissociation between DNAAF5 and DNAAF2 or SPAG1, but does result in the accumulation of aggregates of the DNAAF5, DNAAF2, and SPAG1 complex which will eventually be degraded (Horani et al., 2018; Fabczak and Osinka, 2019). Thus, it is possible that DNAAF5 is involved in dynein arm assembly by forming a complex with DNAAF2 and SPAG1 that facilitates the formation of the DICs or the DIC complex, independent of HSP70 or HSP90.

DNAAF6 (PIH1D3) is one of the X-linked genes associated with causing PCD. It is a protein with a total of 214 amino acids and a N-terminal PIH1 domain (a.a. #1 – 111) and a C-terminal CS domain (a.a. #111 – 208) (Figure 2.5). It is one of the four proteins present in humans that is part of the PIH1 domain-containing protein family, other than DNAAF2, PIH1D1 and PIH1D2 (Fabczak and Osinka, 2019). In *DNAAF6*-deficient PCD-affected individuals, respiratory cilia had reduced ciliary beat frequency due to consistent reduction of ODAs and IDA *I1/f, g,* and *d* measured by electron tomography

averaging. Knock out mice of the *DNAAF6* pseudogene, which is only expressed in the testis, were infertile, with no obvious other phenotype, and had immotile sperm due to defects in ODAs and IDAs (Dong et al., 2014). Additionally, *Danio rerio DNAAF6* mutants had immotile cilia that lead to the development of pronephric cysts, abnormal body axis curvature, and reversed heart looping caused by defects in ODAs, IDA *c*, IDA *g*, and IDA *d* (Olcese et al., 2017; Yamaguchi et al., 2018). DNAAF6 can interact with DNAI2, HSP70, and/or HSP90 as demonstrated by co-immunoprecipitation studies, and mutations in DNAAF6's CS domain can interrupt its interaction with DNAI2 (Dong et al., 2014; Olcese et al., 2017). It has also been proposed that DNAAF6 can form an HSP70 or HSP90 co-chaperone complex with DNAAF4 (DYX1C1), suggesting it is involved in the folding or interaction of dynein chains (Olcese et al., 2017; Maurizy et al., 2018; Fabczak and Osinka, 2019). While decreases in both dynein intermediate chains and heavy chains have been reported in *DNAAF6* mouse mutants, it was shown that the DIC complex can still form by co-immunoprecipitation studies, suggesting that the proposed DNAAF6 co-chaperone complex is more likely involved in stabilizing the DIC complex, the folding of the DHCs, and/or the interaction of the DHCs with the DIC/LC complex (Dong et al., 2014).

DNAAF7 (ZMYND10) is a 440 amino acid protein with three LxxLL motifs (a.a. #4 – 8, #215 – 223, and #266 – 270) and a Mynd-type zinc finger domain (a.a. #394 – 430) at its C-terminus (Figure 2.5) (Fabczak and Osinka, 2019). In humans, genetic variants in *DNAAF7* lead to completely immotile cilia and a typical PCD phenotype due to absence of both ODAs and IDAs in ciliary axonemes as shown by TEMs and immunofluorescence (Zariwala et al., 2013; Moore et al., 2013). In zebrafish and medaka fish morphants of *DNAAF7*, significantly reduced ciliary beating or immotile cilia and sperm led to common defective cilia phenotypes, including pronephric cysts with dilated kidney tubules, heart looping defects, scoliosis, and infertility (Zariwala et al., 2013; Kobayashi et al, 2017). Additionally, a knockout model of *DNAAF7* in *Drosophila* had uncoordinated locomotion and infertility due to motility defects of cilia in Ch neurons and sperm caused by absence of ODAs and IDAs (Moore et al., 2013). Finally, *DNAAF7* knockout mouse models have defects in dynein arms and immotile cilia that result in PCD phenotypes, including hydrocephalus, laterality defects, and mucus plugging in the upper airways (Cho et al., 2018;

Mali et al., 2018). Co-immunoprecipition assays and GST-pull down experiments have demonstrated that DNAAF7 can interact with LRRC6 through its CS domain (Zariwala et al., 2013; Moore et al., 2013). Additional interactors of DNAAF7, including CFAP298, DNAAF4, RUVBL2, HSP70, HSP90, FKBP8, DNAI1, DNAH5, DNAH17, and DYNLT5, have been revealed by co-immunoprecipitation, GST-pull down, and AP-MS studies (Cho et al., 2018; Mali et al, 2018). Cycloheximide chase assays to study protein stability showed that ZMYND10 helped stabilize the LRRC6 protein, as well as the DIC DNAI1 which in turn stabilized DNAI2 when bound in the DIC complex (Cho et al., 2018). In situ proximity ligation assays showed a reduction in the interaction between DNAI2 and DNAH5, while co-immunoprecipitations showed that DNAI1 and DNAI2 could still interact in the absence of DNAAF7 in murine cells (Mali et al., 2018). Interestingly, label-free quantitative proteomic analysis demonstrated that all dynein heavy chains were reduced in *DNAAF7* mutant mice (Mali et al., 2018). Thus, it was recently proposed that DNAAF7 can form a HSP90 co-chaperone complex with FKBP8 to facilitate the folding and maturation of DHCs (Mali et al., 2018). Due to the multitude of protein interactions discovered for ZMYND10 and the evidence for it being necessary for the folding and stabilization of both DICs and DHCs, it is possible ZMYND10 plays a role in transferring or relaying dynein chain clients to multiple chaperone complexes.

DNAAF11 (LRRC6) is a protein with a total of 466 amino acids and a N-terminal LRRC domain with six LRR repeats (a.a. #23 – 146), a coiled-coil domain (a.a. #178 – 204), and a C-terminal CS domain (a.a. #332 – 381) (Figure 2.5) (Fabczak and Osinka, 2019). Cilia and sperm were found to be immotile due to the absence of ODAs and IDAs in *LRRC6* mutant mice with hydrocephalus and laterality defects, as well as in PCD-affected individuals with deficient *LRRC6* (Kott et al., 2012; Horani et al., 2013). Through co-immunoprecipitation, yeast two-hybrid, and GST pull down assays, LRRC6 has been shown to directly interact with RUVBL2 and DNAAF7, most likely through its CS domain (Moore et al., 2013; Zariwala et al., 2013; Zhao et al., 2013; Fabczak and Osinka, 2019). Additionally, co-immunoprecipitations for DNAI2 in *LRRC6* mutant mice testes show that the formation of the DIC complex for ODAs (i.e, DNAI1-DNAI2) does not require LRRC6 (Inaba et al., 2016). Taken together,

these data suggests that LRRC6 interacts with DNAAF7 (ZMYND10) to facilitate dynein arm assembly, possibly at a stage involving the dynein heavy chains.

DNAAF16 (CFAP298/C21orf59) was identified as a DNAAF, and a causative PCD gene, through an *in vivo* zebrafish morpholino knockdown screen. Knockdown of *CFAP298* in *Danio rerio* and planarian *Schmidtea mediterranea* disrupts ODA assembly, and mutations in *CFAP298* in humans results in PCD with defects in both outer and inner dynein arms (Austin-Tse et al., 2013). CFAP298 has only been shown to interact with DNAAF7 (ZMYND10) (Cho et al., 2018). DNAAF17 (CFAP300/C11orf70) has also been identified as a causative PCD gene, with genetic variants resulting in a typical PCD phenotype with immotile cilia due to loss of inner and outer dynein arms (Fassad et al., 2018). CFAP300 has also been shown to interact with DNAAF2 (Höben et al., 2018). RNAi knockdown of *CFAP300* in *Paramecium* reduced cilia beating, altered ciliary waveform, and reduced swimming velocity as a result of the loss of both outer and inner dynein arms. It was also discovered that CFAP300 localization is mostly in the cytoplasm in *Paramecium* and *Chlamydomonas*; however, there was a small amount detected in the ciliary axoneme. When retrograde IFT was defective, CFAP300 was found to accumulate at the ciliary tip, suggesting it could be involved in both dynein arm assembly or IFT of mature axonemal dyneins (Fassad et al., 2018).

Genetic variants in *DNAAF12* (*LRRC56*) have been identified in PCD-affected individuals with dyskinetic cilia but had no detectable axonemal defects on TEMs (Bonnefoy et al., 2018). Interestingly, however, *LRRC56* mutants in *Trypanosoma brucei* or *Chlamydomonas oda8* mutants showed outer dynein arm defects, more so in the distal axoneme, with abnormal ciliary beat patterns and reduced swimming velocities. In these model organisms, it was shown that LRRC56 is partially localized in the axoneme and cytoplasm, and LRRC56 interacts with the IFT protein IFT88, suggesting that LRRC56 is more involved in IFT of mature dynein arms than the cytoplasmic preassembly of axonemal dyneins (Desai et al., 2015; Bonnefoy et al., 2018). The final DNAAF is DNAAF13 (SPAG1), which is the focus of this project, and its protein domains, interactions, and proposed functions are described in section 2.3.
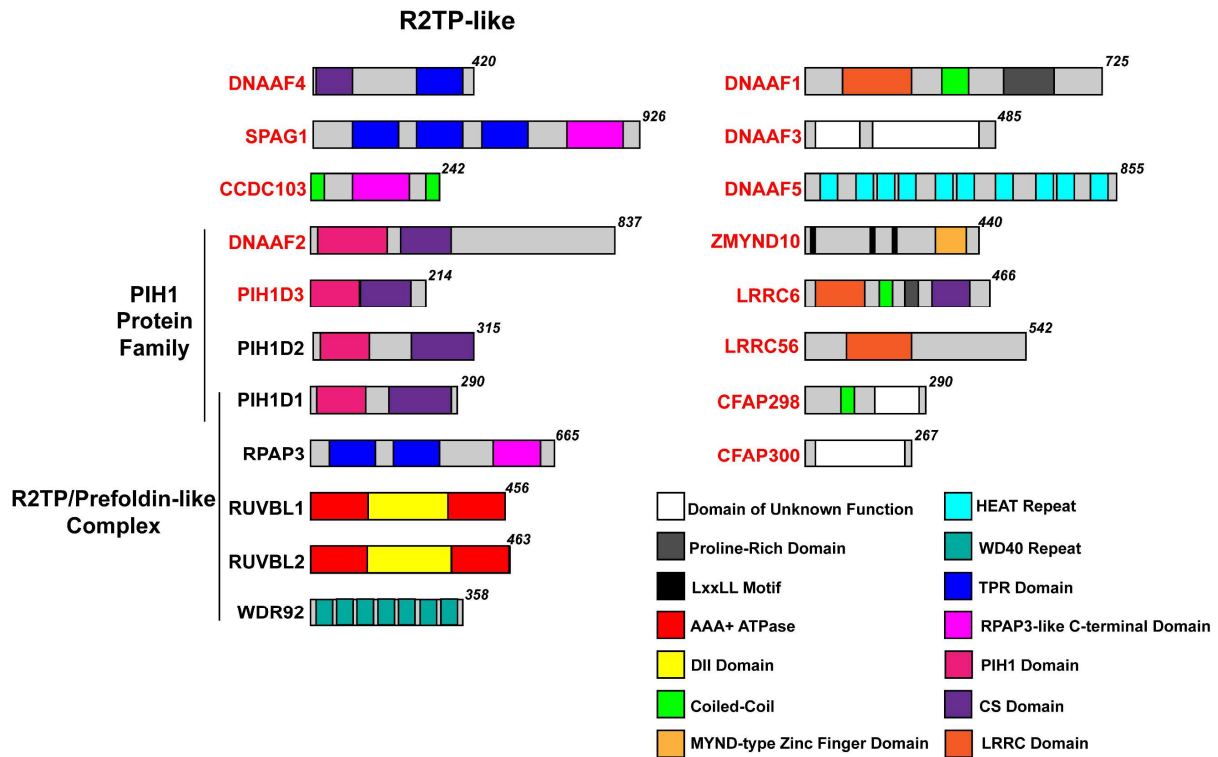
Figure 2.5 DNAAFs and their conserved protein domains

Schematic showing the different dynein arm assembly factors (DNAAFs) and their associated conserved

protein domains that are color-coded. Several DNAAFs identified in other model organisms are canonical

components of the heat shock protein 90 co-chaperone complex, R2TP/Prefoldin-like complex. These

canonical components include PIH1D1, RPAP3, RUVBL1, RUVBL2, and WDR92. Several other

DNAAFs have similar protein domains, including TPR, PIH1, CS, and RPAP3 C-terminal domains, that

are present in this R2TP/Prefoldin-like complex. These DNAAFs have been proposed to form R2TP-like

complexes with other DNAAFs to facilitate the assembly of axonemal dyneins. DNAAFs listed in red

text are those that have been associated with causing PCD in humans when mutated.

2.2.2 Potential Chaperones and the R2TP complex

  Molecular chaperones are proteins that facilitate the folding of newly synthesized proteins,

refolding of misfolded proteins, and assembly of multiprotein complexes. Multiple DNAAFs have been

shown to interact with different chaperones that include heat shock protein 70 (HSP70), heat shock

protein 90 (HSP90), and the chaperonin containing TCP-1 (CCT) complex. HSP70 and HSP90 have been

shown to interact with DNAAF2 (KTU), DNAAF4 (DYX1C1), DNAAF6 (PIH1D3), and DNAAF7

(ZMYND10). HSP90 has been shown to additionally interact with DNAAF1 (LRRC50) (Fabczak and

Osinka, 2019). These interactions may suggest that DNAAFs interact with chaperones to either recruit the

dynein chains to the chaperones or facilitate the folding of the dynein chains in the cytoplasm.

HSP70 is known to be involved in the early stages of protein folding, while HSP90 is involved in

the late stages of protein folding and the formation of quaternary structures of molecular complexes

(Houry et al., 2018). Interestingly, a vital HSP90 co-chaperone complex, known as the R2TP/Prefoldin-

like complex, has been proposed to associate with all three chaperones, including HSP70, HSP90, and the

CCT complex, that have been shown to interact with DNAAFs. The R2TP/Prefoldin-like complex, also

called the PAQosome, facilitates the folding, interactions, and formation of several multiprotein

complexes involved in cell growth, gene expression, and protein synthesis, including L7Ae

ribonucleoproteins, U5 small nuclear ribonucleoproteins, phosphatidylinositol-3-kinase-related kinases,

and RNA polymerases (Houry et al., 2018). The R2TP complex in humans is composed of a hexamer of

two AAA+ ATPases, RUVBL1/Pontin and RUVBL2/Reptin, associated with a heterodimer of RPAP3

and PIH1D1. The two ATPases, RUVBL1 and RUVBL2 are the catalytic components of the co-

chaperone complex, while RPAP3 and PIH1D1 are proposed to regulate their enzymatic activity (Houry

et al., 2018; Lynham and Houry, 2018). Both RUVBL1 and RUVBL2 contain AAA+ ATPase rings with

a DII domain, which is involved in the mediation of the hexamer or dodecamer formations of RUVBL1

and RUVBL2 complexes or binding to client proteins. RUVBL1 and RUVBL2 facilitate client complex

folding and assembly and also regulates the dissociation of the R2TP complex from substrates through

ATP binding and hydrolysis (Lynham and Houry, 2018). RPAP3 contains two N-terminal

tetratricopeptide repeat (TPR) domains and a C-terminal domain formed by eight α-helices. The two TPR

domains in RPAP3 are known to bind to the conserved MEEVD motif of the C-terminal tails of HSP70

and HSP90, and the C-terminal domain is involved in binding to RUVBL2 (Benbahouche et al., 2014;

Martino et al., 2018; Maurizy et al., 2018). PIH1D1, which is part of the PIH1 domain-containing protein

family, has been suggested to scaffold interactions between the R2TP complex and adaptor and client proteins. It has a conserved PIH1 domain on its N-terminus and a CS domain towards its C-terminus, where the CS domain is used to directly interact with RPAP3. The N-terminal PIH1 domain has been proposed to determine substrate specificity of the R2TP complex by binding to a DpSDD/E motif in client proteins phosphorylated by casein kinase 2 (Hořejší et al., 2014; Lynham and Houry, 2018). In addition to the canonical components of the R2TP complex, additional components that can associate with the R2TP complex include those of the non-canonical Prefoldin-like complex and its accessory proteins. WDR92/Monad is known to associate with the R2TP complex and is composed of seven WD40-repeats that form a β-propeller structure. WDR92 interacts with RPAP3 through its C-terminal domain, and has been suggested to interact with PIH1D1 as well, to stabilize the RPAP3-PIH1D1 interaction (Itsuki et al., 2008; Lynham and Houry, 2018). Some WD40-repeat proteins have been reported to interact with the chaperonin containing TCP-1 (CCT) complex for folding, and the R2TP/Prefoldin-like complex has also been suggested to interact with the CCT chaperone, suggesting that WDR92 may scaffold the interaction between R2TP and CCT for protein folding (Liu et al., 2019; Lynham and Houry, 2018).

Interestingly, the canonical components of the R2TP/Prefoldin-like complex have been implicated in axonemal dynein arm assembly in various model organisms (Fabczak and Osinka, 2019). Both *RUVBL1/Pontin* and *RUVBL2/Reptin* zebrafish mutants developed motile ciliopathy phenotypes, including abnormal body axis curvature, pronephric cysts, and dilated pronephric ducts, which resulted from ODA and IDA defects and immotile cilia that was examined in the pronephric tubule by HSVMA and TEMs (Zhao et al, 2013; Li et al., 2017). Genetic variants in *RUVBL1/Pontin* in mice were embryonic lethal; however, when *RUVBL1/Pontin* was specifically deleted in premeiotic male gametes, mice became infertile due to immotile sperm (Li et al., 2017). In a tamoxifen-inducible *RUVBL1* knockout mouse model that was specific to multiciliated cells, mice developed renal disease and hydrocephalus, where it was shown that they have dynein arm defects in their cilia by TEMs and immunofluorescence (Dafinger et al., 2018). Additionally, RNAi depletion of *RPAP3* in *Drosophila* testis and chordotonal sensory

43

neurons lead to male infertility and poor proprioception, which are typical defective motile ciliary phenotypes in fruit flies, due to immotile sperm and cilia caused by dynein arm defects (zur Lage et al., 2018). By performing co-immunoprecipitations and immunoblots of testis lysates from WT and *RUVBL1* mutant mice, RUVBL1 and RUVBL2 were shown to stabilize each other and that they associated with HSP90; however, the interaction with HSP90 was disrupted when *RUVBL1* was mutated. Additionally, the levels of DNAI1 and DNAI2 were substantially reduced in *RUVBL1* mutant mice, suggesting that RUVBL1 and RUVBL2 could be involved in the folding and stabilization of the DIC complex (Li et al., 2017). Co-immunoprecipitations and yeast two-hybrid assays also demonstrated that RUVBL1 and RUVBL2 interact with other known dynein arm assembly factors, specifically DNAAF1 and LRRC6, which further provides evidence that RUVBL1 and RUVBL2 are involved in axonemal dynein assembly (Zhao et al., 2013; Dafinger et al., 2018).

The R2TP canonical component PIH1D1 has also been implicated in dynein arm assembly in *Chlamydomonas*, *Drosophila*, and *Danio rerio* mutants (Fabczak and Osinka, 2019). *Chlamydomonas ida10* mutants had a slower swimming velocity with a reduced beat frequency due to a drastic reduction in IDAs *b*, *c*, and *d*, as well as a slight reduction of ODAs and IDA *e*, as shown by urea gels, immunoblots, and electron microscopy. It was discovered that the *ida10* mutants were deficient in MOT48, which is the ortholog of PIH1D1. Additionally, co-immunoprecipitations for DNALI1 in *ida10* mutant cells demonstrated a decrease in the association of IDA *c* heavy chain and actin with DNALI1 (Yamamoto et al., 2010). Mutations in *PIH1D1* in *Danio rerio* caused a reduced beat frequency of sperm flagella and Kupffer's vesicle motile cilia due to ODA and IDA *c* defects as shown by cryo-electron tomography and immunofluorescence staining of cilia (Yamaguchi et al, 2018). Tissue-specific RNAi depletion of *PIH1D1* in *Drosophila* chordotonal sensory neurons and testis resulted in impaired proprioception and infertility due to immotile cilia and sperm caused by defective dyneins (zur Lage et al., 2018).

WDR92, a component of the Prefoldin-like complex that associates with the R2TP complex, has also been associated with axonemal dynein arm assembly. RNAi knockdown of *WDR92* in planarian

Schmidtea mediterranea resulted in immotile or slower, uncoordinated ciliary beating due to defects in ciliary architecture that included loss of ODAs, occluded A-tubules, incomplete B-tubules, and defective central pairs (Patel-King and King, 2016). RNAi depletion or CRISPR/Cas9 knockout of *WDR92* in chordotonal sensory neurons and germline cells of *Drosophila* lead to defective proprioception demonstrated by climbing assays, impaired auditory function shown by tonal hearing assays, and infertility. These motile ciliopathy phenotypes were caused by outer and inner dynein arm defects in chordotonal sensory cilia and sperm flagella as shown by TEMs and immunofluorescence studies (zur Lage et al., 2018). DNA mutagenesis screens in *Chlamydomonas* resulted in a *WDR92* mutant that either did not produce flagella or had extremely short flagella that were immotile due to the loss of ODAs and IDAs. There was a drastic decrease in the levels of dynein heavy chains and the association of the DHCs to the DIC/LC complex in the *WDR92 Chlamydomonas* and *Drosophila* mutants as demonstrated by comparative mass spectrometry, immunoblots, and gel-filtration chromatography fractionation studies, which suggests that WDR92 is necessary for the folding and/or stabilization of the DHCs (zur Lage et al., 2018; Liu et al., 2019; Patel-King et al., 2019). Additionally, co-immunoprecipitations for WDR92 in S2 *Drosophila* cells suggest WDR92 interacts with RPAP3, PIH1D1, RUVBL1, RUVBL2, DNAAF1 (LRRC50), and SPAG1 (zur Lage et al., 2018). In *Chlamydomona*s, WDR92 was shown to interact either directly or indirectly with IDA heavy chains, RUVBL1, RPAP3, MOT48 (PIH1D1), ODA7 (DNAAF1), and DNAAF4 (DYX1C1) by co-immunoprecipitations followed by mass spectrometry, yeast two-hybrid assays, or GST-pull down assays (Liu et al., 2019; Fabczak and Osinka, 2019).

Due to these canonical PAQosome components being involved in the dynein arm assembly and their interactions with other DNAAFs, it has been suggested that DNAAFs may interact together in R2TP-like complexes to facilitate the folding and assembly of axonemal dyneins. Interestingly, several DNAAFs have similar protein domains as the R2TP canonical components (Figure 2.5). DNAAF2 (KTU), DNAAF6 (PIH1D3), and PIH1D2 are part of the PIH1 domain-containing protein family and share a similar protein structure as PIH1D1, with an N-terminal PIH1 domain and a CS domain further downstream or at the C-terminus. SPAG1 contains three tetratricopeptide repeat (TPR) domains towards

the N-terminus and a RPAP3-like C-terminal domain, which is similar to RPAP3 that has two N-terminal

TPR domains and a RPAP3 C-terminal domain. Additionally, DNAAF4 (DYX1C1) also contains a N-

terminal CS domain and a C-terminal TPR domain (Fabczak and Osinka, 2019). Interestingly, by

performing luminescence-based mammalian interactome (LUMIER) co-immunoprecipitation assays and

stable isotope labeling by amino acids in cell culture (SILAC) proteomic analysis in HEK293T and HeLA

cells, SPAG1 was shown to interact with the RUVBL1/2 hexamer, with this interaction being

strengthened by the addition of PIH1D2, suggesting it forms an R2TP-like complex named R2SP

(RUVBL1-RUVBL2-SPAG1-PIH1D2) (Maurizy et al., 2018). While PIH1D2 is not a known causative

PCD gene in humans, it has been associated with dynein arm assembly in *Danio rerio*, where *PIH1D2*

mutants have either immotile or reduced ciliary beat frequency of Kupffer's vesicle cilia due to defective

ODAs (Yamaguchi et al., 2018). Another putative R2TP-like complex named R2SD has been proposed

and is comprised of the RUVBL1/2 hexamer with SPAG1 and DNAAF2 (Maurizy et al., 2018). SPAG1

has also been shown to be able to interact with HSP70 and HSP90 through its first and third TPR

domains, similar to RPAP3 interacting with HSP70 and HSP90 through its first and second TPR domains

(Chagot et al., 2019; Dermouche et al, 2021). Additional interactions between the different DNAAF4

isoforms and PIH1 domain containing proteins, DNAAF2 and DNAAF6, were shown to occur without

the presence of the RUVBL1/2 hexamer (Maurizy et al., 2018). These RT2P-like complexes composed of

either DNAAF2-DNAAF4 or DNAAF4-DNAAF6 have been previously suggested to interact with

HSP70 or HSP90 to facilitate folding and assembly of dynein arms (Omran et al., 2008; Tarkar et al.,

2013; Dong et al., 2014; Paff et al., 2017). However, whether these R2TP-like complexes containing

DNAAFs, including SPAG1, DNAAF2, DNAAF4, and DNAAF6, are assembled endogenously in

multiciliated cells in humans and their potential roles in axonemal dynein arm assembly remains to be

explored.

2.3 Sperm-associated antigen 1 (SPAG1)

Sperm-associated antigen 1 (SPAG1) is a protein known to cause PCD in humans when

dysfunctional (Knowles et al., 2013). It was first discovered using the antiserum of an infertile woman

where SPAG1 antibodies caused sperm aggregation leading to infertility (Zhang et al., 1992). Since then, multiple isoforms, expression patterns, interactions, and functions in various model organisms and cell lines have been identified for SPAG1, which are described below.

2.3.1 Structure of SPAG1 Protein

In humans, the *SPAG1* gene [NCBI, NG_033834.2] is located on chromosome 8 and contains 19 exons with 18 of them encoding at least three distinct SPAG1 isoforms (Lin et al., 2001; Knowles et al., 2013). The nucleotide sequence consists of 3,818 bp with a reading frame of 2,778 bp that is translated into a full-length protein with a total of 926 amino acids (106-kDa) (Lin et al., 2001). Another common isoform of SPAG1 is derived from a transcript that alternatively splices out exon 11 and 12 of SPAG1, and when translated, is an 857 amino acid protein (94-kDa) (Neesse et al., 2007). Additionally, according to the Universal Protein Database, SPAG1 also has another isoform that produces a 416 amino acid protein (48-kDa) that is missing amino acids #417 – #926 of the canonical full-length protein [Uniprot Q07617-2]. The full-length 106-kDa SPAG1 protein contains a total of nine tetratricopeptide repeat (TPR) motifs, arranged in three domains (a.a. #209 – 309, #445 – 554, and #623 – 724), on the N-terminus and middle of the protein. Full-length SPAG1 also contains a RPAP3-like C-terminal domain (Knowles et al., 2013; Maurizy et al., 2018) (Figure 2.6). Tetratricopeptide repeat motifs are 34-amino acid motifs that are usually repeated in arrangements of two to twenty repeats. These repeated motifs give rise to two antiparallel α-helices that fold in a helix-turn-helix conformation. A short loop of usually four residues connects adjacent repeats to each other. TPR motifs can form scaffolds to mediate protein-protein interactions, particularly with HSPs, and are involved in assembly of multiprotein complexes (Perez-Riba and Itzhaki, 2019). The SPAG1 C-terminal domain is similar to that found in RPAP3, and it has been suggested to be involved in the interaction between RPAP3, or SPAG1, and the RUVBL1/RUVBL2 hexamer (Maurizy et al., 2018). The mouse homolog Spag1, also known as Tpis, also has distinct isoforms: the full-length protein has 901 amino acids (101-kDa) with 8 TPR motifs [GenBank AF181252; Uniprot Q80ZX8-2], while the second isoform is transcribed from an alternative start site at

exon 11 and translates as a 529 amino acid protein (59-kDa) with 5 TPR motifs [GenBank AF181253;

Uniprot Q80ZX8-2] (Takaishi et al., 1999).
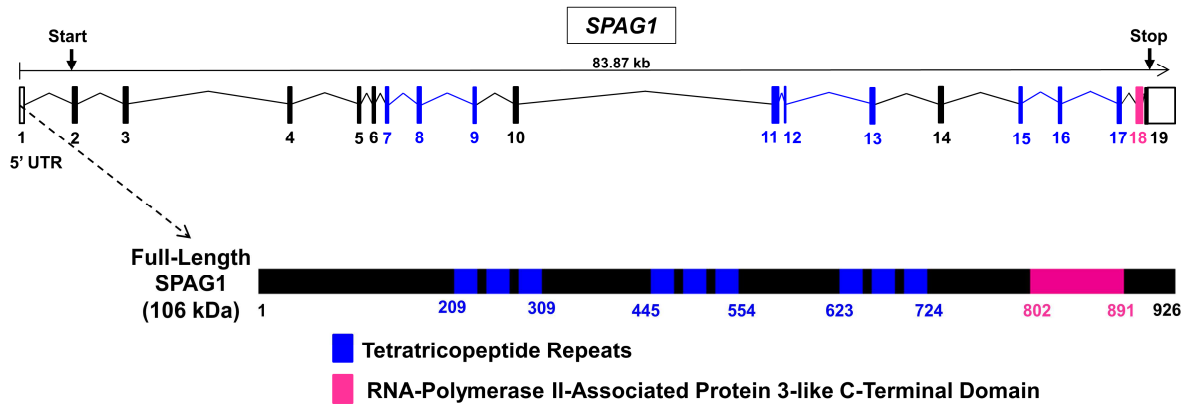
Figure 2.6 The gene sperm-associated antigen 1 (*SPAG1*) and its translated protein

Schematic showing the *SPAG1* gene, which contains 19 exons, with 18 of those exons being coded. The

canonical start codon is in exon 2, and full-length SPAG1 is translated into a 926 amino acid protein

(106-kDa) with three tetratricopeptide repeat (TPR) domains (blue) and a RPAP3-like C-terminal domain

(pink). Exons in the gene diagram are color-coded to correspond with their translated protein domains.

2.3.2 Functions of SPAG1

SPAG1 has been suggested to be involved in spermatogenesis and to play a role in fertilization.

In mice, it was discovered that the two isoforms of Tpis had differential expression in different tissues,

where the shorter isoform (59-kDa) was expressed in tissues with stratified squamous epithelium

including skin, tongue, esophagus, and forestomach and the longer isoform (101-kDa) was most strongly

detected in testis. *In situ* hybridization also showed expression of Tpis corresponding with differentiation

of spermatogenic cells, suggesting Tpis is involved in spermatogenesis (Takaishi et al., 1999).

Immunofluorescence studies using SPAG1 antibodies have also shown that SPAG1 localizes to the head

and midpiece of human sperm and also in germ cells located in the human testis. A recombinant protein

of SPAG1 was used to evaluate its ability to inhibit fertilization in adult female rats, where immunized

rats were shown to be infertile or had a significant reduction in their fertility. Additionally, SPAG1 has phosphorylation sites that could be phosphorylated by protein kinase C (PKC) and the presence of a putative P-loop structure was discovered at amino acid residues 781 – 788 that may bind GTP or ATP and have GTPase activity to regulate spermatogenesis (Lin et al., 2001), but this finding has recently been challenged (Chagot et al., 2019).

Additional studies that have shown SPAG1 is involved in the production, proliferation, and differentiation of gametes have been performed in Large White boars and mice (Huang et al., 2016; Hu et al., 2017). SPAG1 was discovered to be a target of the microRNA miR-638 in immature porcine Sertoli cells, where SPAG1 is downregulated by miR-638. This downregulation inhibits cell proliferation, as shown by S phase arrest, suppression of cell cycle regulators, and MTT assays showing a slower proliferation rate, and promoted apoptosis by ATP depletion indirectly through inhibition of the PI3K/AKT signal transduction pathway (Hu et al., 2017). SPAG1 was also shown to be required for spindle morphogenesis and spermatogonia proliferation by showing co-localization with α-tubulin and localization into astral structures in metaphase and spindle midzone in telophase, as well as a decrease in ɣ-tubulin, SOX2, and POU5F expression in *SPAG1* siRNA transfected Sertoli cells (Hu et al., 2017). In mouse oocytes, SPAG1 was shown to co-localize with meiotic spindles and revealed to be necessary for spindle morphogenesis. Knockdown of *SPAG1* by RNAi in mouse oocytes lead to a higher incidence of cellular arrest at the germinal vesicle stage and disrupted polar body extrusion. This is due, in part, to activation of AMP-dependent kinase (AMPK) that leads to an increase in cAMP and a decrease in ATP in oocytes which disrupts germinal vesicle breakdown. *SPAG1* RNAi depletion also resulted in mislocalization of MAPK, due to reduced phosphorylation and activation, which leads to mislocalization of the microtubule organizing center-associated protein ɣ-tubulin and defective spindle morphogenesis. Additionally, *SPAG1*-depleted mouse oocytes failed to form actin caps, cortical granule free domains, and contractile rings and were shown to have a reduced expression of actin (Huang et al., 2016). Thus, these studies suggest that SPAG1 is necessary for regulating microtubules and actin during cell division, specifically in spindle morphogenesis.

SPAG1 has also been suggested to be a potential oncogene and prognostic biomarker due to the high expression of SPAG1 in multiple cancers, including pancreatic ductal adenocarcinomas (PDAC), testicular germ cell seminomas, renal, breast, colon, prostate, and acute myeloid leukemia (Neesse et al., 2007; Biermann et al., 2007; Siliņa et al., 2011; Shamsara and Shamsara, 2020; Lin et al., 2021; Gu et al., 2022). SPAG1 being considered an oncogene may suggest it is involved in cell growth, cell proliferation, and/or cell motility. Intriguingly, immunocytochemical and immunoprecipitation analysis demonstrated that SPAG1 colocalized and interacted with microtubules in two pancreatic cancer cell lines, Panc1 and IMIM-PC2. siRNA knockdown of *SPAG1* in Panc1 cells followed by transwell migration assays and wound healing assays showed a severe impairment of cell motility following SPAG1 expression reduction (Neesse et al., 2007). Knockdown of *SPAG1* by shRNA in breast cancer cell lines MDA-MB-231 and T-47D significantly lowered cell proliferation as measured by green fluorescent protein (GFP)-based imaging and MTT assay, cell colony formation as measured by light microscopy, and the number of cells in G1 phase of the cell cycle as measured by fluorescence activated cell sorting (Lin et al., 2021). It was also demonstrated that subfragments of SPAG1 and G protein β 1 subunit (Gβ1) interact as shown by yeast two-hybrid, co-localization, and co-immunoprecipitation analysis in HEK293 cells, and overexpression of a subfragment of SPAG1 with the TPR region and putative P-loop in COS-7 cells activated extracellular signal regulated kinases 1/2 (ERK1/2) in a PKC-mediated pathway. ERK are known to be involved in a variety of cellular processes, but mainly cell growth, development, and division (Liu et al., 2006).

In 2013, Knowles and colleagues identified genetic variants in the *SPAG1* gene in at least 14 different PCD-affected individuals with a typical PCD phenotype, including neonatal respiratory distress, bronchiectasis, sinusitis, otitis media, and situs abnormalities. These PCD-affected individuals all had ODA and IDA defects with completely immotile cilia as shown by TEM, videomicroscopy, and immunofluorescent staining for DNAH5 and DNALI1. Morpholinos for *SPAG1* in zebrafish produced defective motile cilia phenotypes, including dorsal body axis curvature and hydrocephalus. Interestingly, SPAG1 was only detected in undifferentiated and differentiated human airway epithelial cell lysates, but

not in isolated ciliary axonemes, by immunoblots. These immunoblot findings and the defects in dynein arms in *SPAG1*-deficient PCD-affected individuals suggest that SPAG1 is involved in the axonemal dynein arm assembly in the cytoplasm as a DNAAF, and is also known as DNAAF13 (Knowles et al., 2013). Additionally, Horani and colleagues demonstrated that SPAG1 is expressed in primary human airway epithelial cells, along with DNAAF2 and DNAAF5 (HEATR2), at an earlier stage in differentiation than other DNAAFs by RT-PCR, immunoblots, and immunofluorescence studies. SPAG1 was shown to interact with DNAAF5 (HEATR) and DNAAF2 by co-immunoprecipitation for GFP-DNAAF5 in human bronchial epithelial cells and by fluorescence resonance energy transfer (FRET) analysis in HEK293T cells, suggesting these three DNAAFs form a complex in an early stage of dynein arm assembly. Mutations in DNAAF5 did not disrupt the interaction between SPAG1 and DNAAF5 as shown by FRET analysis, but instead lead to aggregates of the members of the early preassembly complex which then undergo degradation (Horani et al., 2018). Additionally, *Spag1* (*CG18472*) mutants in *Drosophila* had defective larval hearing and adult proprioception and complete male infertility due to immotile cilia caused by a lack of ODAs and IDAs shown by TEMs. Spag1 was shown to interact with Wdr92 and Dnah10 in *Drosophila* testes by affinity purification-mass spectrometry (AP-MS), where the Spag1 and Wdr92 interaction was strengthened by the overexpression of Pontin and Reptin (zur Lage et al., 2018). SPAG1 has also been shown to localize in the cytoplasm of multiciliated cells, specifically in large puncta known as DynAPs, and has been shown to co-localize with DNAAF5, DNAAF2, ZMYND10, and RUVBL2 in these DynAPs in human airway epithelia cells and Xenopus embryos (Horani et al., 2018; Huizar et al., 2018; Lee et al., 2020).

SPAG1 has been suggested to be a paralog of RPAP3, one of the canonical components of the R2TP complex, that arose from a duplication event between the evolution of fungi and metazoans. LUMIER co-immunoprecipitation assays in HEK293 cells for FLAG-tagged RPAP3-like proteins uncovered a strong association between SPAG1 and PIH1D2 or DNAAF2, as well as an association of SPAG1 with RUVBL1 and RUVBL2 that is increased in the presence of co-expressed PIH1D2. Quantitative SILAC proteomic analysis using GFP-tagged DNAAF2 as bait in HeLa cells found strong

interactions with SPAG1, RUVBL1, and RUVBL2, along with HSP70. This suggests that SPAG1 forms

an R2TP-like complex with PIH1D2 or DNAAF2 and the hexamer RUVBL1/RUVBL2, named R2SP or

R2SD, respectively. Additionally, multiple potential clients of the R2SP complex, including liprin-α2,

were identified through yeast two-hybrid screens and LUMIER co-immunoprecipitation assays.

Interestingly, liprin-α2 and its complex components have higher expression and immunoprecipitation

efficiencies measured by luciferase assays and LUMIER co-immunoprecipitation assays in HeLa cells

expressing PIH1D2 compared to HeLa cells without PIH1D2 expression. Thus, liprin-α2 complexes are

more efficiently expressed and formed in presence of PIH1D2, which is present in the R2SP co-chaperone

complex, suggesting that the R2SP complex has chaperone activity (Maurizy et al., 2018). SPAG1 has

been shown to directly interact with HSP70 and HSP90 through its first and third TPR domain with

similar affinities, suggesting that none of the TPR domains preferentially interact with one specific

chaperone. Additionally, GTPase catalytic activity was not seen in this study of SPAG1's putative P-loop

motif, and SPAG1 was instead suggested to be an activator or regulator of the ATP hydrolysis on the

hexamer RUVBL1/2 (Chagot et al., 2019).

2.4 Statement of Research Aims

While SPAG1 has been suggested to be a DNAAF, the exact function it performs in this poorly

understood, complex process is unknown. While several studies have explored SPAG1's role in various

model organisms and cell lines, very few of these studies have been in human airway epithelial cells and

focused on its role in axonemal dynein arm assembly for human respiratory motile cilia, even though

genetic variants in *SPAG1* result in PCD which suggests a vital function in respiratory motile cilia.

Additionally, several of these studies involved exogenous over-expression of epitope-tagged proteins to

study SPAG1 protein interactions and functions. To gain a better understanding of why genetic variants in

SPAG1 leads to ODA and IDA defects resulting in PCD, we propose to explore endogenous SPAG1's

isoforms, expression patterns, and protein interactions in normal human airway epithelia. Additionally,

we also propose to examine the dynein chain defects in motile cilia, the levels of dynein chains, and any

disruptions of the dynein arm subunit complexes in *SPAG1*-deficient PCD human airway epithelia

compared to healthy controls. Thus, we intend to accomplish the following aims, which are discussed in

Chapter 3 and Chapter 4, respectively:

1. Identify and characterize SPAG1 isoforms, their expression patterns, and their protein interactions in

human airway epithelial cells

2. Determine the role of SPAG1 in axonemal dynein arm assembly

CHAPTER 3: IDENTIFICATION AND CHARACTERIZATION OF SPAG1 ISOFORMS, EXPRESSION, AND PROTEIN INTERACTIONS

3.1 Introduction[1]

To further elucidate SPAG1's function in axonemal dynein arm assembly, we aimed to examine its isoforms, expression patterns, and protein interactions in human airway epithelia. SPAG1 is known to have at least three different isoforms: the full-length 106-kDa isoform with 926 amino acids with three TPR domains and a RPAP3-like C-terminal domain, the 94-kDa isoform derived from exon splicing of exons 11 and 12, and the 48-kDa that is truncated from amino acid residue 416 of the canonical full-length protein (Neesse et al., 2007; Knowles et al., 2013). Expression of SPAG1 is present in the majority of cell types examined with enrichment in tissues expressing motile cilia, specifically in respiratory epithelium and testis (Takaishi et al., 1999; Knowles et al., 2013). SPAG1 has been shown to be expressed in early stages of differentiation in primary human airway epithelial cells before other known DNAAFs (Horani et al., 2018). Additionally, several protein interactions have been identified for SPAG1, including DNAAF2, DNAAF5 (HEATR2), RUVBL1, RUVBL2, PIH1D2, WDR92, HSP70, and HSP90 (Horani et al., 2018; zur Lage et al., 2018; Maurizy et al., 2018; Chagot et al., 2019). However, these interactions were identified in model organisms, such as *Drosophila*, or in human primary cells or cell lines through exogenous expression studies. There have been minimal studies examining the protein interactions of endogenous SPAG1 in human respiratory epithelia.

---

[1] The results, figures, and text contained in this chapter has previously appeared, in part, in an article published in the *Journal of Cell Science*. The original citation is as follows:
Smith, A. J., Bustamante-Marin, X. M., Yin, W., Sears, P. R., Herring, L. E., Dicheva, N. N., López-Giráldez, F., Mane, S., Tarran, R., Leigh, M. W. et al. (2022). The role of SPAG1 in the assembly of axonemal dyneins in human airway epithelia. *J. Cell Sci.* **135**, jcs259512. https://doi.org/10.1242/jcs.259512

Here, we identified a 60-kDa isoform of SPAG1 derived from an alternative start site through immunoblots, targeted proteomics, and 5' rapid amplification of cDNA ends (5' RACE) in normal human bronchial epithelial cells. Additionally, droplet digital PCR (ddPCR), immunoblots, and immunofluorescence studies demonstrate that the different SPAG1 isoforms have distinct expression patterns in differentiating human bronchial epithelial cells, where the full-length and 94-kDa SPAG1 expression corresponds with ciliogenesis, while the 60-kDa SPAG1 is expressed constitutively. Co-immunoprecipitations for endogenous SPAG1 in human bronchial epithelial cells followed by mass spectrometry and immunoblot analysis suggests that SPAG1 interacts with R2TP complex components and other DNAAFs to scaffold an R2TP-like complex to interact with dynein chain clients. The evidence for the most abundant DNAAF interactions is strengthened by their co-expression in differentiating human airway epithelial cells shown by ddPCR, immunoblots, and reverse co-immunoprecipitations. Additionally, a time course of co-immunoprecipitations for endogenous SPAG1 in differentiating human bronchial epithelial cells suggests that the 60-kDa SPAG1 isoform has distinct protein interactions from other isoforms and possibly a cilia-independent function. However, DNAAF2 was shown to still be able to interact with SPAG1 using co-immunoprecipitations for full-length or the 60-kDa isoform of SPAG1 co-expressed with DNAAF2 in HEK293T cells. Finally, using co-immunoprecipitations in HEK293T cells co-expressing WT or mutant SPAG1 and/or WT or mutant DNAAF2, the CS-domain in DNAAF2 was discovered to be required for the interaction between DNAAF2 and SPAG1. Multiple domains toward the C-terminus of SPAG1 appear to be involved in the interaction between SPAG1 and DNAAF2.

3.2 Materials and Methods

3.2.1 Overview

In this chapter, we aimed to examine SPAG1's isoforms, expression patterns, localization, and protein interactions in airway epithelia. Model systems used to study motile cilia structure and function include *Chlamydomonas reinhardtii*, two species of *Paramecium*, *Drosophila melanogaster*, two species of *Xenopus*, *Danio rerio*, and *Mus musculus,* as well as *in vitro* cell cultures of airway epithelia (Niziolek et al., 2022). While *SPAG1* knockout models in zebrafish and *Drosophila* have been created previously,

we wanted to acquire a more physiologically relevant model to study dynein arm assembly (Knowles et al., 2013; zur Lage et al., 2018). Thus, we chose to use primary human airway epithelial cells (HAECs), which includes human bronchial epithelial cells (HBECs) and human nasal epithelial cells (HNECs), as our model system as it is the most physiologically relevant out of the PCD model systems to determine the pathophysiology that occurs in PCD-affected individuals at a molecular level. When primary HAECs are fully differentiated in 2D cultures at an air/liquid interphase (ALI), they consist of multiple differentiated cell types found in the pseudostratified airway epithelium, including progenitor basal cells, club cells, mucus-producing goblet cells, and multiciliated cells. These 3D cultures can be composed of up to 50 – 70% of multiciliated cells, unlike that of airway-derived cell lines A10, 16HBE, and Calu-3, and can be maintained for several weeks or months (Rayner et al., 2019; Lodes et al., 2020; Niziolek et al., 2022).

Using HBEC cultures, we examined the expression of SPAG1 mRNA and protein throughout differentiation through a combination of molecular and biochemistry techniques, including droplet digital polymerase chain reaction (ddPCR), immunoblots, and immunofluorescence. We identified a previously uncharacterized SPAG1 isoform through immunoblots and targeted proteomics, and we used 5' rapid amplification of cDNA ends (5' RACE) to sequence this isoform's transcript. We identified potential SPAG1 protein interactions by performing co-immunoprecipitation for endogenous SPAG1 followed by mass spectrometry analysis and immunoblots. We further characterized the top three potential interactors of SPAG1 by performing reverse co-immunoprecipitations, a time course of endogenous SPAG1 co-immunoprecipitations, and examining their expression during differentiation of HBEC cultures using ddPCR and immunoblots. Finally, we explored which domains in SPAG1 and its most abundant interactor were necessary for their interaction using co-immunoprecipitations for epitope-tagged mutant plasmids co-expressed in HEK293T cells, due to this cell line's easy maintenance and high transfection efficiency (Thomas and Smart, 2005).

3.2.2 Cell Culture

All cells and cell lines were acquired from the Marsico Lung Institute Tissue Procurement and Cell Culture Core, unless otherwise stated. HBECs were acquired using protocols approved by the University of North Carolina Institutional Review Board. Normal HBECs obtained from non-smoking male and female donors were cultured at an ALI as described previously (Fulcher and Randell, 2013).

Normal HBECs were acquired and processed by the Marsico Lung Institute Tissue Procurement and Cell Culture Core. Briefly, airway tissue explants were dissected into 1 x 2 longitudinally cut sections, removing all connective tissues and lymph nodes. Sections were then washed three times with Joklik Minimum Essential Medium (JMEM) (Sigma-Aldrich) plus desired antibiotics. Sections were incubated rocking at 4°C for 24 hours in JMEM supplemented with 0.1% Protease XIV (Sigma-Aldrich), 0.001% DNase (Sigma-Aldrich) and antibiotics. After dissociation, contents were added to a 150 mm plastic dish, and fetal bovine serum (FBS) (Corning) was added to a final concentration of 10%. The epithelial surface of sections was scrapped with a #10 scalpel blade, rinsed with phosphate-buffered saline (PBS), and collected into conical tubes. Cells were centrifuged and then incubated for 15 minutes to 1 hour at 37°C in declumping solution (2 mM EDTA, 0.05 mg/mL DTT, 0.25 mg/mL collagenase, 0.75 mg/mL calcium chloride, 1 mg/mL magnesium chloride, and 10 μg/mL DNase in PBS). Dissociated P0 HBECs were centrifuged, resuspended, counted, and plated on type 1/III collagen-coated plastic dishes in antibiotic-supplemented bronchial epithelial growth medium (BEGM). Media was changed every 2 – 3 days until cultures were 70 – 90% confluent. Cultures were passaged using the double-trypsinization technique using 0.025% trypsin-EDTA to collect P1 HBECs (Fulcher and Randell, 2013).

We obtained P1 HBECs from the Marsico Lung Institute Tissue Procurement and Cell Culture Core. P1 HBECs were plated on human type IV placental collagen-coated 12 mm or 30 mm Millicell porous support inserts (Millipore) at ~1.5 x 10^5 cells/cm^2 in ALI media. ALI media was changed, with PBS washes of the apical surface, 3X weekly for the first two weeks, followed by 2X weekly. Once cultures reached confluence, the apical media was removed and cultures were allowed to differentiate at

an ALI. HBEC cultures are considered fully differentiated after 28 days in culture (ALI Day 26). The compositions of BEGM and ALI media are listed in Appendix 1 (Randell and Fulcher, 2013).

HEK293T cells were cultured on 100 mm plastic dishes in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented with 4.5 g/L D-glucose, L-glutamine, 110 mg/L sodium pyruvate, 10% FBS (Corning), and 1X Penicillin-Streptomycin (Thermo Fisher Scientific). Media was changed every 2 – 3 days, and cells were passaged every 4 – 5 days using 0.025% trypsin-EDTA when cultures were 70 – 90% confluent.

3.2.3 Immunoblots

HBECs were lysed in RIPA buffer (Thermo Fisher Scientific) supplemented with protease inhibitor cocktail (Sigma-Aldrich). Cilia were isolated from HBECs using methods described previously (Ostrowski et al., 2002). Briefly, HBEC cultures were washed with PBS, followed by addition of deciliation buffer (10 mM Tris pH 7.5, 50 mM NaCl, 10 mM CaCl2, 1 mM EDTA, 0.1% Triton X-100, 7 mM β-mercaptoethanol, 1% protease inhibitor cocktail) to the cultures' surfaces. The cultures were rocked gently for 1 minute, and the supernatants were collected. This procedure was repeated once. The two supernatants were pooled together, centrifuged at 1,000 x g to pellet debris, and then centrifuged again at 16,000 x g to collect the ciliary axonemes. Ciliary pellets were resuspended in NuPAGE 4X LDS sample buffer (Invitrogen). Using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific), the protein concentration of lysates or cilia isolates was determined.

Samples (5 – 20 μg) of either whole cell lysates, isolated cilia, or co-immunoprecipitation eluates were electrophoresed on Novex NuPAGE 4 – 12% Bis-Tris gels in 1X NuPAGE MOPS-SDS Running Buffer (Invitrogen) and then transferred to 0.45-μm nitrocellulose membranes using 1X NuPAGE Transfer Buffer (Invitrogen). After transfer, membranes were stained with REVERT Total Protein Stain (LI-COR), imaged using a LI-COR Odyssey Scanner, destained, and then blocked in 5% non-fat milk in 1X TBS solution for 1 hour at room temperature (RT). Membranes were then incubated in 5% non-fat milk in 1X TBS plus 0.1% Tween 20 (TBST) with either primary antibodies or secondary antibodies for 1 hour at RT or 4°C overnight, with four washes with 1X TBST for 5 minutes each after each incubation.

Fluorescent membranes were then scanned using a LI-COR Odyssey Scanner, or chemiluminescent membranes were probed using Amersham ECL Plus reagents (GE Healthcare) according to manufacturer's protocol. All antibodies used and their dilutions are listed in Appendix 2.

Quantitative analysis was performed on fluorescent immunoblots scanned by a LI-COR Odyssey Scanner using Image Studio Lite software version 5.2 (LI-COR). For time-course expression studies on HBECs, background-subtracted signals were normalized to a total protein stain, and then normalized to a positive control (Diff. Control). Immunoblot images were processed and their brightness/contrast adjusted with FIJI software version 2.1.0 and Photoshop 2020 version 21.1.0 (Adobe) for publication.

3.2.4 5' Rapid Amplification of cDNA Ends (5' RACE)

The 5'RACE assay was performed on undifferentiated and differentiated HBEC total RNA using a SMARTer 5'/3' RACE Kit according to manufacturer's protocols (TaKaRa). In brief, total RNA was isolated using an RNeasy kit (QIAGEN). The 5' RACE-ready cDNA was synthesized using SMARTScribe Reverse Transcriptase (TaKaRa). Using two different gene-specific reverse primers (GSP1 and GSP2) and the Universal Primer Mix (UPM), 5' RACE was performed following the manufacturer's touchdown PCR cycling conditions. Reactions contained 25 uL 2X SeqAmp Buffer, 2.5 uL 5'-RACE-Ready cDNA, 5 uL 10X UPM, 1 uL 10 uM GSP, 1 uL SeqAmp DNA Polymerase, and molecular grade H_2O to a total volume of 50 uL. Reactions were run at 94°C for 30 seconds and 72°C for 4 minutes for 5 cycles; 94°C for 30 seconds, 70°C for 30 seconds, and 72°C for 4 minutes for 5 cycles; and 94°C for 30 seconds, 68°C for 30 seconds, and 72°C for 4 minutes for 25 cycles. The 5' RACE products were electrophoresed on a 1% agarose gel, and isolated using a NucleoSpin Gel and PCR Clean-Up Kit (TaKaRa). Isolated 5' RACE products were cloned into the provided pRACE vector using an In-Fusion Cloning Kit (TaKaRa). Stellar Competent Cells (TaKaRa) were transformed with 5' RACE product vectors using heat shock and cultured overnight at 37°C. Colonies were selected and subjected to colony PCR, and positive colonies were liquid cultured overnight at 37°C. Plasmids were isolated using a QIAprep Spin Midiprep Kit (QIAGEN) and Sanger sequenced. The 5' RACE sequences were aligned and

compared to *SPAG1* genomic sequence (NCBI, NG_033834.2) using BLAST (NCBI) and SnapGene software version 6.0.2 (GSL Biotech). Sequences of the primers used are listed in Appendix 3.

3.2.5 Droplet Digital Polymerase Chain Reaction (ddPCR)

ddPCR was used to measure mRNA expression as described previously (Heetvelde et al., 2017). Briefly, total RNA was isolated from differentiating HBEC cultures on various days in culture using an RNeasy kit (QIAGEN). Using SuperScript II Reverse Transcriptase (Thermo Fisher Scientific), first-strand cDNA was synthesized. At least 10,000 droplets were generated, using a Bio-Rad QX200 Automated Droplet Generator (Bio-Rad), from 20 μL of ddPCR reaction mixes consisting of 10 ng of cDNA, 250 nM primers, 900 nM probes and 1X ddPCR Supermix for Probes (No dUTP) (Bio-Rad). Target gene probes were fluorescently labelled with FAM, whereas the reference gene (Calreticulin) probe was labelled with HEX. PCR reactions were performed using the cycling conditions provided in the manufacturer's protocols. Reactions were run at 95°C for 10 minutes to activate the enzyme; at 94°C for 30 seconds and then at 60°C for 1 minute for 40 cycles for denaturation, annealing, and extension; and then heated at 98° for 10 minutes to deactivate the enzyme. After PCR, the fluorescent positive and negative droplets were identified using a Bio-Rad QX200 Droplet Reader (Bio-Rad). Calculation of the target-concentration was completed using Quantasoft Analysis Pro software version 1.0.596 (Bio-Rad), after normalization to the concentration of the reference gene, Calreticulin.

The specific primer/probe assay for all SPAG1 isoforms was generated using Primer3Plus, and custom ordered from Bio-Rad. Validation of this custom assay was performed by running ddPCR on undifferentiated and differentiated HBEC cDNA samples, but instead of reading the droplets using a Bio-Rad QX200 Droplet Reader, amplicons were isolated using chloroform, as described previously, and Sanger sequenced (Bio-Rad, 2018). To isolate amplicons from droplets, the bottom oil phase of samples were removed, 10 μL of 1X Tris-EDTA (TE) buffer and 40 μL of chloroform (Sigma-Aldrich) were added to each sample, and mixed. The samples were then vortexed at maximum speed for 1 minute, centrifuged at 15,500 x *g* for 10 minutes, and the PCR products from the upper aqueous phase were isolated. Quantification of the levels of the 60-kDa SPAG1 isoform was calculated by subtracting the

levels of the full-length and 94-kDa isoforms measured by BioRad Assay dHsaCPE5033992 from the levels of all SPAG1 isoforms measured by our custom Bio-Rad Assay for exons 14 and 15. All primers and probes were obtained from Bio-Rad, and their sequences are listed in Appendix 3.

3.2.6 Immunofluorescence

Whole-mount staining of ALI cultures was performed as described previously (Bustamante-Marin et al., 2019). Briefly, washed ALI cultures were fixed with 4% paraformaldehyde (PFA) or 100% methanol for 20 minutes, washed with 1X TBS for 5 minutes three times, and permeabilized with 0.2% Triton X-100 in 1X TBS for 30 minutes at RT. Cultures were then incubated in blocking solution (1% BSA, 1% fish gelatin, 0.1% Triton X-100, and 5% donkey serum in 1X TBS) for 1 hour at RT. Cultures were incubated in primary antibodies diluted in blocking solution for 1 hour at RT, followed by three washes with washing solution (0.25% BSA, 0.25% fish gelatin, 0.025% Triton X-100, and 1.25% donkey serum in 1X TBS) for 15 minutes each. Cultures were then incubated in secondary antibodies diluted in blocking solution for 1 hour at RT, followed by four washes with 1X TBS for 15 minutes each.

Nuclei were stained using DNA dyes SYTO13 or Hoechst33342 (Thermo Fisher Scientific). Whole-mount membranes were mounted on slides with ProLong Gold or Diamond Antifade Reagent (Thermo Fisher Scientific). Slides were imaged using a Leica SP5 inverted confocal microscope or a Zeiss 800 upright confocal microscope with a 63X/1.4 oil objective. No detectable staining was observed for isotype-matched antibody or no primary antibody control slides. All images were processed using FIJI software version 2.1.0 and pseudocolored, and brightness/contrast was adjusted evenly across experimental groups. All antibodies used and their dilutions are listed in Appendix 2.

3.2.7 Creation of SPAG1-FLAG and DNAAF2-HA Constructs

The pcDNA3.1(+)-hSPAG1[WT]-FLAG and pcDNA3.1(+)-hDNAAF2[WT]-HA vectors were designed and purchased from GenScript Biotech. The vector expressing the 60-kDa isoform of SPAG1 and the SPAG1 and DNAAF2 domain deletion vectors were produced with a Q5 Site-Directed Mutagenesis Kit (New England Biolabs) from these original WT vectors according to manufacturer's protocols. The pcDNA3.1(+)-hSPAG1[60-kDa]-FLAG vector was created using primers that flank

nucleotides that correspond to amino acids #1 - #381 in full-length SPAG1. The pcDNA3.1(+)-hSPAG1[-TPR1]-FLAG, pcDNA3.1(+)-hSPAG1[-TPR2]-FLAG, pcDNA3.1(+)-hSPAG1[-TPR3]-FLAG, pcDNA3.1(+)-hSPAG1[-PIHIS]-FLAG, and pcDNA3.1(+)-hSPAG1[-RPAP3]-FLAG vectors were created using primers that flank nucleotides that correspond to amino acids #209 - #309, #445 - #554, #623 - #724, #732 - #784, #802 - #891 in full-length SPAG1, respectively. The pcDNA3.1(+)-hDNAAF2[-PIH1]-HA and pcDNA3.1(+)-hDNAAF2[-CS]-HA vectors were created using primers that flank nucleotides that correspond to amino acids #10 - #182 and #254 - #349 in full-length DNAAF2, respectively.

Briefly, PCR using Q5 Hot Start High-Fidelity 2X Master Mix and the primers that flank the region of SPAG1 absent in the 60-kDa SPAG1 or known domains on the pcDNA3.1(+)-hSPAG1[WT]-FLAG construct led to a linearized vector containing a SPAG1 insert that only contained a transcript for the 60-kDa isoform or linearized vectors that contained SPAG1 inserts without the respective tetratricopeptide domains, PIH1D2 interaction site, or RPAP3 C-terminal domain. PCR using Q5 Hot Start High-Fidelity 2X Master Mix and the primers that flank the region of known domains on the pcDNA3.1(+)-hDNAAF2[WT]-HA construct led to linearized vectors that contained DNAAF2 inserts without the respective PIH domain or CS domain. PCR reactions were composed of 1X Q5 Hot Start High-Fidelity Master Mix, 0.5 μM forward primer, 0.5 μM reverse primer, 10 ng of template plasmid DNA, and nuclease-free H$_2$O to a total volume of 25 μL. PCR reactions were run at the following conditions: 98°C for 30 seconds; 98°C for 10 seconds, 58°C – 72°C for 30 seconds, and 72°C for 4 to 5 minutes for 25 cycles; and then 72°C for 2 minutes followed by a 4°C hold. Annealing temperatures varied based on the primers to create the deletions in the vectors and were 68°C, 58°C, 67°C, 72°C, and 72°C for creating the 60-kDa SPAG1 vector, deleting the first TPR domain in SPAG1, deleting the PIH1D2 interaction site in SPAG1, deleting the PIH1 domain in DNAAF2, and deleting the CS domain in DNAAF2, respectively. The vectors without the second TPR domain in SPAG1, without the third TPR domain in SPAG1, and without the RPAP3 domain in SPAG1 were created using a touch-up PCR approach where the annealing temperature increased by 0.5 degrees for 10 cycles, and then had a constant

temperature for 25 cycles. The ranges of these annealing temperatures to create these vectors were 67°C –

72°C, 57°C – 62°C, and 56°C – 61°C, respectively. Incubation of these PCR products with 1X KLD

Enzyme and 1X KLD Reaction Buffer for 5 minutes at RT led to the ligation of these constructs and

depletion of the original WT constructs. The ligated constructs were transformed into 5-α Competent

Escherichia coli cells (New England Biolabs) using heat shock, and transformed cells were cultured on

selective Luria–Bertani agar plates with 50 μg/mL ampicillin overnight at 37°C. Colonies were picked

and subjected to colony PCR, and positive colonies with expected transcript size were incubated in liquid

cultures overnight at 37°C. Plasmids were isolated using a Plasmid Maxiprep Kit (QIAGEN). Isolated

plasmids had their sequences confirmed by Sanger sequencing. Primers that were used in the creation and

confirmation of these constructs are listed in Appendix 3.

Transfection and expression of these constructs in HEK293T cells using Lipofectamine 2000

Transfection Reagent (Thermo Fisher Scientific) were optimized and confirmed by immunoblotting.

HEK293T cells were plated in two 6-well plates at 5 x 10^5 cells per well for a total of 12 wells. The next

day after seeding, DNA-lipid complexes were created in Opti-MEM medium (Gibco) by adding various

concentrations (6, 9, 12, or 15 μL per well) of Lipofectamine 2000 and 2.5 μg of DNA plasmids per well

and incubating for 20 minutes at RT. These DNA-lipid complexes were then added to HEK293T cells in

fresh HEK293T media and incubated for 1 – 2 days at 37°C. For comparing the expression of WT, 60-

kDa SPAG1, and domain deletion plasmids, each plasmid DNA was added at a 1:1 ratio with its

respective potential binding partner (i.e., WT SPAG1 and DNAAF2 without CS domain or WT DNAAF2

and SPAG1 without TPR1 domain) in the same respective well. Co-transfected HEK293T cells were

lysed with RIPA buffer (Thermo Fisher Scientific) supplemented with protease inhibitor cocktail (Sigma-

Aldrich) at 24 and 48 hours after transfection. After determining the protein concentration of lysates using

a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific), 10 μg of lysates were examined by

immunoblots. Expression of the epitope-tagged proteins were compared by measuring and calculating

ratios of the background-subtracted signals of each protein using Image Studio Lite software version 5.2

(LI-COR).

3.2.8 SPAG1/FLAG-tag/HA-tag Co-Immunoprecipitation

Immunoprecipitations for endogenous SPAG1, DNAAF1, DNAAF2, PIH1D2, and exogenous FLAG-tag and HA-tag were performed using a Pierce Crosslink IP Kit (Thermo Fisher Scientific) according to manufacturer's protocols. In essence, 5 – 10 μg of bait antibody or control IgG was coupled and crosslinked to protein A/G plus agarose resin in spin columns by incubating with Coupling Buffer and then 2.5 mM disuccinimidyl suberate (DSS) in DMSO for 45 minutes at RT for both incubations. HBEC and HEK293T cultures were lysed using IP Lysis/Wash Buffer supplemented with protease inhibitor cocktail (Sigma-Aldrich), and protein concentration was determined using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Cell lysates were precleared by incubating with control agarose resin for 1 hour at 4°C. Similar amounts of protein from precleared lysates were added to columns containing resin crosslinked to bait antibody or IgG control, and these columns were incubated overnight (12 – 18 hours) at 4°C. Columns were washed multiple times with IP Lysis/Wash Buffer and 1X Conditioning Buffer, followed by elution of eluates using Elution Buffer or 4X LDS Sample Buffer. Eluates were examined by immunoblotting as described above or by using mass spectrometry as described below.

Immunoprecipitations for exogenous FLAG-tagged SPAG1 or HA-tagged DNAAF2 in co-transfected HEK293T cells were also performed using the Pierce Crosslink IP Kit (Thermo Fisher Scientific). Before immunoprecipitation, HEK293T cells were transfected with FLAG-tagged SPAG1 and HA-tagged DNAAF2 constructs using Lipofectamine 2000 Transfection Reagent (Thermo Fisher Scientific). HEK293T cells were plated in two to three 6-well plates at 5 x 10^5 cells per well for a total of 10, 12, or 16 wells. The next day after seeding, DNA-lipid complexes were created in Opti-MEM medium (Gibco) by adding 6 or 9 μL of Lipofectamine 2000 and 0.4 - 2.5 μg of DNA plasmids per well and incubating for 20 minutes at RT. These DNA-lipid complexes were then added to HEK293T cells in fresh HEK293T media and incubated for 24 hours at 37°C. Negative controls included a non-transfected control and WT plasmids incubated with either an empty vector or a HA-tagged green fluorescent protein (GFP). The respective concentrations of plasmid DNA added to wells were 5 μg/well for WT SPAG1, 0.8

µg/well for WT DNAAF2, 0.4 µg/well for GFP, 2.5 µg/well for domain-deleted SPAG1, and 0.8 µg/well for domain-deleted DNAAF2. HEK293T cells were then lysed using IP Lysis/Wash Buffer supplemented with protease inhibitor cocktail (Sigma-Aldrich) 24 hours after co-transfection, followed by immunoprecipitation for the FLAG-tag or the HA-tag.

Quantitative analysis of protein interactions was performed by analyzing the level of protein co-precipitated with the bait protein compared to a control on immunoblots using Image Studio Lite software version 5.2 (LI-COR). Specifically, background-subtracted signals were normalized to the amount of bait protein that was eluted in each sample, and then compared with the control, which was normalized to 1.0.

3.2.9 Mass Spectrometry

For targeted proteomics on 60-kDa SPAG1, undifferentiated and differentiated HBEC cultures were lysed using RIPA buffer (Thermo Fisher Scientific) supplemented with protease inhibitor cocktail (Sigma-Aldrich). Protein lysates or co-immunoprecipitation eluates were loaded and electrophoresed on Novex 4 – 12% Bis-Tris NuPAGE gradient gels (Invitrogen), and stained with GelCode Coomassie Blue Stain Reagent (Thermo Fisher Scientific). For targeted proteomics, gel sections were excised corresponding to proteins between 50 – 75 kDa. The UNC Proteomics Core then reduced, alkylated, and in-gel digested proteins with trypsin overnight at 37°C. Peptides were extracted, desalted with C18 spin columns (Pierce) and dried via vacuum centrifugation.

Peptide samples were analyzed by the UNC Proteomics Core with liquid chromatography-tandem mass spectrometry (LC-MS/MS) using an Easy nLC 1000 coupled to a QExactive HF (Thermo Scientific) as described previously (Sun et al., 2020). Samples were injected onto an Easy Spray PepMap C18 column (75 µm id × 25 cm, 2 µm particle size) (Thermo Fisher Scientific) and separated over a 120 minute method. The gradient for separation consisted of 5 – 35% mobile phase B at a 250 nl/min flow rate, where mobile phase A was 0.1% formic acid in water and mobile phase B consisted of 0.1% formic acid in acetonitrile. For untargeted analysis, the QExactive HF was operated in data-dependent mode, where the 15 most intense precursors were selected for subsequent higher energy C-trap dissociation (HCD) fragmentation. Resolution for the precursor scan (m/z 400 – 1600) was set to 120,000 with an

automatic gain control (AGC) target value of 3×10^6 ions, 100 milliseconds inject time. MS/MS scans

resolution was set to 15,000 with an AGC target value of 5×10^4 ions, 50 milliseconds inject time. The

normalized collision energy was set to 27% for HCD, with an isolation window of 1.6 m/z. Peptide match

was set to preferred, and precursors with unknown charge or a charge state of 1 and ≥ 7 were excluded.

Targeted analysis of SPAG1 peptides was conducted on the same LC-MS system as above using parallel

reaction monitoring (PRM). Peptides were eluted over a 65 or 120 minute method, and the QExactive was

operated in PRM mode; an inclusion list was used to target ten previously identified SPAG1 peptides.

MS1 scans (350 - 1600 m/z) were collected at a 120,000 resolution, a 2×10^6 AGC target value and a

200-millisecond injection time. MS/MS (PRM) scans resolution was set to 15,000, with an isolation

window of 2 m/z, a 1×10^6 AGC target value, and a 100-millisecond injection time.

Raw data files were processed using Proteome Discoverer version 1.4 or 2.1 (Thermo Scientific)

by the UNC Proteomics Core. Peak lists were searched against a reviewed UniProt human database,

appended with a common contaminants database, using Sequest. The following parameters were used to

identify tryptic peptides for protein identification: 10 ppm precursor ion mass tolerance; 0.02 Da product

ion mass tolerance; up to two missed trypsin cleavage sites; carbamidomethylation of Cys was set as a

fixed modification; and oxidation of Met and acetylation of the N-terminus were set as variable

modifications. Scaffold software version 4.7.3 (Proteome Software) was used to validate MS/MS-based

peptide and protein identifications. Peptide identifications were accepted if they could be established at

greater than 95% probability to achieve a false discovery rate (FDR) less than 0.1% by the Scaffold Local

FDR algorithm. Protein identifications were accepted if they could be established at greater than 99%

probability and contained at least two identified peptides. Relative quantitation was performed using the

calculated quantitative values (spectral counts) within Scaffold. Positive hits for co-immunoprecipitations

were narrowed down using the following parameters: a significant P-value ($P<0.05$) calculated by a

Fisher's exact test; a fold change of at least 2 between the target and IgG eluates; and at least two unique

peptides per protein in each replicate.

3.2.10 Statistics

Data are expressed as the mean ± standard deviation (s.d.) in all experiments. Experiments involving HBECs and HEK293T cells were performed with at least three independent biological replicates, except proteomic studies which were performed twice. Owing to the small sample size and the possibility of a non-normal distribution, statistical methods were non-parametric tests for the experiments $n \leq 6$. Figure legends specify the statistical method and *P*-values for each experiment. *P*<0.05 was considered significant. The Fisher's exact test for the co-immunoprecipitation studies was calculated using Scaffold software version 4.7.3 (Proteome Software). The following statistical methods were calculated using Prism software version 9.3.1 (GraphPad): Kruskal-Wallis test with uncorrected Dunn's multiple comparisons, linear regression analysis, and a two-tailed one-sample t-test. Graphs were created with Prism software version 9.3.1 (GraphPad), and figures were created and processed using Photoshop 2020 software version 21.1.0 (Adobe).

3.3 Results

3.3.1 Identification of a novel 60-kDa SPAG1 isoform

To begin to explore the isoforms and expression patterns of SPAG1, we examined its expression in HBECs. Immunoblots on whole-cell lysates from undifferentiated HBECs, differentiated HBECs, and isolated ciliary axonemes using an anti-SPAG1 antibody targeting its C-terminus detected the full-length and the 94-kDa isoforms of SPAG1 in differentiated HBECs, whereas a 60-kDa band was present in both undifferentiated and differentiated HBECs (Figure 3.1; Figure 3.2A). No SPAG1 protein was detected in isolated axonemes, as was shown in previous studies (Knowles et al., 2013). Immunoblots for DNAI2 were used as a ciliogenesis control. While this 60-kDa band is present in HBEC lysates, it is not expected to be the predicted 48-kDa SPAG1 isoform (Universal Protein Database Q07617-2), as the SPAG1 antibody's epitope sequence is not present in this isoform. To determine whether the 60-kDa band is a truncated isoform of SPAG1 that contains the C-terminus, we performed targeted proteomics for ten peptides present in SPAG1 on gel-isolated 60-kDa bands from whole-cell lysates. In undifferentiated HBEC lysates, all peptides detected in this 60-kDa band corresponded to the C-terminus half of SPAG1,

whereas no peptides corresponding to the N-terminus half were detected (Figure 3.2; Table 3.1). These

data demonstrate that the 60-kDa band is a SPAG1 isoform that contains the C-terminal half of the

protein but is missing the N-terminal half.
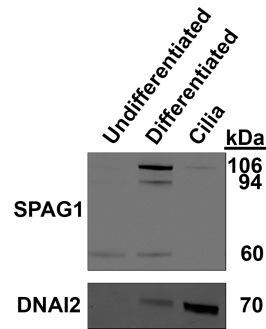
Figure 3.1 SPAG1 expression in undifferentiated and differentiated HBECs, and isolated cilia

Representative immunoblots for SPAG1 and DNAI2 on undifferentiated HBEC whole-cell lysates,

differentiated HBEC whole-cell lysates, and isolated cilia. DNAI2 was used as a ciliogenesis control. [n =

3 independent experiments]
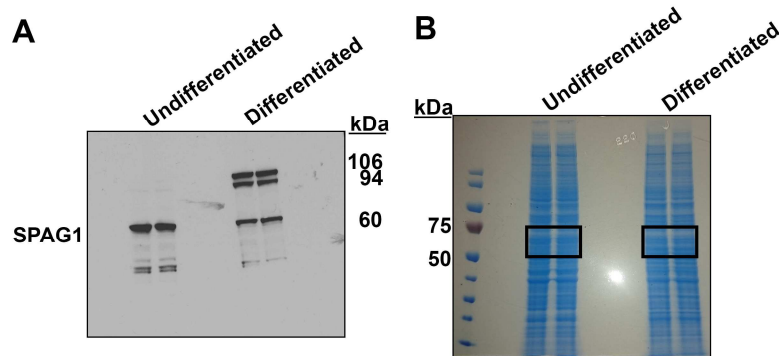
Figure 3.2 Targeted proteomic analysis of 60-kDa SPAG1 isoform

(A) Representative immunoblot for SPAG1 and **(B)** coomassie blue stained SDS-PAGE gel on

undifferentiated and differentiated HBEC lysates used in targeted proteomic analysis of the 60-kDa

SPAG1 isoform. The black outlined areas of the gel were isolated and processed for mass spectrometry

analysis. [*n* = 2 independent experiments]

Table 3.1 Peptides identified in the 60-kDa Band Correspond to the Carboxyl Terminal Half of SPAG1

A.A. # in SPAG1	Peptide Sequence	m/z	Charge	Undiff. IDed	Undiff. # of SC[1]	Diff. IDed	Diff. # of SC[1]
216 - 234	GNEAFNSGDYEEAVMYYTR	1108.47	2				
235 - 248	SISALPTVVAYNNR	752.91	2				
255 - 267	LQNWNSAFQDCEK + carb	820.36	2				
339 - 353	MVIQEIENSEDEEGK + ox	883.4	2			YES	4
339 - 356	MVIQEIENSEDEEGKSGR + ox	689.32	3				
513 - 526	ALELHPFSMKPLLR + ox	417.74	4	YES	5		
546 - 563	TVLQIDCGLQLANDSVNR + carb	672.68	3	YES	13	YES	4
567 - 577	ILMELDGPNWR + ox	680.34	2	YES	5	YES	5
625 - 637	ALKEEGNQCVNDK	752.85	2	YES	1		
791 - 819	SPEDPEKLPIAKPNNAYEFGQIINALSTR	803.92	4	YES	3	YES	2
859 - 878	NNLIEKDPSLVYQHLLYLSK	597.58	4	YES	1		

Abbreviations are as follows: Undiff., Undifferentiated; Diff., Differentiated; IDed, Identified.
[1]SC = Spectral Counts

To characterize the transcript of the 60-kDa SPAG1 isoform, we performed 5′ rapid amplification of cDNA ends (5′ RACE) on total RNA isolated from undifferentiated and differentiated HBECs using reverse primers for SPAG1 exons 15 and 19. The exon 19 reverse primer amplified major transcripts that were ~1700 bp and ~2800 bp long in undifferentiated and differentiated HBEC samples, respectively. The exon 15 reverse primer amplified ~1000-bp and ~2000-bp transcripts in undifferentiated and differentiated HBEC samples, respectively (Figure 3.3A). As there was very little full-length SPAG1 expressed in undifferentiated HBECs, the shorter transcript was predicted to be responsible for the 60-kDa SPAG1 isoform detected on immunoblots. Subsequent cloning and sequencing of these transcripts revealed that undifferentiated HBECs expressed a SPAG1 transcript that contained exons 11-19. Interestingly, this transcript initiated 79 bp upstream of the start of exon 11 in intron IVS10 (NG_033834.2, base 59,977). Kozaq sequence analysis predicts the next most likely start codon for SPAG1 to be the methionine in exon 11 at amino acid 382 (NG_033834.2, bases 60,103 – 60,105) with a reliability of 0.38, compared to 0.62 for the canonical start codon of full-length SPAG1 (Table 3.2). Thus, the ATG at bases 127-129 of this transcript is likely to be the start codon for the 60-kDa isoform. When

translated, this transcript remains in-frame with the full-length isoform of SPAG1, and produces a 545-amino acid protein (60-kDa) that is identical to amino acids 382-926 of full-length SPAG1 (Figure 3.3B). Therefore, the full-length isoform is a 926-amino acid protein (106-kDa) with three TPR domains and a RPAP3-like C-terminal domain, while the 60-kDa isoform of SPAG1 only contains the second and third TPR domains and the RPAP3-like C-terminal domain (Figure 3.4).
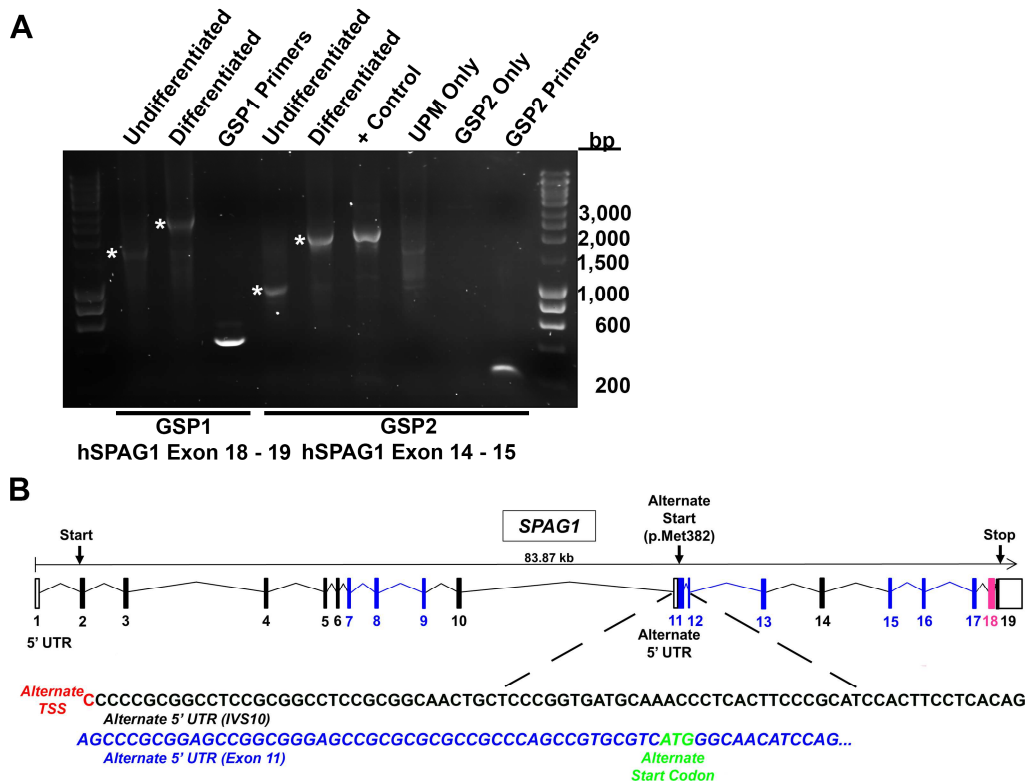
Figure 3.3 60-kDa SPAG1 isoform is transcribed from an alternate start site

(A) Gel of 5' RACE products (white asterisks) produced from total RNA from undifferentiated and differentiated HBEC cultures using gene specific reverse primers (GSP1 Ex. 19, GSP2 Ex. 15) and Universal Primer Mix (UPM) provided by the SMARTer 5'/3' RACE Kit (TaKaRa). GSP1/GSP2 Primers samples were positive controls to test quality of cDNA from HBECs, while UPM Only and GSP2 Only were negative controls. + Control was produced from mouse heart RNA and control primers provided by the SMARTer 5'/3' RACE Kit (TaKaRa). [$n = 2$ independent experiments] **(B)** Schematic of *SPAG1* gene (NCBI, NG_033834.2) indicating the alternate transcription start site (red text) in intron

IVS10, alternate 5' UTR that corresponds to 79 nucleotides of IVS10 (black text) and 47 nucleotides of exon 11 (blue text), and the alternate start codon (p.Met382) (green text) of the 60-kDa SPAG1 isoform determined by sequencing of 5' RACE products from undifferentiated HBECs. Blue and pink boxes represent the exons that translate into tetratricopeptide domains and the RPAP3-like C-terminal domain, respectively.

Table 3.2 Kozak Sequence Analysis[1] of SPAG1 cDNA to Identify Alternative Start Codon

Number of ATG from 5' End	Reliability	Frame	Identity to Kozak Rule A/GXXATGG	Start (bp)	Finish (bp)	ORF Length (aa)	Stop Codon Found	Sequence
1	0.62	3	GXXATGa	33	2810	926	Yes	MTTKDYPSLWGFGTTKTFKIPIEHLDFKYIEKCSDVKHLEKILCVLRSGEEGYYPELTEFCEKHLQALAPESRALRKDKPAATAASFTAEEWEKIDGDIKSWVSEIKKEEDKMHFHETETFPAMKDNLPPVRGSNSCLHVGKEKYSKRPTKKKTPRDYAEWDKFDVEKECLKIDEDYKEKTVIDKSHLSKIETRIDTAGLTEKEKDFLATREKEKGNEAFNSGDYEEAVMYYTRSISALPTVVAYNNRAQAEIKLQNWNSAFQDCEKVLELEPGNVKALLRRATTYKHQNKLREATEDLSKVLDVEPDNDLAKKTLSEVERDLKNSEAASETQTKGKRMVIQEIENSEDEEGKSGRKHEDGGGDKKPAEPAGAARAAQPCVMGNIQKKLTGKAEGGKRPARGAPQRGQTPEAGADKRSPRRASAAAAAGGGATGHPGGGQGAENPAGLKSQGNELFRSGQFAEAAGKYSAAIALLEPAGSEIADDLSILYSNRAACYLKEGNCSGCIQDCNRALELHPFSMKPLLRRAMAYETLEQYGKAYVDYKTVLQIDCGLQLANDSVNRLSRILMELDGPNWREKLSPIPAVPASVPLQAWHPAKEMISKQAGDSSSHRQQGITDEKTFKALKEEGNQCVNDKNYKDALSKYSECLKINNKECAIYTNRALCYLKLCQFEEAKQDCDQALQLADGNVKAFYRRALAHKGLKNYQKSLIDLNKVILLDPSIIEAKMELEEVTRLLNLKDKTAPFNKEKERRKIEIQEVNEGKEEPGRPAGEVSMGCLASEKGGKSSRSPEDPEKLPIAKPNNAYEFGQIINALSTRKDKEACAHLLAITAPKDLPMFLSNKLEGDTFLLLIQSLKNNLIEKDPSLVYQHLLYLSKAERFKMMLTLISKGQKELIEQLFEDLSDTPNNHFTLEDIQALKRQYEL
22	0.38	3	GXXATGG	1176	2810	545	Yes	MGNIQKKLTGKAEGGKRPARGAPQRGQTPEAGADKRSPRRASAAAAAGGGATGHPGGGQGAENPAGLKSQGNELFRSGQFAEAAGKYSAAIALLEPAGSEIADDLSILYSNRAACYLKEGNCSGCIQDCNRALELHPFSMKPLLRRAMAYETLEQYGKAYVDYKTVLQIDCGLQLANDSVNRLSRILMELDGPNWREKLSPIPAVPASVPLQAWHPAKEMISKQAGDSSSHRQQGITDEKTFKALKEEGNQCVNDKNYKDALSKYSECLKINNKECAIYTNRALCYLKLCQFEEAKQDCDQALQLADGNVKAFYRRALAHKGLKNYQKSLIDLNKVILLDPSIIEAKMELEEVTRLLNLKDKTAPFNKEKERRKIEIQEVNEGKEEPGRPAGEVSMGCLASEKGGKSSRSPEDPEKLPIAKPNNAYEFGQIINALSTRKDKEACAHLLAITAPKDLPMFLSNKLEGDTFLLLIQSLKNNLIEKDPSLVYQHLLYLSKAERFKMMLTLISKGQKELIEQLFEDLSDTPNNHFTLEDIQALKRQYEL
8	0.18	3	GXXATGa	402	2810	803	Yes	MKDNLPPVRGSNSCLHVGKEKYSKRPTKKKTPRDYAEWDKFDVEKECLKIDEDYKEKTVIDKSHLSKIETRIDTAGLTEKEKDFLATREKEKGNEAFNSGDYEEAVMYYTRSISALPTVVAYNNRAQAEIKLQNWNSAFQDCEKVLELEPGNVKALLRRATTYKHQNKLREATEDLSKVLDVEPDNDLAKKTLSEVERDLKNSEAASETQTKGKRMVIQEIENSEDEEGKSGRKHEDGGGDKKPAEPAGAARAAQPCVMGNIQKKLTGKAEGGKRPARGAPQRGQTPEAGADKRSPRRASAAAAAGGGATGHPGGGQGAENPAGLKSQGNELFRSGQFAEAAGKYSAAIALLEPAGSEIADDLSILYSNRAACYLKEGNCSGCIQDCNRALELHPFSMKPLLRRAMAYETLEQYGKAYVDYKTVLQIDCGLQLANDSVNRLSRILMELDGPNWREKLSPIPAVPASVPLQAWHPAKEMISKQAGDSSSHRQQGITDEKTFKALKEEGNQCVNDKNYKDALSKYSECLKINNKECAIYTNRALCYLKLCQFEEAKQDCDQALQLADGNVKAFYRRALAHKGLKNYQKSLIDLNKVILLDPSIIEAKMELEEVTRLLNLKDKTAPFNKEKERRKIEIQEVNEGKEEPGRPAGEVSMGCLASEKGGKSSRSPEDPEKLPIAKPNNAYEFGQIINALSTRKDKEACAHLLAITAPKDLPMFLSNKLEGDTFLLLIQSLKNNLIEKDPSLVYQHLLYLSKAERFKMMLTLISKGQKELIEQLFEDLSDTPNNHFTLEDIQALKRQYEL
6	0.16	3	AXXATGc	369	2810	814	Yes	MHFHETETFPAMKDNLPPVRGSNSCLHVGKEKYSKRPTKKKTPRDYAEWDKFDVEKECLKIDEDYKEKTVIDKSHLSKIETRIDTAGLTEKEKDFLATREKEKGNEAFNSGDYEEAVMYYTRSISALPTVVAYNNRAQAEIKLQNWNSAFQDCEKVLELEPGNVKALLRRATTYKHQNKLREATEDLSKVLDVEPDNDLAKKTLSEVERDLKNSEAASETQTKGKRMVIQEIENSEDEEGKSGRKHEDGGGDKKPAEPAGAARAAQPCVMGNIQKKLTGKAEGGKRPARGAPQRGQTPEAGADKRSPRRASAAAAAGGGATGHPGGGQGAENPAGLKSQGNELFRSGQFAEAAGKYSAAIALLEPAGSEIADDLSILYSNRAACYLKEGNCSGCIQDCNRALELHPFSMKPLLRRAMAYETLEQYGKAYVDYKTVLQIDCGLQLANDSVNRLSRILMELDGPNWREKLSPIPAVPASVPLQAWHPAKEMISKQAGDSSSHRQQGITDEKTFKALKEEGNQCVNDKNYKDALSKYSECLKINNKECAIYTNRALCYLKLCQFEEAKQDCDQALQLADGNVKAFYRRALAHKGLKNYQKSLIDLNKVILLDPSIIEAKMELEEVTRLLNLKDKTAPFNKEKERRKIEIQEVNEGKEEPGRPAGEVSMGCLASEKGGKSSRSPEDPEKLPIAKPNNAYEFGQIINALSTRKDKEACAHLLAITAPKDLPMFLSNKLEGDTFLLLIQSLKNNLIEKDPSLVYQHLLYLSKAERFKMMLTLISKGQKELIEQLFEDLSDTPNNHFTLEDIQALKRQYEL
26	0.15	3	GXXATGG	1617	2810	398	Yes	MAYETLEQYGKAYVDYKTVLQIDCGLQLANDSVNRLSRILMELDGPNWREKLSPIPAVPASVPLQAWHPAKEMISKQAGDSSSHRQQGITDEKTFKALKEEGNQCVNDKNYKDALSKYSECLKINNKECAIYTNRALCYLKLCQFEEAKQDCDQALQLADGNVKAFYRRALAHKGLKNYQKSLIDLNKVILLDPSIIEAKMELEEVTRLLNLKDKTAPFNKEKERRKIEIQEVNEGKEEPGRPAGEVSMG

[1]Kozak sequence analysis was performed on *Homo sapiens SPAG1*, transcript variant 1, mRNA (NCBI, NM.003114.5) using the online program ATG^{pr} (https://atgpr.dbcls.jp/) (Nishikawa et al., 2000).
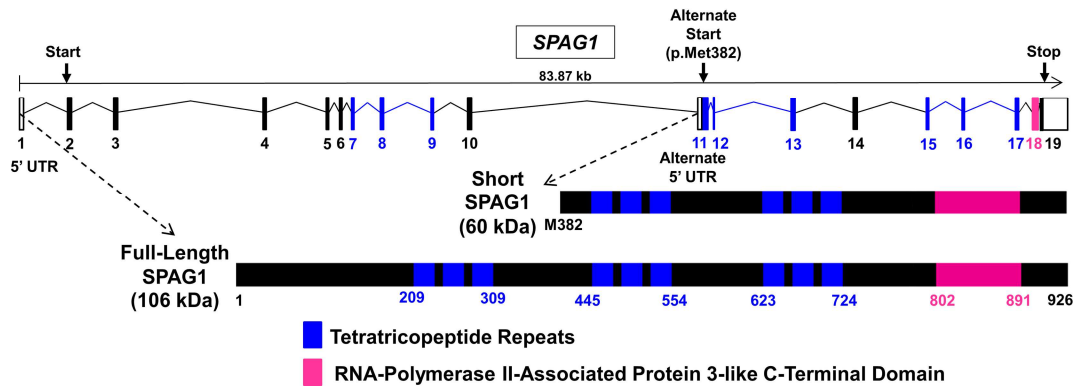
Figure 3.4 Full-length and 60-kDa isoforms of SPAG1

Schematic of the canonical and alternate start sites of SPAG1 and the two different SPAG1 isoforms that are produced. The full-length SPAG1 is a 926-amino acid protein (106-kDa) with three TPR domains (blue boxes) and a RPAP3-like C-terminal domain (pink box). The 60-kDa SPAG1 isoform, or the short isoform, starts at the methionine at amino acid 382 of the full-length SPAG1 and produces a 545-amino acid protein (60-kDa) that has only two of the TPR domains and the RPAP3-like C-terminal domain.

3.3.2 Full-length SPAG1 expression is induced with ciliogenesis

To explore and compare the expression patterns of the full-length and uncharacterized 60-kDa isoforms of SPAG1, we examined their expression throughout HBEC differentiation. The mRNA levels of full-length and 60-kDa isoforms of SPAG1 were measured using droplet digital PCR (ddPCR) and isoform-specific primers in HBEC cultures during differentiation. We detected very low levels of full-length and 94-kDa SPAG1 transcripts in undifferentiated HBECs (ALI day 0). The expression of these transcripts was induced by ALI day 9, which correlated with the initiation of ciliogenesis, and increased throughout differentiation (Fulcher and Randell, 2013). Interestingly, the 60-kDa SPAG1 isoform was expressed at similar levels throughout airway cell differentiation (ALI day 0-28) (Figure 3.5), which could indicate a cilia-independent function.
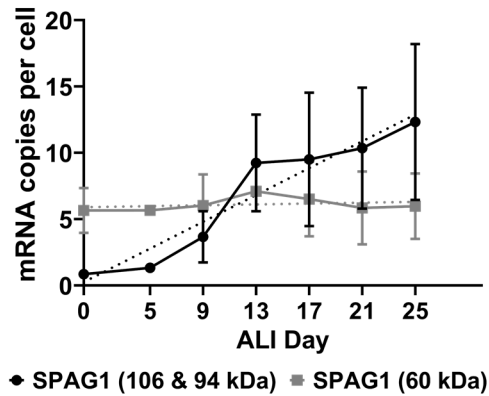
Figure 3.5 Different SPAG1 isoforms have distinct mRNA expression patterns

Quantification of mRNA expression of 106-kDa and 94-kDa (black) or 60-kDa (gray) *SPAG1* isoforms in differentiating HBEC cultures using ddPCR. Calreticulin (*CALR*) was used as a reference gene. Linear regression analysis (dashed lines) determined that the slope of 60-kDa SPAG1 was not significantly non-zero (P=0.603) compared to the slope of 106-kDa and 94-kDa SPAG1 (P<0.001). Data shown are mean ± s.d. across timepoints; *n* = 3 independent experiments using three distinct cell codes.

To measure the abundance of SPAG1 at the protein level, immunoblots for SPAG1 on whole-cell lysates of differentiating HBECs using an antibody targeting the C-terminus of SPAG1 were performed. Immunoblots for DNAI1 were used for a differentiation and ciliogenesis control. Similar to the ddPCR findings, there were very low levels of full-length and 94-kDa SPAG1 protein in undifferentiated HBECs (ALI day 0). The abundance of these isoforms increased starting at ALI day 9, and continued to increase until ALI day 21, where it appears they plateaued. The 60-kDa SPAG1 isoform was present at similar levels throughout HBEC differentiation (Figure 3.6), which corresponds with our ddPCR data.
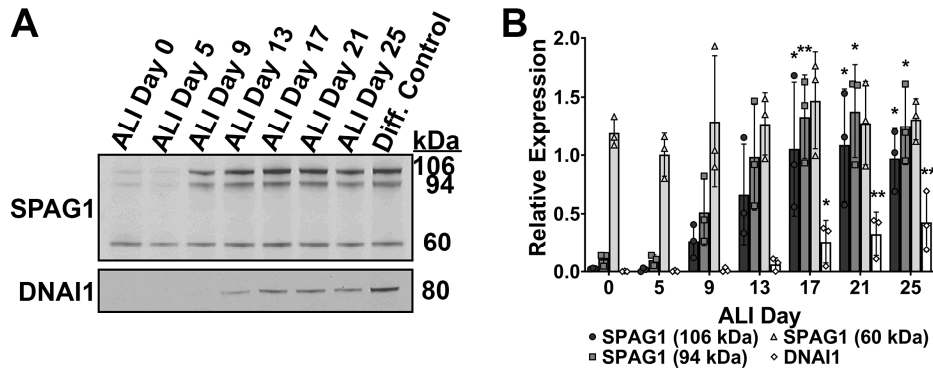
Figure 3.6 Different SPAG1 isoforms have distinct protein expression patterns

(A) Representative immunoblot for SPAG1 and DNAI1 on differentiating HBEC whole-cell lysates. DNAI1 was used as a ciliogenesis control. **(B)** Quantification of SPAG1 isoforms and DNAI1 protein levels analyzed by immunoblots on differentiating HBEC whole-cell lysates. Raw signals were normalized to a total protein stain and then a separate differentiated HBEC whole-cell lysate control (Diff. Control). Data shown are mean ± s.d. across timepoints; n = 3 independent experiments using three distinct cell codes; *P<0.05, **P<0.01 compared to ALI day 0 (Kruskal–Wallis with uncorrected Dunn's multiple comparisons).

We performed immunofluorescent staining of whole-mount HBEC cultures at different stages of differentiation using an antibody against the N-terminus of SPAG1 to detect the full-length isoforms, and an antibody against γ-tubulin to mark basal bodies (Vertii et al., 2016). In undifferentiated HBEC cultures (ALI day 0), no SPAG1-specific staining was observed. When ciliogenesis started to occur (ALI day 5-9), the number of cells co-expressing SPAG1 and basal bodies increased. In fully differentiated HBEC cultures (ALI day 26), a large number of cells co-expressed SPAG1 and basal bodies (Figure 3.7A). Interestingly, there were no cells that expressed basal bodies without SPAG1 expression throughout differentiation. Quantification of these images by counting cells expressing SPAG1 or both SPAG1 and basal bodies, identified as clusters of γ-tubulin-positive puncta, showed continuous increases in cells expressing both SPAG1 and basal bodies throughout differentiation of HBECs, whereas cells expressing only SPAG1increased until ALI day 21 (Figure 3.7B). Thus, the expression of full-length and 94-kDa

SPAG1 suggests it is primarily involved in ciliogenesis, and the constitutive expression of the 60-kDa

isoform indicates that it possibly has a cilia-independent function.
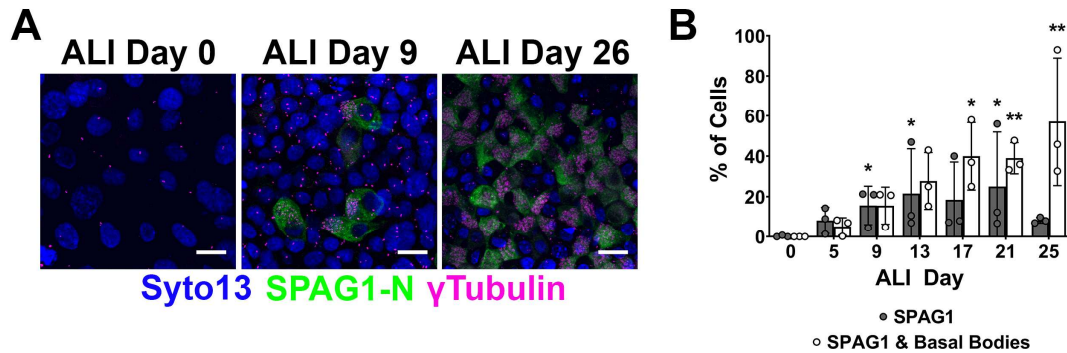
Figure 3.7 Expression of full-length and 94-kDa SPAG1 corresponds with ciliogenesis

(A) Whole-mount immunofluorescence images on differentiating HBEC cultures stained for the N-

terminus of SPAG1-N (green), γ-tubulin (magenta) to mark basal bodies, and nuclei (blue). Images are

representative of a total of six fields of view across two cultures per timepoint per independent

experiment. Scale bars: 10 μm. $n = 3$ independent experiments using three distinct cell codes. **(B)**

Quantification of the percentage of cells expressing either just SPAG1 (grey) or both SPAG1 and basal

bodies (white) in whole-mount immunofluorescent micrographs of differentiating HBEC cultures. Basal

bodies were defined as a cluster of γ-tubulin-positive puncta. There was no full-length or 94-kDa SPAG1,

detected by an antibody that recognizes the amino-terminus, or basal bodies being expressed in

undifferentiated HBECs at ALI day 0. Expression of SPAG1 and basal bodies increased during

differentiation. Of note, throughout differentiation, there were no cells that were detected to have basal

bodies without SPAG1 expression. Percentages of cells were calculated per field of view for a total of six

field of views across two cultures per timepoint per independent experiment. Data shown are mean ± s.d.;

$n = 3$ independent experiments using 3 distinct cell codes; *$P<0.05$, **$P<0.01$ compared to ALI Day 0

(Kruskal-Wallis test with uncorrected Dunn's multiple comparisons).

3.3.3 SPAG1 interacts with dynein chains, DNAAFs, and R2TP/Prefoldin-like complex components

To examine potential protein interactions of SPAG1, immunoprecipitations for endogenous SPAG1 in HBECs were performed using an anti-SPAG1 antibody targeting the N-terminus, and were analyzed by mass spectrometry and immunoblotting (Figure 3.8A; Table 3.3). A total of 52 positive hits were identified (Appendix 4). Intriguingly, two DHCs, DNAH6 and DNAH5, were co-precipitated with SPAG1. A number of previously identified DNAAFs also co-precipitated with SPAG1, including DNAAF2, PIH1D2, DNAAF1, RPAP3, WDR92, DNAAF3, RUVBL1 and RUVBL2 (Table 3.3). Interestingly, RPAP3, WDR92, RUVBL1 and RUVBL2 are canonical components of the R2TP/Prefoldin-like complex, a heat shock protein 90 (HSP90) co-chaperone complex (Houry et al., 2018). According to Database for Annotation, Visualization, and Integrated Discovery (DAVID) gene ontology analysis, the top five most significant terms associated with potential interactors of SPAG1 were: cullin deneddylation; nucleotide-excision repair, DNA damage recognition; protein binding; transcription-coupled nucleotide-excision repair; and axonemal dynein complex assembly. Other terms that were considered significant ($P<0.05$) included cell cycle, regulation of growth, ATPase activity, left/right pattern formation, outer dynein arm assembly, and motile cilium assembly (Table 3.4).

As DNAAF2, PIH1D2 and DNAAF1 were the most abundant DNAAFs identified in the co-immunoprecipitations, we further characterized their interactions with SPAG1. Immunoblots of the immunoprecipitated eluates confirmed that DNAAF2, PIH1D2 and DNAAF1 co-precipitate with SPAG1 (Figure 3.8A). Reverse co-immunoprecipitation for DNAAF2 co-precipitated the full-length and 94-kDa SPAG1 with small amounts of PIH1D2 and DNAAF1 (Figure 3.8B). Reverse co-immunoprecipitation for PIH1D2 only co-precipitated minimal amounts of SPAG1 and DNAAF1, but this may be because of the low binding affinity of the PIH1D2 antibody, as higher concentrations of antibody are needed to detect PIH1D2 on immunoblots, and less than optimal immunoprecipitation conditions (Figure 3.8C). Reverse co-immunoprecipitation for DNAAF1 co-precipitated full-length SPAG1 with small amounts of PIH1D2 and DNAAF2 but not the 94-kDa SPAG1 isoform, suggesting that the binding site for DNAAF1 is absent

in the 94-kDa SPAG1 isoform (Figure 3.8D). These results indicate that SPAG1 specifically interacts with these three proteins, either individually or in a complex.
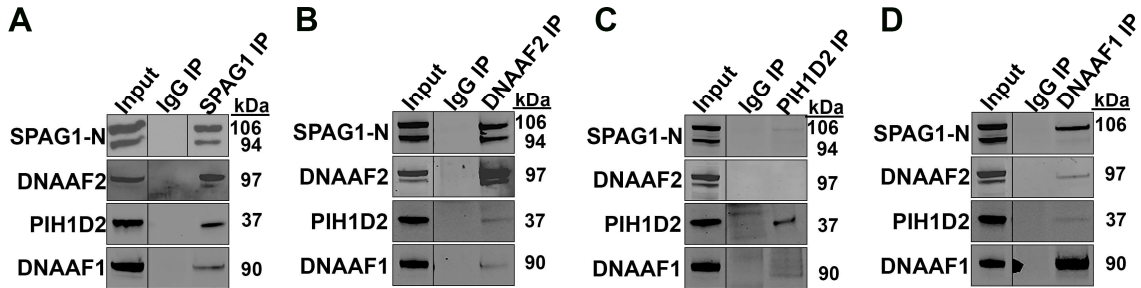
Figure 3.8 SPAG1 interacts with other known DNAAFs

(A) Representative immunoblots for SPAG1, DNAAF2, PIH1D2 and DNAAF1 analyzing co-immunoprecipitation samples for endogenous SPAG1 in HBEC cultures. $n = 3$ independent experiments using three distinct cell codes. **(B – D)** Representative immunoblots for SPAG1, DNAAF2, PIH1D2 and DNAAF1 analyzing reverse co-immunoprecipitation samples for endogenous DNAAF2 **(B)**, PIH1D2 **(C),** and DNAAF1 **(D)** in HBEC cultures. $n = 2$ independent experiments using two distinct cell codes.

Table 3.3 Identified Interactors of SPAG1 Involved in Axonemal Dynein Arm Assembly[1]

Identified Proteins	Gene ID	P-value[2]	Fold Change[3]	Replicate #1		Replicate #2	
				IgG Spectral Counts (Unique Peptides)	SPAG1 Spectral Counts (Unique Peptides)	IgG Spectral Counts (Unique Peptides)	SPAG1 Spectral Counts (Unique Peptides)
Dynein Chains							
Dynein heavy chain 6, axonemal	DNAH6	<0.0001	INF	0 (0)	2 (2)	0 (0)	26 (25)
Dynein heavy chain 5, axonemal	DNAH5	0.04	5	2 (2)	4 (4)	0 (0)	6 (6)
Dynein Arm Assembly Factors							
Kintoun	DNAAF2	<0.0001	INF	0 (0)	88 (45)	0 (0)	117 (48)
PIH1 domain-containing protein 2	PIH1D2	<0.0001	INF	0 (0)	14 (10)	0 (0)	29 (15)
Dynein assembly factor 1, axonemal	DNAAF1	<0.0001	INF	0 (0)	6 (6)	0 (0)	12 (12)
RNA polymerase II-associated protein 3	RPAP3	0.015	INF	0 (0)	3 (3)	0 (0)	4 (3)
WD repeat-containing protein 92	WDR92	<0.0001	14	2 (2)	10 (9)	0 (0)	17 (13)
Dynein assembly factor 3, axonemal	DNAAF3	0.037	8	0 (0)	4 (3)	1 (1)	4 (3)
RuvB-like 2	RUVBL2	<0.0001	5	17 (15)	58 (28)	9 (8)	73 (31)
RuvB-like 1	RUVBL1	<0.0001	4.5	16 (13)	45 (20)	8 (6)	62 (21)
Chaperones							
Heat shock protein 105 kDa	HSP1H1	<0.0001	3.6	13 (10)	25 (19)	3 (3)	32 (24)

[1]This table shows 11 out of 52 positive hits. The complete list of positive hits is shown in Appendix 4, and DAVID gene ontology analysis results are shown in Table 3.5.
[2]P-values were calculated by Fisher's exact test between the SPAG1 and IgG samples. Positive hits were narrowed down using $P<0.05$.
[3]Fold changes were calculated using the spectral counts of the SPAG1 and IgG samples, averaged between the two replicates. Positive hits were narrowed down using a fold change of at least two. INF, infinity or undefined due to zero being denominator.

Table 3.4 DAVID Gene Ontology Analysis on SPAG1 Interactors in Differentiated Human Bronchial Epithelial Cells[1,2]

Term	Count	%	*P*-Value	Benjamini
cullin deneddylation	5	9.4	8.70E-09	3.40E-06
nucleotide-excision repair, DNA damage recognition	5	9.4	5.90E-07	1.20E-04
protein binding	43	81.1	2.20E-06	3.10E-04
transcription-coupled nucleotide-excision repair	6	11.3	2.60E-06	3.40E-04
axonemal dynein complex assembly	4	7.5	2.90E-06	2.90E-04
box C/D snoRNP assembly	3	5.7	1.80E-04	1.40E-02
branched-chain amino acid catabolic process	3	5.7	1.40E-03	9.00E-02
oxidoreductase activity	5	9.4	2.90E-03	1.80E-01
cell cycle	5	9.4	3.90E-03	2.00E-01
3-oxoacyl-[acyl-carrier-protein] reductase (NADH) activity	2	3.8	5.90E-03	2.40E-01
methylcrotonoyl-CoA carboxylase activity	2	3.8	5.90E-03	2.40E-01
oxidation-reduction process	7	13.2	8.00E-03	3.30E-01
regulation of growth	3	5.7	1.10E-02	3.80E-01
protein deneddylation	2	3.8	1.20E-02	3.70E-01
peptidyl-arginine N-methylation	2	3.8	1.20E-02	3.70E-01
COP9 signalosome assembly	2	3.8	1.20E-02	3.70E-01
ATP-dependent 5'-3' DNA helicase activity	2	3.8	1.50E-02	4.10E-01
leucine catabolic process	2	3.8	1.50E-02	4.10E-01
ATPase activity	4	7.5	1.70E-02	3.80E-01
regulation of cellular response to heat	3	5.7	2.10E-02	5.00E-01
3-hydroxyacyl-CoA dehydrogenase activity	2	3.8	2.30E-02	4.30E-01
left/right pattern formation	2	3.8	2.40E-02	5.10E-01
protein-arginine N-methyltransferase activity	2	3.8	2.60E-02	4.20E-01
fatty acid catabolic process	2	3.8	2.60E-02	5.30E-01
lateral ventricle development	2	3.8	3.50E-02	6.10E-01
S100 protein binding	2	3.8	3.80E-02	4.90E-01
biotin metabolic process	2	3.8	4.10E-02	6.40E-01
histone H2A acetylation	2	3.8	4.40E-02	6.40E-01
outer dynein arm assembly	2	3.8	4.40E-02	6.40E-01
motile cilium assembly	2	3.8	4.40E-02	6.40E-01
positive regulation of telomerase RNA localization to Cajal body	2	3.8	4.40E-02	6.40E-01
cell-cell adhesion	4	7.5	4.70E-02	6.50E-01

[1]DAVID gene ontology analysis on proteins identified as positive hits in endogenous SPAG1 co-immunoprecipitations in differentiated human bronchial epithelial cells, listed in Appendix 4.
[2]A functional annotation chart was created using the categories, GOTERM_BP_DIRECT and GOTERM_MF_DIRECT, with the background set to *Homo Sapiens*.

3.3.4 SPAG1 is co-expressed with DNAAF2, DNAAF1, and PIH1D2

There are minimal studies measuring the level of mRNA expression of DNAAFs quantitatively in differentiating HBECs. There also are limited studies examining PIH1D2 and its functions, especially in human airway epithelia. Therefore, we examined and compared SPAG1's expression in differentiating HBECs with other known ciliogenesis-related genes and its expected interactors, including PIH1D2. mRNA levels of full-length *SPAG1*, *DNAAF2*, *PIH1D2*, *DNAAF1*, *DNAI1* and *FOXJ1*, a major

transcription factor controlling ciliogenesis, were measured in differentiating HBECs using ddPCR. Full-length *SPAG1*, *DNAAF2*, *PIH1D2* and *DNAAF1* were induced by ALI day 9, after the expression of the transcription factor *FOXJ1*, and increased throughout differentiation of HBECs (Figure 3.9). This expression pattern is similar to that of DNAI1, which suggests all of these proteins are involved in ciliogenesis.
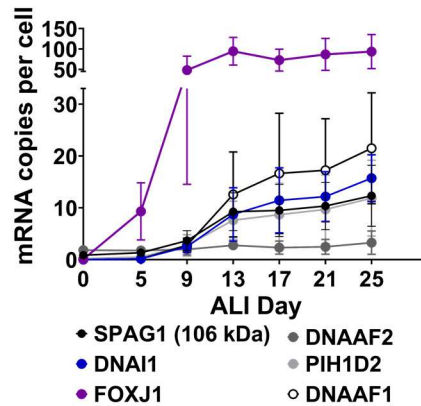
Figure 3.9 Co-expression of *SPAG1* and potential interactors at the RNA level

Quantification of mRNA co-expression between *SPAG1, FOXJ1, DNAI1*, *DNAAF1*, *DNAAF2*, and *PIH1D2* in differentiating HBEC cultures. *FOXJ1* and *DNAI1* were used as ciliogenesis controls. Calreticulin (*CALR*) was used as a reference gene. Data shown are mean ± s.d. across timepoints; $n = 3$ independent experiments using three distinct cell codes.

To measure the level of protein of SPAG1, DNAAF2, PIH1D2, DNAAF1, HEATR2, DNAI1 and FOXJ1 during differentiation of HBECs, immunoblots were performed on whole-cell lysates from differentiating HBEC. SPAG1, DNAAF2, PIH1D2, DNAAF1, and FOXJ1 proteins are expressed around ALI day 9 and then increased throughout differentiation of HBECs. All of these proteins were expressed before the expression of a dynein chain, DNAI1, which was expressed starting at ALI day 13. This is a common characteristic of DNAAFs expression (Horani et al., 2018). These results revealed that the expression of these proteins during differentiation corresponds with their respective mRNA levels (Figure

3.10). Interestingly, HEATR2 is expressed at ALI day 0, decreased in abundance from ALI day 0 to ALI day 9, and then increased from ALI day 9 to ALI day 25. Previously, it was suggested that SPAG1, DNAAF2 and HEATR2 are co-expressed, and form a complex at an earlier stage in dynein arm assembly than other DNAAFs (Horani et al., 2018). Here, we show that only 60-kDa SPAG1 and HEATR2 are highly expressed at early stages of differentiation (ALI day 0 - 5), and HEATR2 was not a positive hit in any of the immunoprecipitation studies for endogenous SPAG1, which is different from previous results. Therefore, it is possible that the earlier expression of HEATR2 in undifferentiated cells (ALI day 0 - 5) is due to an additional function other than dynein arm assembly, similar to the 60-kDa isoform of SPAG1. However, these differences in results could be due to differences in methods, antibodies used, and cell culture conditions. Importantly, the expression of PIH1D2 follows a similar expression pattern as the other DNAAFs, suggesting PIH1D2 is involved in ciliogenesis in human airway epithelium.
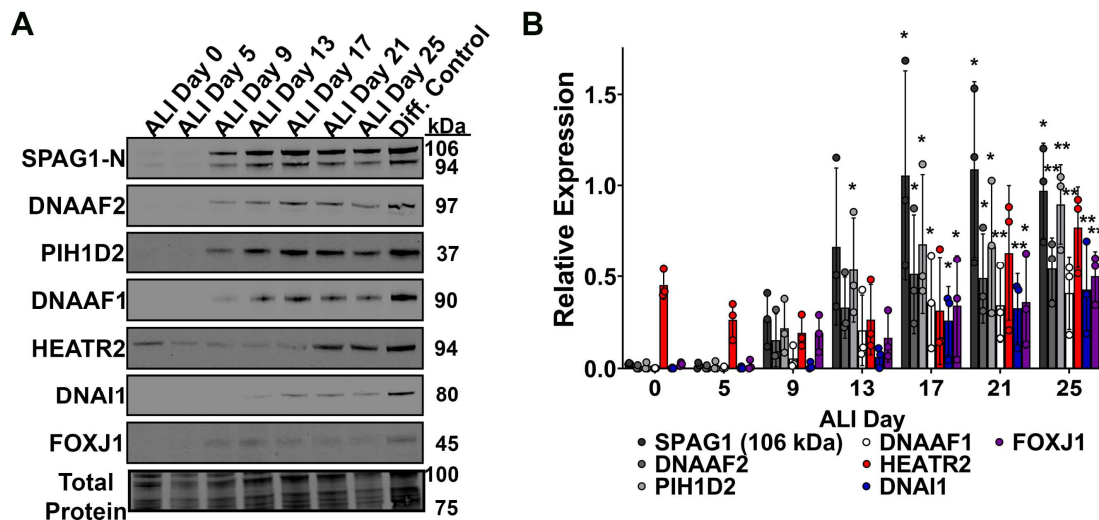
Figure 3.10 Co-expression of SPAG1 and potential interactors at the protein level

Representative immunoblots **(A)** and quantification **(B)** of protein co-expression of SPAG1, DNAAF2, PIH1D2, DNAAF1, HEATR2, DNAI1 and FOXJ1 in differentiating HBEC culture lysates. DNAI1 and FOXJ1 were used as ciliogenesis controls. Total protein stain was used as a loading control. Raw signals were normalized to a total protein stain and then a separate differentiated HBEC lysate control (Diff.

Control). Data are mean ± s.d.; $n = 3$ independent experiments using three distinct cell codes; *P<0.05,

**P<0.01 compared to ALI day 0 (Kruskal–Wallis with uncorrected Dunn's multiple comparisons).

3.3.5 60-kDa SPAG1 has distinct protein interactions from full-length SPAG1

To determine whether there were proteins more likely to interact with the 60-kDa isoform instead of other isoforms of SPAG1, and to further characterize its interactions with DNAAF1, DNAAF2, and PIH1D2, we performed a time course of co-immunoprecipitations using an anti-SPAG1 antibody targeting the C-terminus in differentiating HBECs on ALI days 0, 13, 18 and 28, and analyzed the eluates by immunoblotting and mass spectrometry (Figure 3.11; Table 3.5 – 3.7; Appendix 5 - 6). On ALI day 0, the 60-kDa SPAG1 co-precipitated with MISP, DBN1, IMPDH2, RPN2 and SVIL: proteins associated with actin binding according to Database for Annotation, Visualization and Integrated Discovery (DAVID) gene ontology analysis (Table 3.6 – 3.7). DNAAF2, DNAAF1 and PIH1D2 co-precipitated with SPAG1 by ALI day 13 or 18, and maintained interactions with SPAG1 on ALI day 28 (Figure 3.11). WDR92 co-precipitated with SPAG1, detected by proteomic analysis, starting on ALI day 13. Interestingly, proteomic analysis revealed that a number of dynein chains co-precipitated with SPAG1 on ALI day 18, including DNAH6, DNAH9, DNAH5 and DNAI1. In addition, multiple chaperones co-precipitated with SPAG1, starting at ALI day 13 or 18, including HSPA4, HSPH1 and several components of the chaperonin CCT (Table 3.5). In summary, full-length SPAG1 interacts with DNAAF1, DNAAF2, PIH1D2 and several other DNAAFs, dynein chains, and multiple components of chaperone complexes, including the R2TP/Prefoldin-like and CCT complexes. In contrast, the 60-kDa SPAG1 interacts with proteins involved in actin cytoskeleton and regulation of interactions between actin and microtubules, including MISP, DBN1, IMPDH2, RPN2 and SVIL.
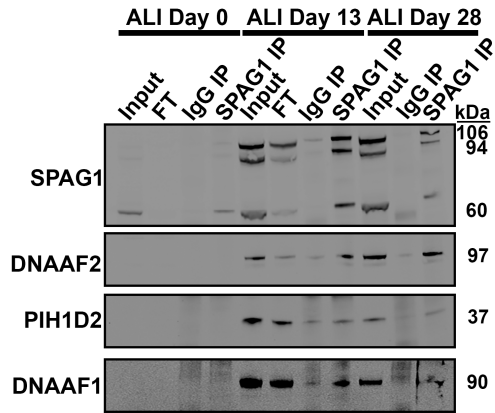
Figure 3.11 SPAG1 interacts with DNAAF2, PIH1D2, and DNAAF1 by ALI day 13

Representative immunoblots for SPAG1, DNAAF2, PIH1D2 and DNAAF1 in time-course co-immunoprecipitation for endogenous SPAG1 samples. HBEC cultures were lysed at various days (ALI days 0, 13, 18 and 28) throughout differentiation, and lysates were used to immunoprecipitate endogenous SPAG1 and interactors. *n* = 2 independent experiments using two distinct cell codes. IP, immunoprecipitation. FT, flow-through. Input lanes represent 5% of the total lysate.

Table 3.5 Identified Interactors of SPAG1 Involved in Axonemal Dynein Arm Assembly in Differentiating Human Bronchial Epithelial Cells[1]

Identified Proteins	Gene ID	ALI Day 0			ALI Day 13			ALI Day 18		
		Fold Change[2]	IgG Spectral Counts (Unique Peptides)	SPAG1 Spectral Counts (Unique Peptides)	Fold Change[2]	IgG Spectral Counts (Unique Peptides)	SPAG1 Spectral Counts (Unique Peptides)	Fold Change[2]	IgG Spectral Counts (Unique Peptides)	SPAG1 Spectral Counts (Unique Peptides)
Dynein Chains										
Dynein heavy chain 6, axonemal	DNAH6	INF	0 (0)	1 (1)	-	-	-	INF	0 (0)	3 (2)
Dynein heavy chain 9, axonemal	DNAH9	-	-	-	-	-	-	INF	0 (0)	3 (2)
Dynein heavy chain 5, axonemal	DNAH5	-	-	-	-	-	-	5	1 (1)	5 (3)
Dynein intermediate chain 1, axonemal	DNAI1	-	-	-	-	-	-	4	2 (2)	8 (7)
Dynein Arm Assembly Factors										
Kintoun	DNAAF2	-	-	-	INF	0 (0)	3.6 (5)	INF	0 (0)	22 (16)
WD repeat-containing protein 92	WDR92	-	-	-	INF	0 (0)	4.3 (5)	INF	0 (0)	15 (11)
Dynein assembly factor 1, axonemal	DNAAF1	-	-	-	-	-	-	INF	0 (0)	11 (10)
RNA polymerase II-associated protein 3	RPAP3	-	-	-	-	-	-	INF	0 (0)	2 (2)
Chaperones										
Heat shock 70 kDa protein 4	HSPA4	1.69[3]	11.6 (9)[3]	19.6 (14)[3]	1.57	2.3 (4)	3.6 (5)	11	1 (1)	11 (9)
T-complex protein 1 subunit θ	CCT8	0.97[3]	17.9 (13)[3]	17.4 (12)[3]	5.01	0.6 (1)	2.9 (4)	2.25	4 (4)	9 (8)
T-complex protein 1 subunit ε	CCT5	1.02	2.3 (1)	2.4 (1)	3.76	0.6 (1)	2.1 (3)	-	-	-
T-complex protein 1 subunit β	CCT2	1.24	17 (18)	21 (20)	2.51	0.6 (1)	1.4 (2)	2.5	2 (2)	5 (4)
T-complex protein 1 subunit ζ	CCT6A	1.02	2.3 (1)	2.4 (1)	2.51	0.6 (1)	1.4 (2)	-	-	-
Heat shock protein 105 kDa	HSPH1	1.42	12 (9)	17 (13)	-	-	-	3	5 (4)	15 (11)

[1]This table shows 14 out of 505 positive hits. The complete list of positive hits is in Appendix 5, and DAVID gene ontology analysis results are in Appendix 6.

[2]Fold changes were calculated using the spectral counts of the SPAG1 and IgG samples. Positive hits were narrowed down using a fold change of at least two and at least two unique peptides detected in at least one time point.

[3]Fold changes, spectral counts, and unique peptides are averages of two ALI Day 0 replicates.

Table 3.6 Identified Interactors of SPAG1 in Undifferentiated Human Bronchial Epithelial Cells

Identified Proteins	Gene ID	Replicate #1			Replicate #2		
		Fold Change[1]	IgG Spectral Counts (Unique Peptides)	SPAG1 Spectral Counts (Unique Peptides)	Fold Change[1]	IgG Spectral Counts (Unique Peptides)	SPAG1 Spectral Counts (Unique Peptides)
Mitotic interactor and substrate of PLK1	MISP	3.07	2.3 (1)	7.2 (3)	2.5	2 (2)	5 (5)
Drebrin	DBN1	2.04	7.0 (3)	14.4 (5)	2	1 (1)	2 (2)
Inosine-5'-monophosphate dehydrogenase 2	IMPDH2	2.04	2.3 (1)	4.8 (2)	2	5 (5)	10 (10)
Dolichyl-diphosphooligosaccharide- protein glycosyltransferase subunit 2	RPN2	2.04	2.3 (1)	4.8 (2)	2	5 (6)	10 (8)
Supervillin	SVIL	2.04	2.3 (1)	4.8 (2)	2	2 (2)	4 (4)

[1]Fold changes were calculated using the spectral counts of the SPAG1 and IgG samples. Positive hits were narrowed down using a fold change of at least two and at least two unique peptides detected in both ALI day 0 co-IP replicates.

Table 3.7 DAVID Gene Ontology Analysis on SPAG1 Interactors in Undifferentiated Human Bronchial Epithelial Cells[1,2]

Term	Genes	Count	%	P-Value	Benjamini
Actin-binding	DBN1, MISP, SVIL	3	50	1.70E-03	7.60E-02
Cell junction	DBN1, MISP, SVIL	3	50	1.00E-02	2.10E-01
Cytoplasm	DBN1, IMPDH2, MISP, SPAG1, SVIL	5	83.3	1.20E-02	1.70E-01
Cell cortex	DBN1, MISP	2	33.3	3.30E-02	6.60E-01

[1]DAVID gene ontology analysis on proteins identified as positive hits in endogenous SPAG1 co-immunoprecipitations in undifferentiated human bronchial epithelial cells (ALI day 0), listed in Table 3.6. [2]A functional annotation chart was created using the categories, GOTERM_BP_DIRECT and GOTERM_MF_DIRECT, with the background set to Homo Sapiens.

3.3.6 DNAAF2 can still interact with 60-kDa SPAG1

We previously characterized the 60-kDa SPAG1 isoform and showed that this isoform contains the C-terminal half of the protein with two TPR domains and the RPAP3-like C-terminal domain. Due to containing the same amino acid sequence and domains as the full-length SPAG1, there is a possibility that this 60-kDa could interact with DNAAFs. To further test this hypothesis, we explored whether the 60-kDa isoform could interact with another DNAAF that interacts with full-length SPAG1. DNAAF2 was the most abundant DNAAF that co-precipitated with SPAG1 in our proteomic studies. Therefore, FLAG-tagged full-length SPAG1 or a FLAG-tagged 60-kDa SPAG1 and HA-tagged DNAAF2 were co-

expressed in HEK293T cells. Immunoprecipitions for the FLAG-tag were performed, and the levels of

HA-tagged DNAAF2 that co-precipitated were analyzed and quantified by immunoblotting. FLAG-

tagged full-length SPAG1 co-precipitated the HA-tagged DNAAF2, suggesting that the addition of the

epitope tags to the C-terminus did not interfere with the function of the protein or its interactions.

Interestingly, the FLAG-tagged 60-kDa isoform of SPAG1 also co-precipitated HA-tagged DNAAF2,

albeit at a level of $7.7 \pm 5.8\%$ (mean $\pm$ s.d.) of interaction compared to full-length SPAG1 (Figure 3.12).

Thus, the 60-kDa SPAG1 isoform can still interact with DNAAFs, and the interaction site for DNAAF2 is

most likely in the C-terminal half of SPAG1. However, a caveat to this experiment is that this interaction

may be driven through the overexpression of constructs, and more known interactors of SPAG1 should be

tested to determine whether the 60-kDa isoform of SPAG1 can interact with them as well.
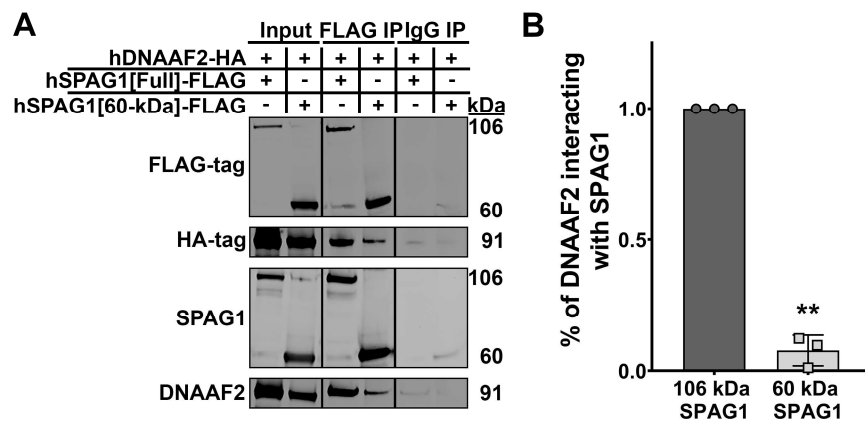
Figure 3.12 The 60-kDa SPAG1 isoform can still interact with DNAAF2

(A) Immunoblots for SPAG1, DNAAF2, FLAG-tag and HA-tag to analyze co-immunoprecipitations of

FLAG-tagged full-length or 60-kDa SPAG1 with HA-tag DNAAF2 in HEK293T cells. IP,

immunoprecipitations. Input lanes represent 5% of the total lysate. **(B)** Quantification of the amount of

DNAAF2 interacting with 106-kDa or 60-kDa SPAG1 isoforms. Background-subtracted HA-tag signals

were divided by FLAG-tag signals per each sample, and then the 106-kDa SPAG1 ratio was set to 1.0.

Data are mean $\pm$ s.d.; $n = 3$ independent experiments; **$P<0.01$ compared to 106-kDa SPAG1 (two-tailed

one-sample t-test).

3.3.7 CS domain in DNAAF2 is required for SPAG1 interaction

Next, we wanted to determine if the conserved domains in DNAAF2 were necessary for the interaction between SPAG1 and DNAAF2. DNAAF2 has two conserved domains, the PIH1 domain (amino acids #10 - #182) and the CS domain (amino acids #254 - #349) (Maurizy et al., 2018). Using site-directed mutagenesis to create deletions in the pcDNA3.1+-hDNAAF2[WT]-HA plasmid, we created plasmids that expressed mutant DNAAF2 either without the PIH1 domain or the CS domain. The DNAAF2 constructs without the PIH1 domain or without the CS domain were able to be expressed, as examined by immunoblotting, and had the correct molecular weights of approximately 73-kDa or 81-kDa, respectively. FLAG-tagged full-length SPAG1 and either a WT or mutant HA-tagged DNAAF2 were co-expressed in HEK293T cells. Immunoprecipitations for the HA-tag were performed, and the levels of FLAG-tagged SPAG1 that co-precipitated were analyzed and quantified by immunoblotting. Negative controls included non-transfected HEK293T cells and HEK293T cells that were co-expressing WT FLAG-tagged SPAG1 and a HA-tagged GFP. The levels of FLAG-tagged SPAG1 increased 3-fold when interacting with a mutant DNAAF2 that lacks the PIH1 domain compared to the levels of interaction with WT DNAAF2. Interestingly, the interaction between DNAAF2 and SPAG1 was completely abolished in absence of the CS domain in DNAAF2 (Figure 3.13). Therefore, the CS domain is necessary for the interaction between DNAAF2 and SPAG1. The PIH1 domain possibly has a negative impact on the interaction between DNAAF2 and SPAG1 due to blocking access to the CS domain, which could explain the increase in interaction levels between SPAG1 and DNAAF2 when lacking the PIH1 domain.
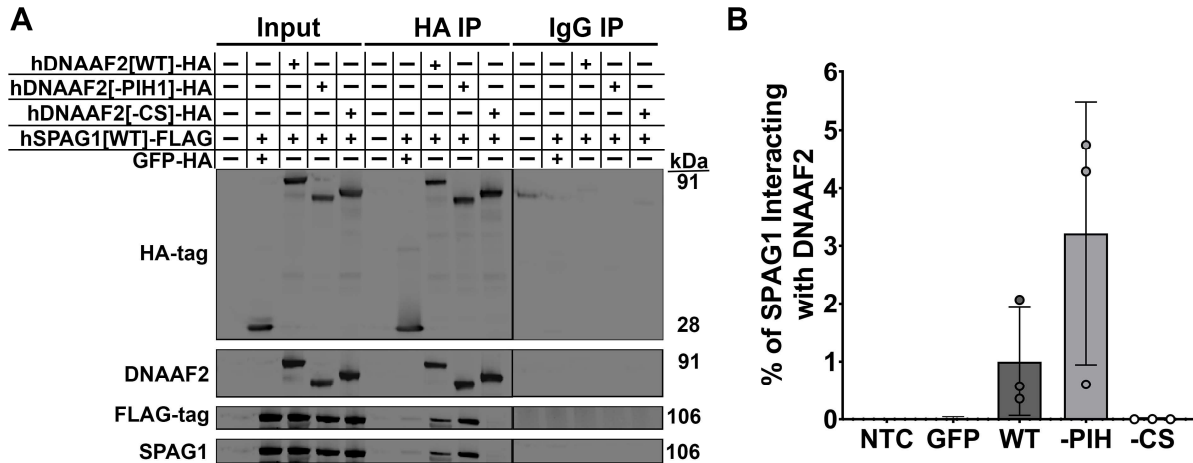
Figure 3.13 CS domain in DNAAF2 is required for its interaction with SPAG1

(A) Immunoblots for HA-tag, DNAAF2, FLAG-tag, and SPAG1 to analyze co-immunoprecipitations of

FLAG-tagged full-length SPAG1 and WT or mutant HA-tagged DNAAF2 in HEK293T cells. IP,

immunoprecipitations. Input lanes represent 5% of the total lysate. **(B)** Quantification of the amount of

SPAG1 interacting with WT or mutant DNAAF2. Background-subtracted FLAG-tag signals were divided

by HA-tag signals per each sample, and then the WT DNAAF2 ratio was set to 1.0. Data are mean ± s.d.;

n = 3 independent experiments; (Kruskal-Wallis test with uncorrected Dunn's multiple comparisons).

3.3.8 C-terminus in SPAG1 is required for DNAAF2 interaction

Finally, we examined if the conserved domains in SPAG1 were necessary for the interaction

between SPAG1 and DNAAF2. SPAG1 has four conserved domains, three TPR domains (amino acids

#209 - #309, #445 - #554, and #623 - #724) and the RPAP3-like C-terminal domain (amino acids #802 -

#891) (Maurizy et al., 2018). Using site-directed mutagenesis to create deletions in the pcDNA3.1+-

hSPAG1[WT]-HA plasmid, we created plasmids that expressed mutant SPAG1 either without the

individual TPR domains or the RPAP3-like C-terminal domain. Due to the similarities of both PIH1D1

and PIH1D2 interacting with the region between the last TPR domain and the RPAP3-like C-terminal

domain in RPAP3 and SPAG1, respectively, we explored if DNAAF2, another PIH1 domain-containing

protein, could interact in this region (amino acids #732 - #784) by creating a construct to express SPAG1

without this region (Maurizy et al., 2018). All of the mutant SPAG1 constructs without the individual

TPR domains, the PIH1D2 interaction site, and the RPAP3-like C-terminal domain were able to be

expressed, as examined by immunoblotting, and had the correct molecular weights of approximately 93-

kDa, 93-kDa, 93-kDa, 99-kDa, and 94-kDa, respectively. HA-tagged full-length DNAAF2 and either a

WT or mutant FLAG-tagged SPAG1 were co-expressed in HEK293T cells. Immunoprecipitations for the

FLAG-tag were performed, and the levels of HA-tagged DNAAF2 that co-precipitated were analyzed and

quantified by immunoblotting. Negative controls included non-transfected HEK293T cells and HEK293T

cells that were co-expressing WT FLAG-tagged SPAG1 and a HA-tagged GFP. Interestingly, the mutant

SPAG1 maintained various levels of interaction with DNAAF2 depending on the domain deleted (Figure

3.14A). Levels of interaction between SPAG1 and DNAAF2 seemed to decrease as the domain deleted

was further downstream towards the C-terminus. There was a significant reduction in the interaction

between SPAG1 and DNAAF2 when SPAG1 was lacking the RPAP3-like C-terminal domain (Figure

3.14B). These results could suggest multiple domains near the C-terminus of SPAG1 are involved in the

interaction between DNAAF2 and SPAG1 to scaffold the proposed R2SD complex, as was suggested by

Horani and colleagues for SPAG1 and HEATR2 (Horani et al., 2018). Another possibility is that the

interaction site for DNAAF2 in SPAG1 is not contained in a conserved domain and was not deleted and

examined in these experiments. Further studies are required to pinpoint the exact amino acid sequences

that is required for the interaction between SPAG1 and DNAAF2.
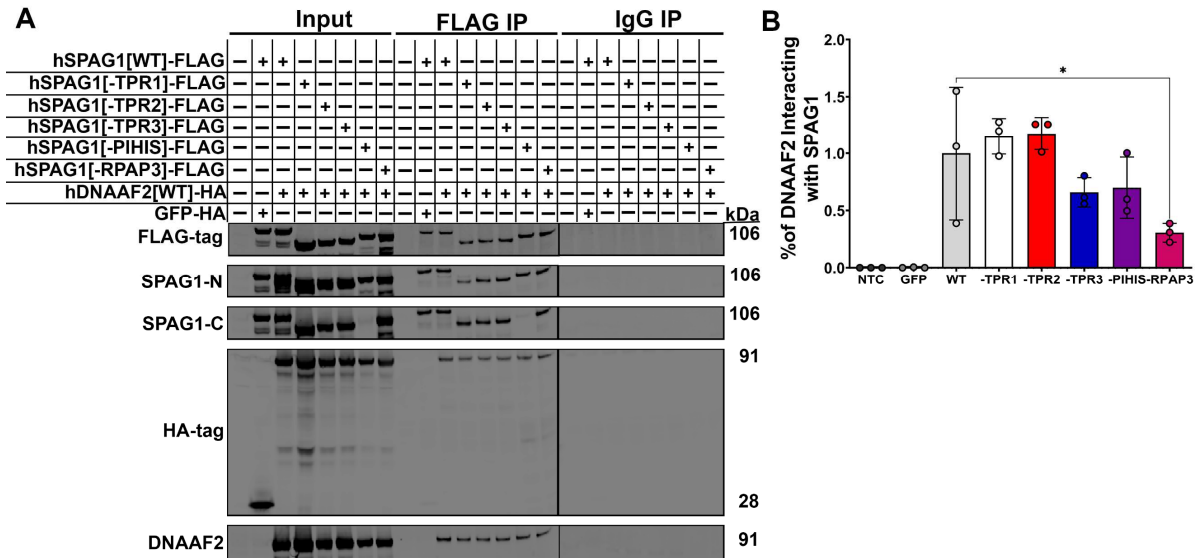
Figure 3.14 Multiple domains in the C-terminus of SPAG1 could be required for its interaction with DNAAF2

(A) Immunoblots for FLAG-tag, N-terminus of SPAG1, C-terminus of SPAG1, HA-tag, and DNAAF2 to analyze co-immunoprecipitations of HA-tagged full-length DNAAF2 and WT or mutant FLAG-tagged SPAG1 in HEK293T cells. IP, immunoprecipitations. Input lanes represent 5% of the total lysate. **(B)** Quantification of the amount of DNAAF2 interacting with WT or mutant SPAG1. Background-subtracted HA-tag signals were divided by FLAG-tag signals per each sample, and then the WT SPAG1 ratio was set to 1.0. Data are mean ± s.d.; $n = 3$ independent experiments; *$P<0.05$ compared to WT (Kruskal-Wallis test with uncorrected Dunn's multiple comparisons).

3.4 Discussion

3.4.1 Full-length SPAG1 expression corresponds with ciliogenesis, while 60-kDa SPAG1 is constitutively expressed

SPAG1 was first isolated as a 528 amino acid protein (~75-kDa) with a cDNA of 2.4 kb [GenBank AF311312.1 1,389 – 3,818 bp] using serum of an infertile woman that caused sperm agglutination (Zhang et al., 1992). This transcript was later considered a truncated form of SPAG1 when the full-length cDNA was discovered as a 3,818 bp transcript with a reading frame of 2,778 bp encoding a

926 amino acid protein (106-kDa) (Lin et al., 2001). Additionally, SPAG1 is also commonly expressed as an 857 amino acid protein (94-kDa) produced from alternative splicing out of exons 11 and 12 (Neesse et al., 2007). According to the Universal Protein Database, SPAG1 also has another isoform that produces a 416 amino acid protein (48-kDa) that is missing amino acids #417 – #926 of the canonical full-length protein [Uniprot Q07617-2]; however, we have not identified this isoform being expressed in airway epithelium. Instead, another isoform of SPAG1 that is a 545 amino acid protein (60-kDa) corresponding to amino acids #382 – #926 of full-length SPAG1 with a cDNA of 2.6 kb was discovered. This isoform was previously shown to be expressed in human testis extracts and HBEC lysates by immunoblots, but was not fully characterized until now (Lin et al., 2001; Neesse et al., 2007; Knowles et al., 2013). 5' RACE and Kozak sequence analysis suggests this 60-kDa isoform is produced from an alternate start site with a 5' UTR including 79 bp of IVS10 [NG_033834.2 59,977 – 60,055 bp] and 47 bp of exon 11 [NG_033834.2 60,056 – 60,102 bp] of *SPAG1* and an alternate start codon at the methionine at amino acid #382 of full-length SPAG1. When translated, the 60-kDa isoform contains six TPR motifs, arranged in two TPR domains, and the RPAP3 C-terminal domain, whereas full-length SPAG1 has nine TPR motifs, arranged in three TPR domains. The mouse homolog of Spag1, also known as Tpis, also has distinct isoforms: the full-length protein has 901 amino acids (101-kDa) with 8 TPR motifs [GenBank AF181252; Uniprot Q80ZX8-2], while the second isoform is transcribed from an alternative start site at exon 11 and translates as a 529 amino acid protein (59-kDa) with 5 TPR motifs [GenBank AF181253; Uniprot Q80ZX8-2] (Takaishi et al., 1999). These two isoforms correspond with the human full-length 106-kDa and 60-kDa isoforms of SPAG1, suggesting the identified 60-kDa SPAG1 protein is a conserved isoform transcribed from a conserved alternative start site.

Differences in the expression patterns of the distinct SPAG1 isoforms were examined in differentiating HBEC cultures using droplet digital PCR, immunoblots, and immunofluorescence. The expression of full-length 106-kDa SPAG1, along with the 94-kDa isoform, correlates with ciliogenesis occurring in HBEC cultures, where there are low levels of mRNA and protein in undifferentiated cells (ALI Day 0 & 5) but increasing expression in differentiating and differentiated cells (ALI Day 9 – 26). It

was previously suggested that SPAG1 is co-expressed with HEATR2 and DNAAF2 at an earlier time point in HBEC differentiation than other DNAAFs (Horani et al., 2018). Here, full-length SPAG1 is co-expressed with DNAAF1, DNAAF2, PIH1D2, and FOXJ1 starting around ALI Day 9, before DNAI1 expression, but after HEATR2 expression. However, it is possible previous methods were detecting the 94-kDa isoform or the 60-kDa isoform of SPAG1, as the 60-kDa isoform is constitutively expressed at similar levels throughout differentiating HBECs (ALI Day 0 – 26). Since 60-kDa SPAG1 is constitutively expressed, it possibly has a motile cilia-independent function in normal epithelium.

Previously, immunoblots for full-length SPAG1 identified it as only being present in the cytoplasm of differentiated HBECs, but not in the ciliary axoneme (Knowles et al., 2013). Immunofluorescence on whole mount and isolated multiciliated human airway epithelial cells confirmed this finding, where SPAG1 localized as large puncta, or foci, in the cytoplasm of airway epithelial cells but not in ciliary axonemes. This localization pattern is similar to other DNAAFs that have been found to co-localize in these cytoplasmic foci, known as Dynein Assembly Particles (DynAP). DynAP are postulated to be liquid-like organelles that concentrate DNAAF, dynein chains, and chaperones together to facilitate dynein arm assembly. SPAG1 has been shown to co-localize with DNAAF2, DNAAF5 (HEATR2), ZMYND10, and RUVBL2 in DynAP previously (Horani et al., 2018; Huizar et al., 2018; Lee et al., 2020).

To further elucidate the function of the 60-kDa SPAG1 isoform, co-immunoprecipitations for endogenous SPAG1 were performed on undifferentiated HBECs (ALI Day 0), which is when the 60-kDa isoform is the major form of SPAG1 expressed. Five proteins were identified as positive hits (MISP, DBN1, IMPDH2, RPN2, and SVIL), and DAVID gene ontology analysis resulted in statistically enriched terms including actin-binding, cell junction, cytoplasm, and cell cortex. The majority of these proteins identified are involved in regulating the actin cytoskeleton and/or the interaction of actin with microtubules, either during mitotic/meiotic spindle assembly, cytokinesis, focal adhesion initiation, or cell migration (Maier et al., 2013.; Kasioulis et al., 2017.; Duan et al., 2018; Smith et al., 2010). Interestingly, the 60-kDa Spag1 isoform in mice has been implicated in oocyte meiotic progression, specifically by

associating with meiotic spindles and being necessary for M-phase entry, spindle morphogenesis, cortical granule-free domains, actin caps, and contractile ring formation (Huang et al., 2016). In addition, SPAG1 expression has been found to be highly up-regulated in multiple cancers, including pancreatic adenocarcinomas, seminomas, renal, breast, and colon, suggesting an oncogenic effect of high expression of SPAG1 (Neesse et al., 2007; Biermann et al., 2007; Siliņa et al., 2011; Lin et al., 2021). Numerous studies have previously shown that knock down of *SPAG1* causes impairment of cell proliferation and/or cell motility. Panc1 cells, a human pancreatic cancer cell line, treated with SPAG1 siRNA had impaired migration of cells in transwell migration and wound healing assays (Neesse et al., 2007). In MDA-MB-231 and T47D cells, which are breast cancer cell lines, treated with shRNA to deplete SPAG1, cell proliferation, colony formation, and cell cycle progression were significantly reduced compared to control cells (Lin et al., 2021). In porcine immature Sertoli cells, depletion of SPAG1 by siRNA caused decreased expression of c-MYC, CCNE1, CCND1, and CDK4, proteins that regulate the cell cycle, inhibition of G1/S phase transition, and a reduced growth rate. Additionally, immunofluorescence studies found SPAG1 co-localizing with α-tubulin in bipolar spindles and spindle midzones during metaphase and telophase of gametogenesis (Hu et al., 2017). Taken together, the 60-kDa isoform of SPAG1 is hypothesized to function as a regulator of microtubule and actin dynamics to facilitate cell motility and proliferation, but further studies are required to confirm and explore this theory. Additionally, it is also possible that the 60-kDa SPAG1 plays a role in the assembly of a subset of ODAs in normal airway epithelium.

3.4.2 SPAG1 scaffolds R2TP/Prefoldin-like complexes

To examine the mechanisms by which SPAG1 facilitates dynein arm assembly, co-immunoprecipitations for endogenous SPAG1 were performed on differentiating and differentiated HBECs. Intriguingly, two dynein heavy chains, DNAH6 and DNAH5, previously identified or suspected DNAAF, including DNAAF2, PIH1D2, DNAAF1, RPAP3, WDR92, DNAAF3, RUVBL1, and RUVBL2, as well as a heat shock protein, HSPH1, co-precipitated with SPAG1 in differentiated HBECs. In addition, various dynein chains, including DNAH6, DNAH9, DNAH5, and DNAI1, and chaperones,

including HSPA4 (HSP70), HSPH1, and several components of the CCT complex co-precipitated with endogenous SPAG1 in differentiating HBECs. The interactions, either direct or indirect, between SPAG1 and DNAAF1, DNAAF2, and PIH1D2 were further corroborated by reverse co-IPs in HBECs for endogenous DNAAF1, DNAAF2, and PIH1D2 co-precipitating SPAG1 and each other, as well as co-expression of these proteins during HBEC differentiation determined by ddPCR and immunoblots. Previously, SPAG1 was also shown to interact with DNAAF5 (HEATR2) by immunoprecipitation of EGFP-tagged HEATR2 in human tracheal epithelial cells and FRET analysis in HEK293T cells co-expressing fluorescent tagged-SPAG1 and HEATR2 (Horani et al., 2018). However, in this study, HEATR2 was not identified as a positive hit in any of the endogenous SPAG1 immunoprecipitations analyzed by either mass spectrometry or immunoblots. This suggests that the identified interaction between SPAG1 and HEATR2 is a transient interaction or actually does not occur endogenously in human airway epithelium.

All of the DNAAFs that were identified to be interactors of SPAG1 have been shown to cause PCD in humans or are canonical R2TP complex components that result in dynein arm defects in multiple model systems when defective. DNAAF1 (LRRC50), DNAAF2 (KTU), and DNAAF3 are previously identified DNAAFs in humans, and mutations in these genes cause PCD with defects in ODAs and IDAs (Loges et al., 2009; Omran et al., 2012; Mitchison et al., 2012). RPAP3, RUVBL1, RUVBL2, and WDR92 are canonical components of the HSP90 co-chaperone R2TP/Prefoldin-like complex, also known as the PAQosome. The R2TP complex contains a hexamer of the AAA+ ATPases RUVBL1 and RUVBL2 bound to a heterodimer of RPAP3 and PIH1D1, and WDR92 interacts with PIH1D1 and RPAP3 likely to stabilize this heterodimer. The PAQosome facilitates the formation and stabilization of multiprotein complexes, including L7Ae ribonucleoproteins, U5 small nuclear ribonucleoprotein, RNA polymerase II, phosphatidylinositol-3-kinase-related proteins (PIKKs), etc., by acting as a platform between chaperones, including HSP90, HSP70, and CCT, and their clients, either directly or through interaction with specificity factors or adaptors (Houry et al., 2018; Lynham and Houry, 2018). Previously, deletions of *RUVBL1*/*Pontin* or *RUVBL2*/*Reptin* in zebrafish, targeted ciliated cells in mice, or

Drosophila testes led to ciliary motility defects due to absent or defective ODAs and IDAs (Li et al., 2017; Dafinger et al., 2018; Zhao et al., 2013; zur Lage et al., 2018). RNAi depletion of the homologue of RPAP3, *spaghetti*, in *Drosophila* spermatocytes led to male infertility due to loss of dynein arms (zur Lage et al., 2018). *WDR92* RNAi depletion in *Drosophila* and planarian *Schmidtea mediterranea*, as well as a *wdr92-1* mutant stain of *Chlamydomonas reinhardtii*, also have reduced ciliary motility due to dynein arm defects (Patel-King et al., 2016; zur Lage et al., 2018; Liu et al., 2019; Patel-King et al., 2019). PIH1D2 is a protein in the PIH1-domain-containing protein family, which includes PIH1D1, DNAAF6 (PIH1D3), and DNAAF2 (KTU). DNAAF6 (PIH1D3) and DNAAF2 (KTU) are PCD-associated genes due to causing dynein arm defects in humans, while all of these proteins have been found to cause distinct dynein arm defects in zebrafish mutant models (Olcese et al., 2017; Omran et al., 2012; Yamamoto et al., 2010; Yamaguchi et al., 2018). Intriguingly, PIH1D2 expression is consistent with other DNAAF in human bronchial epithelial cells determined by ddPCR and immunoblots, suggesting that the function of PIH1D2 in dynein arm assembly is conserved in humans. Thus, all of these proteins have been implicated in axonemal dynein arm assembly in various model organisms.

Interestingly, it has been theorized, based on similarities in protein domains and phylogenetic analysis, that SPAG1 is a paralogue of RPAP3 (Maurizy et al., 2018). FRET analysis in HEK293T cells has previously shown that SPAG1 and DNAAF2 directly interact, and quantitative LUMIER co-IP assays in HEK293T cells have identified direct interactions between DNAAF2 and PIH1D2 with SPAG1 (Horani et al., 2018; Maurizy et al., 2018). SPAG1 was also found to interact with RUVBL1-RUVBL2, and this interaction is further stabilized by the addition of PIH1D2. In *Drosophila* S2 cells, Wdr92 was also shown to interact with Spag1, with their association increased by overexpression of Pontin/Ruvbl1 and Reptin/Ruvbl2 (zur Lage et al., 2018). Thus, SPAG1 is suggested to form an R2TP-like complex, called R2SP, containing a hexamer of RUVBL1-RUVBL2 with a heterodimer of SPAG1 and PIH1D2. Studies *in vitro* hypothesized the existence of another R2TP-like complex, called R2SD, that replaces PIH1D2 with DNAAF2; however, this complex was not detected in proteomic experiments *in vivo* (Maurizy et al., 2018). Here, we provide the first evidence that the R2SP complex, and possibly the R2SD

complex, occur endogenously in human airway epithelia by co-immunoprecipitating all components of these complexes with SPAG1. However, further studies are required to determine whether there are two or three complexes occurring separately, which proteins are included in these complexes, and which complex is necessary for which dynein arm subtype.

Previous studies suggest that the canonical R2TP complex selects and specifies substrates and clients via CK2-phosphorylated sites interacting with PIH1D1 (Hořejší et al., 2014; von Morgen et al., 2015). Due to the specificity of defects in dynein arms when examining PIH1 protein null mutants in zebrafish, it's possible the R2TP-like complex substitutes the PIH1 protein with others depending on the specific dynein subtype being assembled, but studies are required to confirm and explore this further (Yamaguchi et al., 2018). The canonical R2TP complex and other possible R2TP-like complexes, such as DNAAF4 -DNAAF2 and DNAAF4-DNAAF6 based on their similar protein domains with the R2TP complex, facilitating the assembly of axonemal dyneins could also increase the amount of complexes demonstrating specificity needed to assemble the multiple dynein chains and subtypes of ODA and IDA (zur Lage et al., 2018; Maurizy et al., 2018).

The R2TP-like complexes are hypothesized to further interact with other DNAAFs, possibly as adaptors or specificity factors for the recruitment of dynein chain clients to the chaperone complexes. In this study, DNAAF1 and DNAAF3 were co-immunoprecipitated with endogenous SPAG1 in HBECs. Although interactions with DNAAF3 have not been described previously, DNAAF1 has been shown to interact with RUVBL1 and RUVBL2 by co-immunoprecipitations and tandem AP-MS assays in HEK293T cells (Hartill et al., 2018). Surprisingly, DPCD, a putative R2TP complex adaptor, also co-precipitated with SPAG1 and has previously been shown to be associated with a PCD phenotype in mice and interact with DNAAF1 and RUVBL1 in NIH-3T3 and HEK293T cells (Cloutier et al., 2017; Zariwala et al., 2004; Dafinger et al., 2018). Other DNAAFs that have potential interactions with R2TP-like complex components, specifically RUVBL1/RUVBL2, include DNAAF4 (DYX1C1), LRRC6, and DNAAF7 (ZMYD10) (Fabczak and Osinka, 2019). Additionally, WDR92 has also been suggested to be a

specificity factor that associates dynein clients to an R2TP-like complex and HSP90 through its interaction with SPAG1 (zur Lage et al., 2018).

Here, we identified various chaperones in endogenous SPAG1 co-immunoprecipitations in HBECs, including HSPA4 (HSP70), HSPH1, and several components of the CCT chaperonin complex. Previously, SPAG1 has been shown to interact with HSP70 and HSP90 through its first and third TPR domain with the (M/I)EEVD motif of HSP70 and HSP90 C-terminal tails (Chagot et al., 2019; Dermouche et al., 2021). RPAP3 is known to interact with both HSP70 and HSP90, possibly to facilitate the exchange of clients between chaperones, and recently, this function has also been suggested for SPAG1 (Benbahouche et al., 2014; Henri et al., 2018; Dermouche et al., 2021). Several DNAAF, other than the canonical PAQosome components, have been shown to potentially interact with HSP70 and HSP90, including DNAAF2 (KTU), DNAAF4 (DYX1C1), DNAAF6 (PIH1D3), and DNAAF7 (ZMYND10), through various methods (Fabczak and Osinka, 2019). Additionally, components of the CCT chaperonin complex, which associates with the canonical R2TP/Prefoldin-like complex through WDR92, have previously been immunoprecipitated with WDR92 and DNAAF4 (DYX1C1) (Cloutier et al., 2017; Liu et al., 2019; Tarkar et al., 2013). Taken together, SPAG1 likely forms R2TP-like complexes with RUVBL1/2, PIH1D2 and/or DNAAF2 that scaffold the interactions of various chaperones, including HSP70, HSP90, HSPH1, and CCT, adaptors, including WDR92, DNAAF1, or DNAAF3, and dynein chain clients (Figure 3.15).
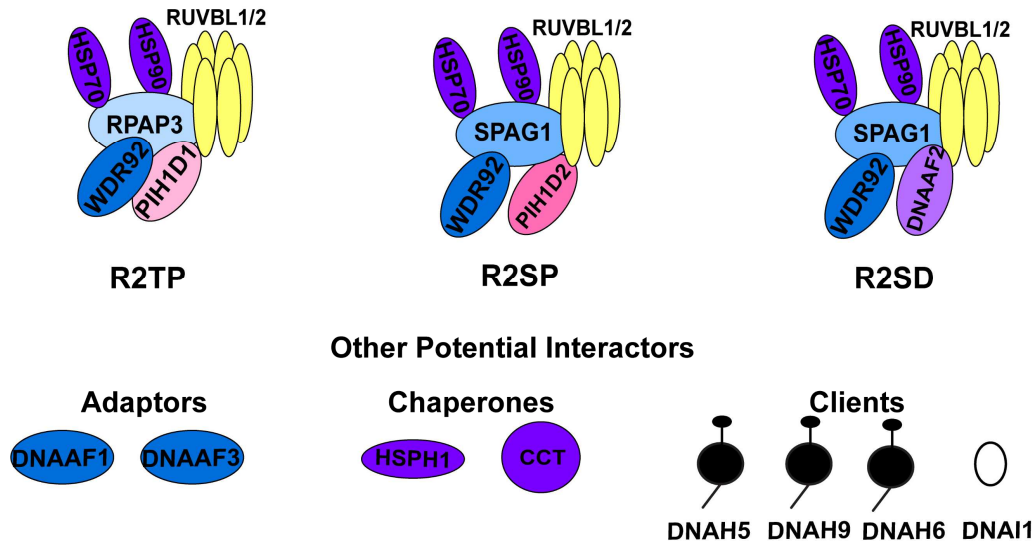
Figure 3.15 Proposed SPAG1 interactors and R2TP-like complexes in human airway epithelia

Schematic showing the components of the canonical R2TP/Prefoldin-like complex (RUVBL1/2-RPAP3-PIH1D1) and the proposed R2TP-like heat shock protein co-chaperone complexes scaffolded by SPAG1, R2SP (RUVBL1/2-SPAG1-PIH1D2) and R2SD (RUVBL1/2-SPAG1-DNAAF2). These two complexes have been theorized to further interact with protein adaptors, including DNAAF1, DNAAF3, and WDR92, to recruit dynein chain clients, including DNAH5, DNAH9, DNAH6, and DNAI1, to chaperones, including HSP70, HSP90, HSPH1, and CCT complex, for assembly of dynein arms in human airway epithelia.

3.4.3 DNAAF2 CS Domain is required to interact with SPAG1, while SPAG1 requires multiple domains to interact with DNAAF2

SPAG1's conserved domains include nine tetratricopeptide repeat (TPR) motifs arranged into three different domains and a RPAP3-like C-terminal domain (Fabczak and Osinka, 2019). TPR motifs give rise to two antiparallel α-helices that fold in a helix-turn-helix conformation and can form scaffolds to mediate protein-protein interactions, particularly with HSPs, that are involved in assembly of multiprotein complexes (Perez-Riba and Itzhaki, 2019). Specifically, HSP70 and HSP90 have been shown to be able to interact with SPAG1 through its first and third TPR domains (Chagot et al., 2019).

Additionally, SPAG1 has been shown to interact with the RUVBL1/2 hexamer (Maurizy et al., 2018). DNAAF2's conserved domains include a N-terminal PIH1 domain, which is present in four proteins all within the PIH1 domain-containing protein family in humans, and a C-terminal CS domain. The CS domain is typically composed of two antiparallel β-sheets with four and three β-strands, also referred to as a β-sandwich fold, and are frequently found in co-chaperones of HSP90 (Lee et al., 2004). Due to the fact that SPAG1 is a paralog of RPAP3 and is similar in its protein domains and protein interactions, it is possible that SPAG1 interacts with DNAAF2 in a similar fashion as RPAP3 interacts with PIH1D1 in the canonical R2TP complex (Fabczak and Osinka, 2019). SPAG1 interacts with HSP70 and HSP90 through its TPR domains, similar to how RPAP3 interacts with HSP70 and HSP90 through its TPR domains (Chagot et al., 2019). PIH1D1 interacts with RPAP3 through a binding motif located between the last TPR domain and the RPAP3 C-terminal domain. P1H1D1 interacts with RPAP3 through its C-terminal CS domain (Martino et al., 2018). Interestingly, PIH1D2 has been shown to interact with SPAG1 through a binding motif (a.a. #732 - #784) located between the third and final TPR domain and the RPAP3 C-terminal domain, similar to the interaction of PIH1D1 with RPAP3 (Maurizy et al., 2018). Due to the similar protein domains in DNAAF2 as PIH1D1 and PIH1D2, it is possible that DNAAF2 can also bind to a motif between the TPR domains and the RPAP3-like C-terminal domain in SPAG1. DNAAF2 could also use its CS domain to interact with SPAG1, similar to how PIH1D1 uses its CS domain to interact with RPAP3.

FRET analysis of the interaction between SPAG1 and DNAAF2 showed a significant increase when fluorescent proteins were present on the C-terminus of SPAG1 (Horani et al., 2018). In addition, using co-immunoprecipitation studies, we have demonstrated that FLAG-tagged 60-kDa SPAG1 can still interact with DNAAF2, albeit weaker compared to full-length SPAG1. Additionally, co-immunoprecipitations for FLAG-tagged mutant SPAG1 showed a greater reduction in the amount of HA-tagged DNAAF2 precipitated when domain deletions became closer to the C-terminus. Specifically, deletion of the RPAP3-like C-terminal domain in SPAG1 caused a significant reduction in the amount of DNAAF2 that co-precipitated compared to the WT interactions. However, none of the domain deletions

caused a complete inhibition of the DNAAF2 and SPAG1 interaction. These findings suggest that multiple domains in the C-terminus of SPAG1 are involved in its interaction with DNAAF2. It is also possible that the DNAAF2 interaction site in SPAG1 was not deleted in these studies. However, a caveat of these studies is that the deletion of the conserved domains could result in the misfolding of either SPAG1 or DNAAF2, which can have an impact on their ability to interact with each other. Co-immunoprecipitation for mutant DNAAF2 showed a complete inhibition of the interaction with FLAG-tagged SPAG1 when DNAAF2 was lacking the CS domain. This suggests that SPAG1 interacts with DNAAF2 through its CS domain, similar to how RPAP3 interacts with PIH1D1 through its CS domain (Martino et al., 2018). Interestingly, in the reverse co-immunoprecipitation for DNAAF1, only the 106-kDa isoform of SPAG1 was pulled down, unlike the DNAAF2 and PIH1D2 reverse co-immunoprecipitations that pulled down both 106-kDa and 94-kDa SPAG1 isoforms. This suggests that the interaction domain(s) for DNAAF1 in SPAG1 is present in the second TPR domain, which corresponds with SPAG1 exons 11 and 12 that are spliced out in the 94-kDa isoform, but this needs to be confirmed. Future studies are required for determining the mechanisms and interaction domains behind the R2SD and R2SP complexes' interactions with other DNAAFs, dynein chains, and chaperones that facilitate the assembly of axonemal dyneins.

CHAPTER 4: SPAG1 IS REQUIRED FOR THE FORMATION OF THE DYNEIN INTERMEDIATE
AND HEAVY CHAIN SUBCOMPLEX

4.1 Introduction[1]

SPAG1, also known as DNAAF13, is suggested to be involved in cytoplasmic assembly of

axonemal dyneins due to its localization in the cytoplasm and dynein arm defects in *SPAG1*-deficient

PCD individuals (Knowles et al., 2013). Further evidence of this is defective motile ciliary phenotypes

found in SPAG1 *Danio rerio* and *Drosophila* mutants due to the absence of ODAs and IDAs (Knowles et

al., 2013; zur Lage et al., 2018). However, studies exploring SPAG1's specific function in axonemal

dynein arm assembly in humans are scarce. Horani and colleagues have suggested that SPAG1 facilitates

axonemal dynein assembly by scaffolding a complex with DNAA5 (HEATR2) and DNAAF2 (KTU) in

an early stage of dynein assembly, and when these proteins are dysfunctional, it leads to accumulation of

large DNAAF5-SPAG1-DNAAF2 aggregates that are eventually degraded, disrupting dynein arm

assembly (Horani et al., 2018). One could also speculate what step SPAG1 is required for the complex

process of axonemal dynein arms through its protein interactions. Due to recent studies suggesting that

SPAG1 forms R2TP-like complexes that can interact with molecular chaperones HSP70 and HSP90, it

can be speculated that SPAG1 is involved in the folding of dynein chains, either the DICs or the DHCs

based on other studies involving protein interactors of SPAG1 (Maurizy et al., 2018; Chagot et al., 2019).

However, these studies were mostly performed in immortalized human cell lines or with exogenous

expression of proteins, with limited focus on examining the defects found in axonemal dynein arms in

respiratory motile cilia when deficient in SPAG1.

---

[1] The results, figures, and text contained in this chapter has previously appeared, in part, in an article
published in the *Journal of Cell Science*. The original citation is as follows:
**Smith, A. J., Bustamante-Marin, X. M., Yin, W., Sears, P. R., Herring, L. E., Dicheva, N. N., López-
Giráldez, F., Mane, S., Tarran, R., Leigh, M. W. et al.** (2022). The role of SPAG1 in the assembly of
axonemal dyneins in human airway epithelia. *J. Cell Sci.* **135**, jcs259512.
https://doi.org/10.1242/jcs.259512

Here, we identified three PCD-affected individuals with genetic variants in *SPAG1*. One of those genetic variants was a variant of uncertain significance (VUS), and RT-PCR studies of this variant demonstrated that it was pathogenic. Two out of the three *SPAG1*-deficient PCD individuals showed an atypical SPAG1 mutant PCD ciliary phenotype with the presence of normal ODAs and some ciliary motility. This abnormal phenotype was attributed to the expression of the 60-kDa isoform of SPAG1 in these two PCD-affected individuals, suggesting that the novel 60-kDa isoform of SPAG1 can partially compensate for the lack of full-length SPAG1 and assemble ODAs. Immunofluorescence staining for DNAI1, DNAH5, and DNALI1 in human nasal epithelial cells confirmed the TEM findings for these PCD patients. A significant decrease in the level of DHCs and the disruption of the addition of the DHCs to the DIC/LC complex in human nasal epithelial cells that lack SPAG1, as shown by immunoblots and co-immunoprecipitations, demonstrate that SPAG1 is required for the folding of the DHCs and/or the addition of the DHCs to the DIC/LC complex.

4.2 Materials and Methods

4.2.1 Overview

In this chapter, we focused on the dynein arm assembly defects found in *SPAG1*-deficient human airway epithelia to elucidate SPAG1's role in the cytoplasmic assembly of axonemal dynein arms. To create a *SPAG1*-deficient human airway epithelial cell culture model, we first tried to knock-out *SPAG1* using clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated endonuclease 9 (CRISPR/Cas9) gene editing technology in normal primary HBECs. Using CRISPR/Cas9 gene editing technology, guide RNA (gRNA) directs Cas9 to a target sequence adjacent to a protospacer adjacent motif where it creates a double-stranded break, and indels are formed by DNA repair from nonhomologous end joining (Alapati and Morrisey, 2017; Everman et al., 2018). To ensure SPAG1 expression was abolished and the CRISPR/Cas9-treated HBECs had a ciliary phenotype, we sequenced the gDNA, analyzed the SPAG1 protein levels by immunoblots, and measured the ciliary beat frequency (CBF) and active area of these cultures. Unfortunately, due to the difficulty of introducing editing reagents into HBECs and expanding a single clone of HBECs without losing their ability to differentiate

into multiciliated cells, our efforts for using CRISPR/Cas9 to completely knock-out *SPAG1* in our HBEC cultures were unsuccessful (Everman et al., 2018). Fortunately, we were able to acquire HNECs from healthy control and *SPAG1*-deficient PCD-affected individuals, and using conditionally reprogrammed cell (CRC) conditions and the commercially available growth media, PneumaCult-Expand Plus and PneumaCult-ALI (Stem Cell Technologies), we were able to expand these HNECs to acquire enough multiciliated cultures to perform our experiments.

Genetic analysis of *SPAG1*-deficient PCD subjects were performed clinically, and transmission electron micrographs of their nasal cilia were previously acquired, which were retroactively analyzed and quantified. Using HNEC cultures, we examined mRNA expression of *SPAG1* using RT-PCR and sequencing to determine the pathogenicity of a novel intronic genetic variant in *SPAG1* identified in one of the subjects. CBF and the level of SPAG1 protein were measured, via immunoblots and immunofluorescence, in these HNEC cultures. Levels of dynein chains were measured using immunoblots and immunofluorescence to further investigate the dynein arm defects in *SPAG1*-deficient HNECs and to determine the step SPAG1 is necessary for in dynein arm assembly. Finally, co-immunoprecipitations for DNAI1 were performed, followed by immunoblots for dynein chains, to determine which axonemal outer dynein arm subcomplexes require SPAG1 to interact.

4.2.2 Subjects

Protocols involving human studies were approved by the Institutional Review Board at the University of North Carolina and were performed in compliance with ethical regulations. Written informed consent was acquired from the PCD-affected individuals, their parents, and other participants. All clinical investigations have been conducted according to the principles expressed in the Declaration of Helsinki.

4.2.3 Genetic Analysis

Using whole-exome sequencing (WES) as described previously (Knowles et al.,2013), genetic variants in *SPAG1* were identified in family UNC-372 by the Zariwala and Knowles labs at UNC-CH. WES was carried out for the probands in families UNC-68 and UNC-1231 at the Yale Center for Genome

Analysis. Briefly, 1 µg of genomic DNA was sheared at ~140 or ~200 bp, captured with Nimblegen

SEqCapEZ Exome Library version 2.0 (Roche) or with a modified Integrated DNA Technologies xGen

Exome Research Panel version 1.0 (Roche), for proband UNC-68 III-2 or UNC-1231 II-1, respectively.

For proband UNC-68 III-2 or UNC-1231 II-1, respectively, captured fragments were sequenced using 76-

bp or 101-bp paired-end sequencing reads with a NovaSeq 2000 according to the manufacturer's

instructions or with an Illumina NovaSeq 6000 with a S4 flowcell according to Illumina protocols.

Sequencing reads were aligned to human genome build 37 (GRCh37/hg19) using the Burrows-Wheeler

aligner Maximal Exact Match (BWA-MEM) (Li, 2013), aggregated into a BAM file, and further

processed to identify variants using the Genome Analysis Toolkit (GATK) version 3.4 (McKenna et al.,

2010) following the GATK best practices workflow (van der Auwera et al., 2013). Variants were

annotated using ANNOVAR (Wang et al., 2010), and MetaSVM (Dong et al., 2015) was used to predict

the deleteriousness of non-synonymous variants. For UNC-68 III-2, an average 61.8% mean fold

coverage was achieved, with 90.4% covered at ≥8X. For UNC-1231 II-1, an average 54.6% mean fold

coverage was achieved, with 98.5% covered at ≥8X. After the removal of duplicates and common

variants [minor allele frequency of <0.1% in the Genome Aggregation database (gnomAD

<https://gnomad.broadinstitute.org/>)], a manual review of the WES dataset was performed for all

currently known PCD-associated genes (Table 4.1). Exonic variants including missense, nonsense,

frameshift, in-frame, 1 bp at the start and end of an exon, and intronic variants ~15 bp on either side of an

exon, including canonical splice-site, were reviewed. Candidate variants were confirmed by PCR

amplification followed by Sanger sequencing.

For UNC-68 III-2, a single pathogenic variant in the *SPAG1* gene was identified. Manual

inspection of BAM files revealed certain gaps in WES, and hence, Sanger sequencing was carried out for

SPAG1 as reported previously (Knowles et al., 2013). This led to the identification of another likely

pathogenic variant. For UNC-1231 II-1, WES analysis identified a single variant of uncertain significance

(VUS) in an intron, close to the extended splice site in SPAG1. PCR to confirm the *POLR2K* and *SPAG1*

deletion in family UNC-1231 was performed as described previously (Knowles et al., 2013). Primers used

for the *POLR2K* and *SPAG1* deletion confirmation sequencing are listed in Appendix 3. Segregation

analysis revealed that genetic variants identified in all three families were inherited in trans (Figure 4.4).

Table 4.1 PCD-associated Genes Checked by WES Screening

#	Current Gene Names (Alias)	Transcript Identifier
1	CCDC103	NM_213607.2
2	CCDC39	NM_181426.1
3	CCDC40	NM_017950.3
4	CCDC65 (DRC2)	NM_033124.4
5	CCNO	NM_021147.3
6	CFAP221 (PCDP1)	NM_001271049.2
7	CFAP298 (C21orf59)	NM_021254.2
8	CFAP300 (C11orf70)	NM_032930.2
9	CFAP57 (WDR65)	NM_001195831.2
10	DNAAF1 (LRRC50)	NM_178452.4
11	DNAAF11 (LRRC6)	NM_012472.3
12	DNAAF2 (KTU)	NM_018139.2
13	DNAAF3	NM_001256714.1
14	DNAAF4 (DYX1C1)	NM_130810.3
15	DNAAF5 (HEATR2)	NM_017802.3
16	DNAAF6 (PIH1D3)	NM_001169154.1
17	DNAH1	NM_015512.4
18	DNAH11	NM_001277115.1
19	DNAH5	NM_001369.2
20	DNAH8	NM_001206927.1
21	DNAH9	NM_001372.3
22	DNAI1	NM_012144.2
23	DNAI2	NM_023036.4
24	DNAJB13	NM_153614.2
25	DNAL1	NM_031427.3
26	DRC1 (CCDC164)	NM_145038.2
27	FOXJ1	NM_001454.4
28	GAS2L2	NM_139285.3
29	GAS8	NM_001481.2
30	HYDIN	NM_032821.2
31	LRRC56 (DNAAF12)	NM_198075.47
32	MCIDAS	NM_001190787.1
33	NEK10	NM_152534.4
34	NME5 (RSPH23)	NM_003551.3
35	NME8 (TXNDC3)	NM_016616.4
36	ODAD1 (CCDC114)	NM_144577.3
37	ODAD2 (ARMC4)	NM_018076.2
38	ODAD3 (CCDC151)	NM_145045.4
39	ODAD4 (TTC25)	NM_031421.2
40	OFD1	NM_003611.2
41	RPGR	NM_000328.2
42	RSPH1	NM_080860.2
43	RSPH3	NM_031924.4
44	RSPH4A	NM_001010892.2
45	RSPH9	NM_152732.4
46	SPAG1 (DNAAF13)	NM_172218.2
47	SPEF2 (KPL2)	NM_024867.4
48	STK36	NM_001243313
49	TP73	NM_005427.4
50	TTC12	NM_017868.4
51	ZMYND10 (DNAAF7)	NM_015896.2

4.2.4 Cell Culture

To create a SPAG1-knock out HBEC culture model using CRISPR/Cas9 gene editing technology, custom lentivirus that expresses an all-in-one construct for *SPAG1* sgRNA and Cas9 (pLenti-U6-sgRNA-SSFV-Cas9-2A-Puro) [Catalog Number K2264015] and a scrambled sgRNA CRISPR/Cas9 All-in-One Lentivirus [Catalog Number K011] was ordered from Applied Biological Materials (abm), Inc. The *SPAG1* sgRNA sequence was *GATAATTTGCCTCCAGTTCG*, which targeted exon 4. Disease-free primary P0 HBECs were obtained from the Marsico Lung Institute Tissue Procurement and Cell Culture Core and were seeded on PureCol-coated 100 mm plastic dish in CRC medium supplemented with 5 μM of the Rho-associated kinase (ROCK) inhibitor, Y-27632 (Sigma-Aldrich), 1X antifungals, and 1X antibiotics. Irradiated 3T3-J2 TripZcon fibroblasts with puromycin resistance were seeded on Purecol-coated 100 mm plastic dishes in Complete Growth Medium (CGM) 12 - 24 hours before transduction and co-culture of HBECs. CGM is composed of DMEM (Gibco) supplemented with 4.5 g/L D-glucose, L-glutamine, 110 mg/L sodium pyruvate, 10% FBS (Corning), and 1X penicillin-streptomycin (Thermo Fisher Scientific), 0.25 μg/mL amphotericin B (Thermo Fisher Scientific), and 0.24 μg/mL fluconazole (Thermo Fisher Scientific). Two days after HBECs were plated, HBECs were passaged using 0.025% trypsin-EDTA followed by soybean trypsin inhibitor, centrifuged, and resuspended in BEGM supplemented with 5 μM Y-27632 (Sigma-Aldrich) and 2 ug/mL polybrene (Sigma-Aldrich). P1 HBECs were divided into 5 separate conical tubes according to the experimental groups: Nontransduced Control with Puromycin, Nontransduced Control without Puromycin, CRISPR Scrambled sgRNA Control, GFP Control, and CRISPR/Cas9 *SPAG1* Knockout. The appropriate volume of lentivirus was added to each experimental HBEC group for a multiplicity of infection of 4, and HBEC solutions were incubated at RT for 10 minutes. Lentivirus transduction was performed by spinoculation at 900 x *g* for 1 hour at 25°C. Transduced HBECs were then seeded onto plates containing the irradiated, puromycin-resistant 3T3-J2 fibroblasts.

Transduced HBECs were washed with PBS and media was changed to CRC+Y medium the day after transduction. Selection of lentivirus-transduced HBECs using 1 ug/mL puromycin started 48 hours

after transduction. Media was changed every other day with one PBS wash per week. Once cultures had become 70 – 90% confluent and the Nontransduced Control with puromycin HBECs had died, HBECs were passaged using the double trypsinization method with 0.025% trypsin-EDTA followed by soybean trypsin inhibitor (Fulcher and Randell, 2013). P2 HBECs were seeded onto human type IV placental collagen-coated 12mm or 30mm Millicell porous support inserts (Millipore) at ~1.5 x 105 cells/cm^2. Once the cultures became confluent, HBECs were cultured at ALI conditions as described previously.

HNECs were obtained from probands UNC-68 III-2, UNC-372 III-1, UNC-1231 II-1 and controls through nasal scrapes as described previously (Müller et al., 2013). Briefly, HNECs were obtained from subjects by drawing a sterile thermoplastic curette (Arlington Scientific) across the mucosal surface of the inferior nasal turbinate visualized with an operating otoscope with a 9 mm reusable polypropylene nasal speculum. Curettes were placed in collection tubes with 10 mL of RPMI 1640 media (Gibco), and rinsed with RPMI 1640 media to dislodge cells. HNEC samples were centrifuged, resuspended in accutase (Sigma-Aldrich), and incubated with agitation at 37°C for 15 minutes. CRC media was added to cells to inactive the accutase, and cells were centrifuged again, resuspended in CRC+Y media, and counted. HNECs were then plated and expanded in CRC conditions, as previously described, or using PneumaCult-Expand Plus medium (Stem Cell Technologies) (Suprynowicz et al., 2012; Bustamante-Marin et al., 2019). Briefly summarizing CRC expansion, 3T3-J2 cells were cultured in CGM; passaged and collected into 50 mL conical tubes; irradiated with a X-RAD 320 biological irradiator (Precision X-Ray Irradiation) with the following parameters: 320.0 KV, 12.50 mA, 3,000 cGy dose, 1.2 mm AI filter, and 47 SSD; and then cryopreserved. Irradiated 3T3-J2 fibroblasts were seeded on PureCol-coated 100 mm culture dishes 12 – 24 hours in CGM before co-culture with HNECs. Processed HNECs were then seeded onto the 100 mm culture dishes with irradiated 3T3-J2, and were expanded in CRC medium supplemented with 5 μM of the ROCK inhibitor, Y-27632 (Sigma-Aldrich). The full composition of CRC media is listed in Appendix 1. HNEC cultures were washed with PBS and the CRC media was replaced every 2 – 3 days. Once colonies were formed and cultures were almost confluent, HNECs were passaged using the double-trypsinization technique with 0.025% trypsin-EDTA and seeded onto human type IV

placental collagen-coated 12mm or 30mm Millicell porous support inserts (Millipore) at ~1.5 x 105 cells/cm^2. Once the cultures became confluent, the basolateral medium was changed to ALI medium, and the cultures were differentiated under ALI conditions as described previously. Media was changed, with PBS washes of the apical surface, three times weekly for the first two weeks, followed by twice weekly. Using PneumaCult media according to manufacturer's protocols (Stem Cell Technologies), PneumaCult-Expand Plus medium replaced the CRC+Y medium, and PneumaCult-ALI medium substituted the ALI medium in these procedures. HNEC cultures are considered fully differentiated after 28 days in culture (ALI Day 26). HNECs were cultured at least twice at separate times for each experiment.

For culturing HNECs from cryopreserved, frozen cell stocks, thawed cells that were expanded and frozen in CRC medium previously were seeded in conditioned fibroblast medium (CFM) first followed by a slow titration to ALI medium supplemented with 1 μM DAPT, a γ-secretase and indirect Notch inhibitor, to ensure ciliation of HNECs. To prepare CFM, 12 – 15 x 10^6 irradiated 3T3-J2 fibroblasts were plated on 150 mm plastic dishes in 25 mL of CRC medium and incubated at 37°C for 72 hours. After 72 hours, medium was collected, pooled, and centrifuged at 350 x *g* to pellet cellular debris. Supernatant was transferred to new conical tubes, and stored at 4°C for up to 3 weeks or at -80°C for up to 6 months. Specifically, HNECs were thawed, centrifuged, resuspended, and counted in CFM supplemented with 5 μM Y-27632 (Sigma-Aldrich) (CFM+Y). HNECs were seeded onto human type IV placental collagen-coated 12mm Millicell porous support inserts (Millipore) at 2 x 10^5 cells per insert in CFM+Y medium. Media was changed every other day until cultures became confluent. Once cultures were confluent, apical media was removed, and basolateral media was switched every other day following the order of these media compositions: 100% CFM + 1 μM DAPT; 100% CFM + 1 μM DAPT; 75% CFM:25% ALI + 1 μM DAPT; 50% CFM:50% ALI + 1 μM DAPT; 25% CFM:75% ALI; 100% ALI. ALI medium was changed three times for one week with PBS washes, and then switched to washes and medium changes twice a week. Thawed cells that were expanded and frozen in PneumaCult medium previously were seeded on Purecol-coated 100 mm plastic dishes in PneumaCult-Expand Plus medium. Cultures were expanded until 70 – 90% confluent, and then passaged using 0.025% trypsin-EDTA.

HNECs were seeded onto human type IV placental collagen-coated 12mm Millicell porous support inserts (Millipore) at 2×10^5 cells per insert in PneumaCult-Expand Plus medium. Media was changed every other day until cultures became confluent. Once cultures were confluent, apical media was removed, and basolateral media was switched to PneumaCult-ALI medium. HNEC cultures were washed with PBS, and PneumaCult-ALI medium was changed according to manufacturer's protocols (Stem Cell Technologies).

4.2.5 Transmission Electron Micrographs of Cilia Cross-Sections

Nasal scrape biopsies were performed as previously described (Olin et al., 2011; Müller et al., 2013). The curettes (Arlington Scientific) with HNECs were immediately immersed in a cold fixative consisting of 2% glutaraldehyde, 2% paraformaldehyde, and 0.5% tannic acid. HNECs were processed and transmission electron micrographs (TEMs) of cilia cross-sections were prepared as previously described (Carson et al., 1994; Carson et al., 2000). Adequacy of the samples was based on the presence of an adequate number of axonemes that were technically sufficient for interpretation, as characterized by a 9 + 2 microtubular structure, intact membranes, and clearly defined ultrastructure components, including microtubule doublets and the central apparatus. Most samples had ≥ 10 adequate cross sections of cilia for quantification. TEMs were examined for absence or shortening of outer and inner dynein arms. Subjects were classified as having either normal ciliary structure, absent or shortened ODA alone, absent or shortened IDA alone, or defects of both the outer and inner dynein arms (Olin et al., 2011). For quantification of the dynein arms, the number of normal ODA and IDA were counted per axoneme per PCD subject. The cross-section of the axoneme was classified as distal or proximal based on the absence or presence of microvilli in the TEM, respectively. The mean of normal proximal or distal ODA and IDA was compared to averages of normal ODA and IDA scored on TEMs of 62 healthy individuals (de long and Rutland, 1995; Shapiro and Leigh, 2017) and across PCD subjects, and analyzed by one-way ANOVA with Tukey's multiple comparisons test.

4.2.6 Reverse Transcription Polymerase Chain Reaction (RT-PCR)

Reverse transcription polymerase chain reaction (RT-PCR) was performed to determine the effect of the VUS [c.1097-11C>G] identified in IVS10 in family UNC-1231 on *SPAG1* RNA transcripts. In brief, total RNA was isolated using a RNeasy kit (QIAGEN) according to manufacturer's protocols. Using SuperScript II Reverse Transcriptase (Thermo Fisher Scientific), first-strand cDNA was synthesized. PCR was performed using Terra PCR Direct Red Dye Premix (TaKaRa). PCR products were electrophoresed on a 1% agarose gel, and isolated using a QIAquick Gel Extraction kit (QIAGEN) according to manufacturer's protocols. Isolated PCR product from proband UNC-1231 II-1 was cloned into a pCR™ 2.1-TOPO® TA vector using a TOPO TA Cloning Kit (Thermo Fisher Scientific). Cloned vectors were transformed into One Shot Chemically TOP10 Competent *Escherichia coli* cells using heat shock, and were cultured at 37°C overnight. Colonies were selected and subjected to colony PCR, and ten positive colonies were liquid cultured overnight at 37°C. Plasmids were isolated using a QIAprep Spin Midiprep Kit (QIAGEN) and were Sanger sequenced. The acquired sequences were compared to SPAG1 mRNA reference sequence (NCBI, NM_003114.5) using BLAST (NCBI) and SnapGene software version 6.0.2 (GSL Biotech) analysis. The primers used are listed in Appendix 3.

4.2.7 Ciliary Beat Frequency (CBF)

Ciliary beat frequency was measured as described previously (Sears et al., 2015). In brief, three cultures of each experimental group were washed with PBS for 5 minutes at 37°C. 60 uL of PBS was added to the apical surface of each culture. Cultures were then individually visualized using a Nikon Eclipse TE2000 inverted microscope with phase optics and a 20× objective (NA = 0.45). Using a Tokai HIT controller (model INU-TIZ-F1) and a microscope stage-heater block, the temperature was maintained at 37°C. High-speed videos (60 frames/s) were recorded using a Basler acA1300-200um camera controlled by SAVA software (Ammons Engineering). CBF videos were analyzed using SAVA whole-field analysis. An average of 12 CBF videos were taken per culture.

4.2.8 Isolated Cell Immunofluorescence

Immunofluorescent staining of isolated human airway epithelial cells was performed as described previously (Blackburn et al., 2017). Briefly, isolated cells from HNEC cultures in 1X PBS were placed on 1% Alcian Blue (Sigma-Aldrich)-treated slides/cover slips and allowed to settle on the slides/cover slips for 20 minutes on ice. Cells were fixed for 10 minutes with 2% PFA or 100% methanol. After washing with 1X PBS twice, cells were permeabilized with 0.2% Triton X-100 in PBS for 15 minutes. Cells were then incubated in blocking solution (5% donkey serum, 3% BSA, and 0.1% Triton X-100 in 1X PBS) for 1 hour at RT. Samples were incubated in primary antibody diluted in blocking solution either for 1 hour at RT or overnight at 4°C. After washing with 1X PBS three times, cells were incubated in secondary antibody solutions for 2 hours at RT, and then washed with 1X PBS four times.

Nuclei were stained using DNA dyes SYTO13 or Hoechst33342 (Thermo Fisher Scientific). Isolated cells were mounted on slides with ProLong Gold or Diamond Antifade Reagent (Thermo Fisher Scientific). Slides were imaged using a Zeiss 800 upright confocal microscope with a 63X/1.4 oil objective. No detectable staining was observed for isotype-matched antibody or no primary antibody control slides. All images were processed using FIJI software and pseudocolored, and brightness/contrast was adjusted evenly across subjects per experiment. All antibodies used and their dilutions are listed in Appendix 2.

4.2.9 Immunoblots

HBECs and HNECs were lysed in RIPA buffer (Thermo Fisher Scientific) supplemented with protease inhibitor cocktail (Sigma-Aldrich). Using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific), the protein concentration of lysates was determined. Samples (5 – 20 μg) of either whole cell lysates or co-immunoprecipitation eluates were electrophoresed on Novex NuPAGE 4 – 12% Bis-Tris gels in 1X NuPAGE MOPS-SDS Running Buffer (Invitrogen) and then transferred to 0.45 μm nitrocellulose membranes using 1X NuPAGE Transfer Buffer (Invitrogen). For immunoblotting for DHCs, samples were electrophoresed on Novex NuPAGE 3 – 8% Tris-Acetate gels in 1X NuPAGE Tris-Acetate Running Buffer (Invitrogen) instead. After transfer, membranes were stained with REVERT

Total Protein Stain (LI-COR), imaged using a LI-COR Odyssey Scanner, destained, and then blocked in 5% non-fat milk in 1X TBS solution for 1 hour at RT. Membranes were then incubated in primary antibody solution in 5% non-fat milk in 1X TBST for 1 hour at RT or 4°C overnight, followed by four washes with 1X TBST for 5 minutes each. Membranes were then incubated in secondary antibody solutions in 5% non-fat milk in 1X TBST for 1 hour at RT, followed by four washes with 1X TBST for 5 minutes each. Fluorescent membranes were then scanned by LI-COR Odyssey Scanner, or chemiluminescent membranes were probed using Amersham ECL Plus reagents (GE Healthcare, Chicago, IL, USA) according to manufacturer's protocol. All antibodies used and their dilutions are listed in Appendix 2.

Quantitative analysis was performed on fluorescent immunoblots scanned by a LI-COR Odyssey Scanner using Image Studio Lite software version 5.2 (LI-COR). For immunoblots examining levels of SPAG1 in HNECs, background-subtracted signals were normalized to a total protein stain and then normalized to a positive HBEC control to control for variations across immunoblots. For immunoblots examining levels of dynein chains in HNECs, background-subtracted signals were normalized to a total protein stain, followed by normalization to levels of RSPH1 as a cilia control, before their normalization to a positive HBEC control to control for variations across immunoblots. In both types of analysis, normal control means were set to 1.0. Immunoblot images were processed and their brightness/contrast adjusted for publication with FIJI software version 2.1.0. Experiments involving HNECs were repeated, in whole beginning from cell culture, twice ($n = 2$), and depending on the immunoblot and availability of lysates, at least two technical replicates were averaged for each n.

4.2.10 DNAI1 Co-Immunoprecipitation

Immunoprecipitations for endogenous DNAI1 were performed using a Pierce Crosslink IP Kit (Thermo Fisher Scientific) according to manufacturer's protocols. In essence, 5 μg of anti-DNAI1 antibody or control IgG was coupled and crosslinked to protein A/G plus agarose resin in spin columns by incubating with Coupling Buffer and 2.5 mM disuccinimidyl suberate in DMSO for 45 minutes each at RT. HNEC cultures were lysed using IP Lysis/Wash Buffer supplemented with protease inhibitor cocktail

(Sigma-Aldrich), and protein concentration was determined using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Cell lysates were precleared by incubating with control agarose resin for 1 hour at 4°C. Similar amounts of protein from precleared lysates were added to columns containing resin crosslinked to anti-DNAI1 antibody or IgG control, and these columns were incubated overnight (12 – 18 hours) at 4°C. Columns were washed multiple times with IP Lysis/Wash Buffer and 1X Conditioning Buffer, followed by elution of eluates using Elution Buffer or 4X LDS Sample Buffer. Eluates were examined by immunoblotting as described above.

Quantitative analysis of protein interactions was performed by analyzing the level of dynein chain protein co-precipitated with DNAI1 compared to a control on immunoblots using Image Studio Lite software version 5.2 (LI-COR). Specifically, background-subtracted signals were normalized to the amount of DNAI1 protein that was eluted in each sample, and then compared with the control, which was normalized to 1.0.

4.2.11 Statistics

Data are expressed as the mean ± standard deviation (s.d.) in all experiments. Experiments involving HNECs were performed as at least two independent replicates, in whole beginning with cell culture at separate times, due to limits on the availability of nasal cells. Owing to the small sample size and the possibility of a non-normal distribution, non-parametric statistical tests were used for the experiments with $n \leq 6$. Figure legends specify the statistical method and P-values for each experiment. $P<0.05$ was considered significant. The following statistical methods were calculated using Prism software version 9.3.1 (GraphPad): Kruskal-Wallis test with uncorrected Dunn's multiple comparisons. Graphs were created with Prism software version 9.3.1 (GraphPad), and figures were created and processed using Photoshop 2020 software version 21.1.0 (Adobe).

4.3 Results

4.3.1 SPAG1 knockout using CRISPR-Cas9 is unsuccessful in HBECs

Previous studies have suggested that SPAG1 is involved in axonemal dynein arm assembly, but its exact role in this process has yet to be determined (Knowles et al., 2013; Horani et al., 2018). To better

understand the role of SPAG1 in this process, we attempted to explore the effect of mutations in *SPAG1* in HBECs by using CRISPR/Cas9 gene editing technology. Normal HBECs were lentiviral transduced by spinoculation with a vector expressing the Cas9 nuclease and a sgRNA targeting exon 4 of *SPAG1*. Transduced cells were selected using puromycin, expanded in CRC conditions, and cultured at an ALI to differentiate into a pseudostratified airway epithelium deficient in *SPAG1*.

To determine if the expression of SPAG1 was successfully abrogated in our CRISPR/Cas9-edited HBECs to produce a ciliary phenotype, we performed PCR to examine the deletion of *SPAG1* and immunoblots to measure the expression of SPAG1, as well as performed active area and CBF measurements on these cultures. PCR using primers for exon 4 of *SPAG1* showed no apparent reduction of *SPAG1* in the *SPAG1* CRISPR knockout HBECs compared to nontransduced control HBECs (Figure 4.1A), and sequencing revealed possible mutant transcripts originating from the site of the sgRNA (data not shown). Immunoblots for SPAG1 and DNAI1 as a ciliogenesis control showed a slight reduction in all isoforms of SPAG1, as well as DNAI1, in both the *SPAG1* CRISPR knockout HBECs and the scrambled control HBECs compared to nontransduced control HBECs (Figure 4.1B - 4.1C). Calreticulin (*CALR*) was used a loading control. Finally, CBF and active area measurements showed that while there was a significant decrease in the active area of the *SPAG1* CRISPR knockout HBEC cultures, there wasn't a significant decrease in the CBF of these cultures (Figure 4.1D – 4.1E). There was a similar decrease in the active area of the scrambled control HBECs, but this decrease was not considered significant. However, there was a significant decrease in the CBF of the scrambled control HBECs. Therefore, our attempt to completely knockout *SPAG1* in our HBEC cultures using CRISPR/Cas9 gene editing to create a *SPAG1*-deficient PCD cell culture model was not successful. This could be due to multiple reasons including the sgRNA not being a sufficient target sequence, not isolating a single homozygous *SPAG1* null cell colony to expand to produce complete *SPAG1* knockout HBEC cultures due to the inability of expanding primary airway cells from a single cell colony, and lentiviral transduction causing cell toxicity and affecting the HBECs' ability to differentiate, etc.
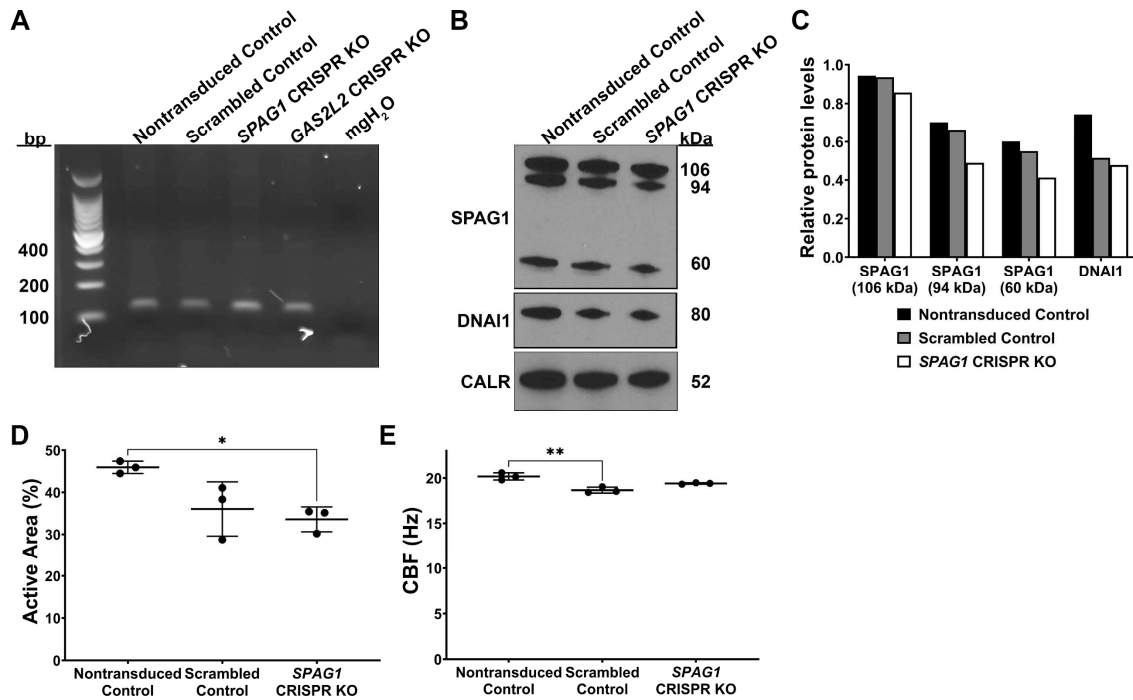
Figure 4.1 *SPAG1* knockout using CRISPR/Cas9 is unsuccessful in HBECs

(A) PCR of *SPAG1* exon 4 on gDNA from nontransduced control, scrambled control, and *SPAG1*

CRISPR knockout HBEC cultures. *GAS2L2* is another PCD gene that was being edited by CRISPR/Cas9

in HBEC cultures. All samples produced bands of similar size to the normal *SPAG1* exon 4 transcript in

the nontransduced control (178 bp). Molecular grade water (mgH$_2$O) was used as a negative control. **(B)**

Immunoblot for SPAG1, DNAI1 as a ciliogenesis control, and CALR as a loading control on

nontransduced control, scrambled control, and *SPAG1* CRISPR knockout HBEC lysates. **(C)**

Quantification of immunoblots probing for SPAG1 and DNAI1 protein levels. Background-subtracted

SPAG1 and DNAI1 signals were normalized to CALR signals. **(D)** Active area and **(E)** CBF

measurements of nontransduced control, scrambled control, and *SPAG1* CRISPR knockout HBEC

cultures. n = 3 cultures; *P<0.05, **P<0.01 compared to nontransduced control (Kruskal–Wallis with

uncorrected Dunn's multiple comparisons).

4.3.2 Pathogenic genetic variants in SPAG1 identified in three PCD samples

Due to the failed attempt to produce a complete knockout of *SPAG1* in HBEC cultures with CRISPR/Cas9 gene editing, we decided to acquire HNECs that were already deficient in *SPAG1* from PCD-affected individuals. We identified and studied three PCD-affected individuals with genetic variants in the *SPAG1* gene. Using whole-exome sequencing (WES), proband III-2 from family UNC-68 was discovered to have compound heterozygous mutations in *SPAG1*: a deletion resulting in a frameshift in exon 11, c.1119delC (p.Ala374Argfs*16), and a substitution resulting in a premature stop in exon 16, c.2014C>T (p.Gln672*) (Figure 4.2A). We identified this same nonsense mutation in exon 16 in proband III-1 from family UNC-372 as homozygous variants (Figure 4.2B). A previously identified large 11,971-bp deletion (c.61+201_POLR2K:c.140+1169_SPAG1del) encompassing exons 1 and 2 of *SPAG1* was discovered in proband II-1 from family UNC-1231 (Figure 4.3). This large deletion included the canonical translation start codon and was predicted to abrogate SPAG1 translation (Knowles et al., 2013). Additionally, a variant of uncertain significance (VUS) was identified in intron IVS10, c.1097-11C>G, in proband UNC-1231 II-1 (Figure 4.2C). These genetic variants follow the autosomal-recessive inheritance model in these three families and are expected to result in the loss of the full-length SPAG1 isoform (Figure 4.4).
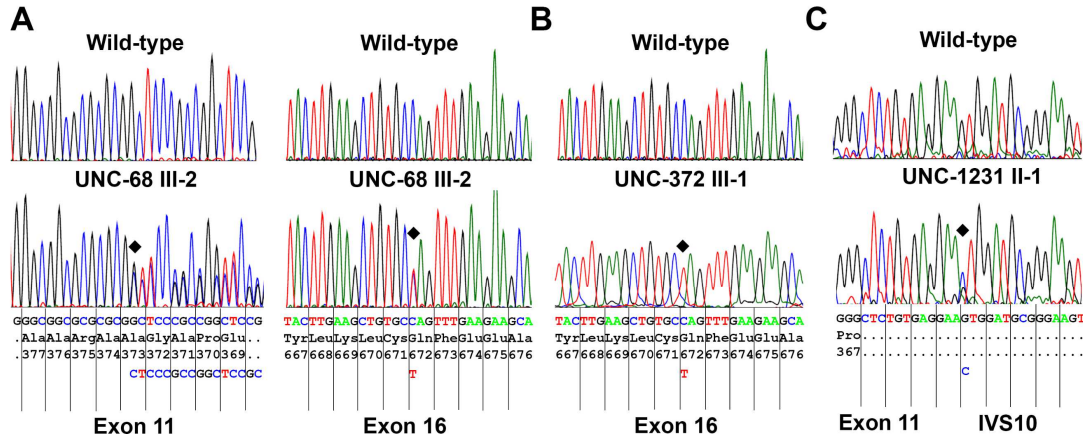
Figure 4.2 Sanger sequencing of pathogenic genetic variants in *SPAG1*-deficient PCD subjects

(A) Reverse sequencing chromatograms of compound heterozygous pathogenic variants [c.1119delC

(p.Ala374Argfs*16); c.2014C>T (p.Gln672*)] in exons 11 (left) and 16 (right) of *SPAG1* in proband

UNC-68 III-2 (bottom) and a wild-type individual (top). **(B)** Forward sequencing chromatograms of

homozygous pathogenic variant [c.2014C>T (p.Gln672*)] in exon 16 of *SPAG1* in proband UNC-372 III-

1 (bottom) and a wild-type individual (top). **(C)** Reverse sequencing chromatograms of compound

heterozygous variant c.1097-11C>G (p.Lys366Ilefs*119) in *SPAG1* in proband UNC-1231 II-1 (bottom)

and a wild-type individual (top). The intronic VUS in IVS10 of SPAG1 substitutes a C for a G ten

nucleotides upstream of the start of exon 11.
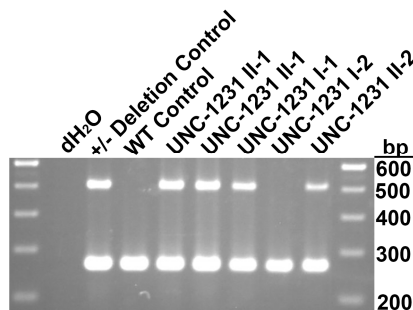
Figure 4.3 PCR confirmation of large *POLR2K* and *SPAG1* deletion in family UNC-1231

PCR of family UNC-1231 gDNA to screen for the 11,971 bp deletion

[c.61+201_POLR2K:c.140+1169_SPAG1del] using *POLR2K* intron 2 sense and *SPAG1* intron 2

antisense primers. 503 bp fragments are detected in individuals carrying deletion (proband UNC-1231 II-1, unaffected father UNC-1231 I-1, and unaffected sister UNC-1231 II-2) and are absent in non-carriers (WT control and unaffected mother UNC-1231 I-2). *DNAI1* primers for exon 3 amplifying a 266 bp fragment was used as an internal control.
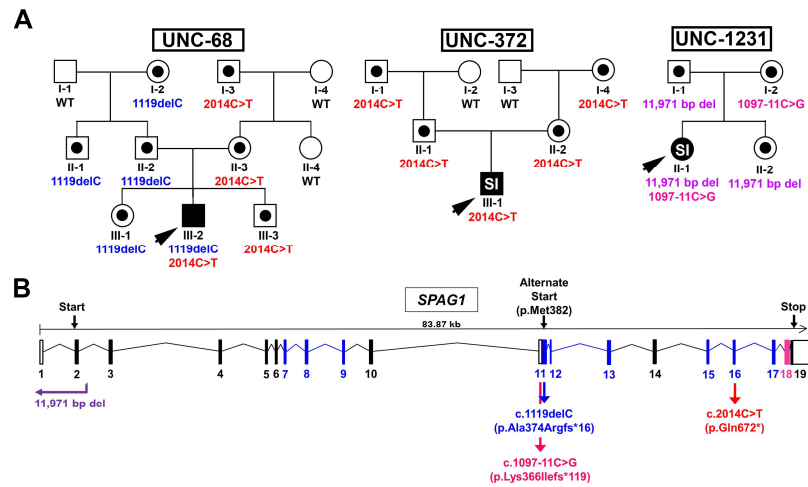
Figure 4.4 Pedigrees of and genetic variants identified in *SPAG1*-deficient PCD subjects

(A) Pedigrees of genetic variants identified in families UNC-68, UNC-372 and UNC-1231. Symbols with black dots indicate carrier status, and black symbols indicate affected individuals with PCD. Probands are designated by black arrowheads, and SI indicates individuals with situs inversus. WT, wild type. **(B)** Diagram of the *SPAG1* gene (NCBI, NG_033834.2). Blue and pink boxes represent the exons that translate into tetratricopeptide domains and the RPAP3-like C-terminal domain, respectively. The positions of all genetic variants found in families UNC-68 (blue and red arrows), UNC-372 (red arrow) and UNC-1231 (purple and magenta arrows) are shown.

To determine whether the identified novel intronic VUS (c.1097- 11C>G) in proband UNC-1231 II-1 is pathogenic, reverse transcription PCR (RT-PCR) was performed on RNA from cultured HNECs for SPAG1 exons 9 - 13. The proband sample produced a major transcript similar in size to the normal control (Figure 4.5A). Gel isolation, cloning, and sequencing of this transcript showed that ~50% of the

plasmids contained a 10-nucleotide insertion (TTCCTCACAG) at the start of exon 11 that matches the ten nucleotides in intron IVS10 immediately downstream of the VUS. This suggests that this substitution shifts the exon 11 splice donor site ten nucleotides upstream and leads to a premature stop codon at the start of exon 12 (p.Lys366Ilefs*119) (Figure 4.5B). Thus, these data support the conclusion that this VUS is a pathogenetic variant that can lead to PCD.
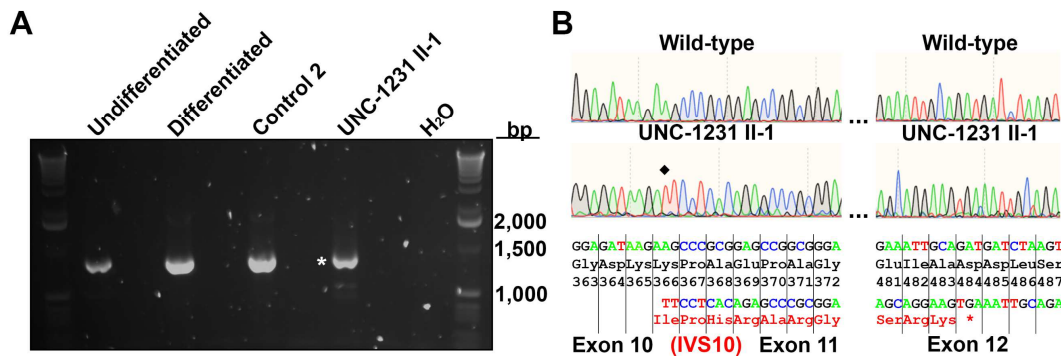
Figure 4.5 RT-PCR confirmation of pathogenicity of VUS identified in *SPAG1*

(A) RT-PCR of *SPAG1* exons 8/9 to exons 13/14 on total RNA from undifferentiated HBEC, differentiated HBEC, normal HNEC (Control 2), and proband UNC-1231 II-1 HNEC cultures. All samples produced bands of similar size to the normal *SPAG1* exons 8/9 to exons 13/14 transcript. The PCR product (white asterisk) from proband UNC-1231 II-1 was gel isolated, cloned into a TOPO TA vector, and sequenced. **(B)** Sequencing chromatograms of the VUS [c.1097-11C>G (p.Lys366Ilefs*119)] in the cloned cDNA of *SPAG1* from proband UNC-1231 II-1 (bottom) and a wild-type individual (top). The intronic substitution (black diamond) causes the splice donor site for exon 11 to shift 10 nucleotides upstream, which then causes a 10-nucleotide insertion at the beginning of exon 11. This insertion causes a frameshift that results in a premature stop codon at amino acid #484 in the beginning of exon 12. Base sequences, amino acid sequences, and codon numbers are shown. Genetic variants are marked by black diamonds. Mutated amino acids are in red.

All three subjects shared common clinical characteristics of PCD, including neonatal respiratory distress, rhinosinusitis, otitis media, and low nasal nitric oxide (Table 4.2; Leigh et al., 2013). Subjects UNC-68 III-2 and UNC-372 III-1 had bronchiectasis, whereas there was no computed tomography imaging available for UNC-1231 II-1. However, the phenotypic severity could be considered milder in UNC-68 III-2, having not developed bronchiectasis until 20 years of age compared to 5 years of age in UNC-372 III-1. UNC-372 III-1 also had worse lung function, with a forced expiratory volume (FEV_1) of 80% predicted at 15 years of age, whereas UNC-68 III-2 had an FEV_1 of 98% predicted at 27 years of age. UNC-1231 II-1 had a FEV_1 of 89% predicted at 6 years of age.

4.3.3 Atypical ultrastructural defect present in SPAG1-deficient PCD samples

While all three of these PCD-affected individuals had genetic variants in the *SPAG1* gene, they had differing ultrastructural changes (Figure 4.6; Table 4.2). Transmission electron micrographs (TEMs) of nasal cilia revealed that proband III-1 from family UNC-372 had absent or defective ODAs and IDAs, with no normal dynein arms present, which is typical for defects in axonemal dynein assembly (Omran et al., 2008; Knowles et al., 2013). Interestingly, both proband III-2 from family UNC-68 and II-1 from family UNC-1231 had few, if any, IDA but many normal ODA present, which is an atypical ciliary ultrastructural defect for PCD subjects with *SPAG1* mutations (Figure 4.6A). Quantification of these TEMs showed an average of 5.3 ± 3.7 and 5.4 ± 3.3 (mean ± s.d.) normal ODA visible in proximal and distal axonemes, respectively, in subject UNC-68 III-2, and 4.9 ± 1.1 and 5.3 ± 2.1 normal ODA visible in proximal and distal axonemes in subject UNC-1231 II-1, compared to 8.5 ± 0.3 in 62 normal controls (de Longh and Rutland, 1995; Shapiro and Leigh, 2017). There was no significant difference between the numbers of normal ODA in proximal and distal ciliary axonemes, indicating that there was not a specific type 1 or type 2 ODA defect in these individuals (Fliegauf et al., 2005) (Figure 4.6B).

Table 4.2 Demographic, Clinical Phenotype, and S*PAG1* Genetic Variants in Three PCD-Affected Individuals

Family #	Sex	Age in Years[1]	Situs Status	Ethnicity	nNO nl/min[2]	Neo RDS	Bxsis	Sinusitis	Otitis Media	Allele #1				Allele #2			
										Exon/ Intron #	Base Change	Amino Acid Change	Segregation[3]	Exon/ Intron #	Base Change	Amino Acid Change	Segregation[3]
Complete Defects in ODA & IDA																	
UNC-372 III-1	M	18	SI	white	11.3	yes	yes	yes	yes	16	c.2014C>T	p.Gln672*	paternal	16	c.2014C>T	p.Gln672*	maternal
Incomplete Defects in ODA & IDA																	
UNC-68 III-2	M	30	SS	white	12.8	yes	yes	yes	yes	11	c.1119delC	p.Ala374Argfs*16	paternal	16	c.2014C>T	p.Gln672*	maternal
UNC-1231 II-1	F	6	SI	white	27.0	yes	NA	yes	yes	multi[4]	large deletion[4]	untranslated	paternal	IVS10[5]	c.1097-11C>G[5]	p.Lys366 Ilefs*119[5]	maternal

Abbreviations are as follows: M, male; F, female; SS, situs solitus; SI, situs inversus; ODA + IDA, outer + inner dynein arms; Neo RDS, neonatal respiratory distress in full-term neonates; Bxsis, bronchiectasis; NA, not available; nNO, nasal nitic oxide.

[1]Current age

[2]Normal nNO levels are 304.6 ± 118.8 nl/min (mean ± SD), calculated from 78 healthy subjects (Leigh et al., 2013).

[3]Mutant allele shown to segregate with either the paternal or maternal side of the family (pedigrees in Figure 4.4).

[4]Large 11,971 bp deletion (c.61+201_POLR2K:c.140+1169_SPAG1del) comprised of multiple introns, exons, and untranslated regions from *POLR2K* and *SPAG1*. This deletion includes translation start codon and is predicted to abrogate SPAG1 translation.

[5]Variant of uncertain significance (c.1097-11C>G) was identified in intron IVS10 of *SPAG1*. This variant leads to abnormal insertion of 10 nucleotides from IVS10 at the exon 10 and exon 11 junction. This 10-nucleotide insertion leads to a shift in the reading frame, which causes a premature stop codon in the beginning of exon 12 at amino acid #484 of full-length SPAG1.
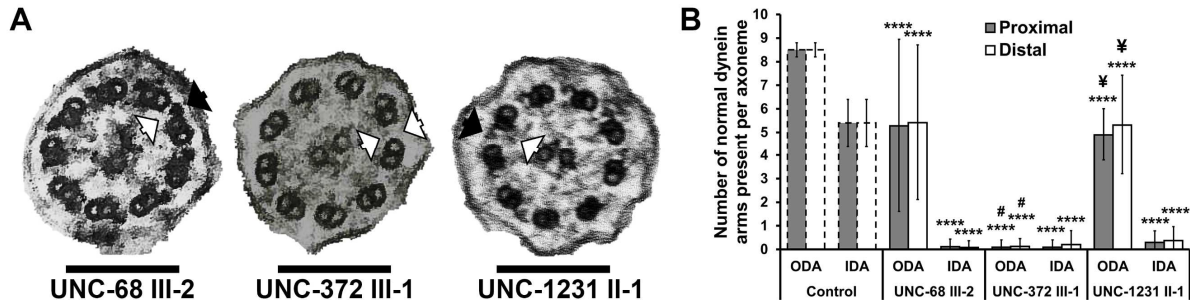
Figure 4.6 UNC-372 III-1 has an ODA + IDA defect, while UNC-68 III-2 and UNC-1231 II-1 have normal ODAs

(A) TEMs of nasal cilia cross-sections from probands UNC-68 III-2 (left), UNC-372 III-1 (middle) and UNC-1231 II-1 (right). UNC-372 III-1 has absent IDAs and ODAs (white arrows), whereas UNC-68 III-2 and UNC-1231 II-1 have normal ODAs (black arrows). Scale bars: 150 nm. **(B)** Quantification of normal ODAs and IDAs present per proximal (gray bars) and distal (white bars) ciliary axoneme examined by TEMs. Control values were based on averages of normal ODAs and IDAs scored on TEMs of 62 healthy individuals (dashed bars) (de Longh and Rutland, 1995). UNC-68 III-2, UNC-372 III-1 and UNC-1231 II-1 n = 18, 31, and 10 proximal axonemes and 24, 24, and 19 distal axonemes, respectively; ****$P<0.0001$ compared to Control; #$P<0.0001$ compared to UNC-68 III-2; ¥$P<0.0001$ compared to UNC-371 III-1 (ANOVA with Tukey's multiple comparisons).

4.3.4 CBF is reduced in PCD samples

To further explore the effect of these *SPAG1* variants on ciliary function, we cultured primary HNECs from these three probands and two healthy controls at the ALI to differentiate into mucociliary airway epithelia (Suprynowicz et al., 2012; Müller et al., 2013; Fulcher and Randell, 2013). The ciliary beat frequencies (CBFs) of cultured HNECs from all PCD subjects (<4 Hz) were significantly reduced compared to their corresponding controls (22.3 ± 0.5 Hz, 17.4 ± 0.5 Hz; mean ± s.d.) (Figure 4.7). Although *SPAG1*-deficient PCD subjects have previously been reported to have essentially immotile cilia (Knowles et al., 2013), ciliated cells from UNC-68 III-2 and UNC-1231 II-1 showed some ciliary motility, with a CBF of 3.7 ± 0.7 Hz and 1.7 ± 1.0 Hz, respectively. In contrast, the cilia of UNC-372 III-

1 HNECs were immotile (0.3 ± 0.2 Hz) (Figure 4.7). Taken together, subjects UNC-68 III-2 and UNC-1231 II-1 have an atypical ciliary phenotype, with normal ODA present and some ciliary motility, compared to other PCD subjects with *SPAG1* variants.
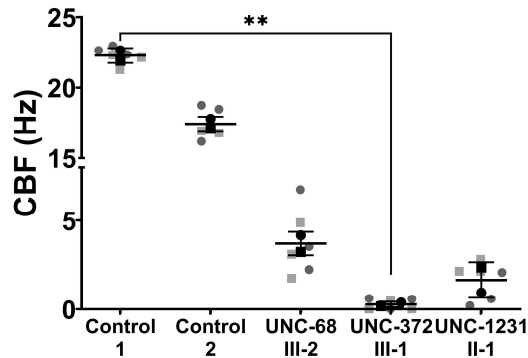
Figure 4.7 Ciliary beat frequency is reduced in *SPAG1*-deficient PCD subjects

CBF measurements of HNEC cultures from control and PCD-affected subjects. Grey circles and light grey squares represent the mean CBF of individual cultures from two distinct experiments. Black circles and squares represent the means of three cultures from each independent experiment. Six cultures were measured across $n = 2$ independent experiments; **$P<0.01$ (Kruskal–Wallis with uncorrected Dunn's multiple comparisons).

4.3.5 Expression of 60-kDa SPAG1 isoform leads to atypical ultrastructural defect

To determine the amount, if any, of SPAG1 proteins produced in these subjects, we performed immunoblots on whole-cell lysates using an antibody that targeted the C-terminus of SPAG1. Notably, the full-length and 94-kDa isoforms of SPAG1 were absent in all three subjects. However, the 60-kDa isoform of SPAG1 was present in subjects UNC-68 III-2 and UNC-1231 II-1, and in control protein lysates from non-PCD subjects (Figure 4.8A). Quantification of these immunoblots revealed that similar levels of the 60-kDa SPAG1 were present in UNC-1231 II-1 compared to a control, whereas the protein levels of UNC-68 III-2 were reduced to 42% expected (Figure 4.8B). Immunoblots using an anti-SPAG1

antibody that targeted the amino-terminus only detected the full-length and 94-kDa isoforms in a control protein lysate and no isoforms in the UNC-68 III-2 and UNC-372 III-1 protein lysates, which could suggest that the predicted 48-kDa SPAG1 isoform is not expressed in human nasal or bronchial epithelial cells (Figure 4.8C).
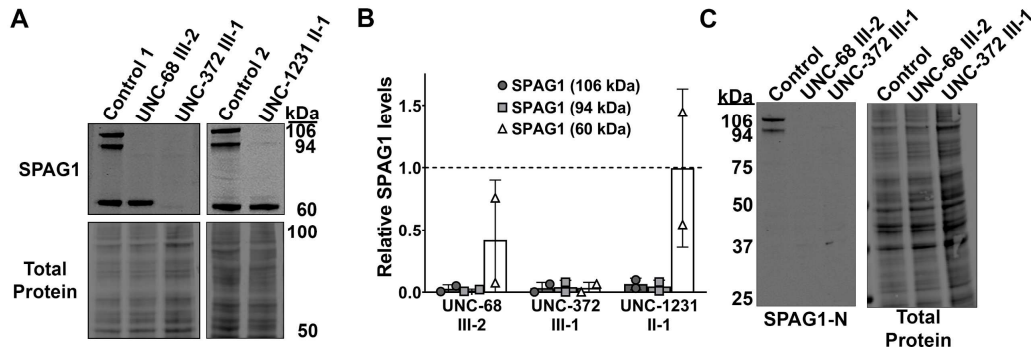
Figure 4.8 Expression of 60-kDa SPAG1 results in atypical ciliary ultrastructural defect

(A) Representative immunoblots for SPAG1, and corresponding total protein stain, on HNEC lysates from normal and PCD-affected individuals. **(B)** Quantification of immunoblots probing for SPAG1 protein levels. Raw signals were normalized to a total protein stain, then a separate differentiated HBEC control to control for variation across immunoblots, and then the control means were normalized to 1.0 (dashed line). $n = 2$ independent experiments. **(C)** Representative immunoblots of SPAG1 using an antibody targeting the amino-terminus and a total protein stain on whole-cell lysates of HNECs isolated from a control, proband UNC-68 III-2, and proband UNC-372 III-1. Only the full-length and 94-kDa isoforms of SPAG1 are detected in the control sample. $n = 2$ independent experiments.

Immunofluorescent staining for SPAG1 and ERICH3, a proximal cilia marker (Blackburn et al., 2017), on isolated multiciliated cells confirmed the absence of any SPAG1 protein in UNC-372 III-1, whereas some SPAG1 staining was still detected in ciliated cells from UNC-68 III-2 and UNC-1231 II-1 (Figure 4.9). Of note, in control nasal multiciliated cells, SPAG1was present only in the cytoplasm as large puncta (Figure 4.9, top panel), recently referred to as dynein axonemal particles (DynAPs) that are

hypothesized to be the location of cytoplasmic dynein arm assembly (Huizar et al., 2018; Lee et al.,

2020). As this 60-kDa SPAG1 isoform still contains two TPR domains and the RPAP3 C-terminal

domain (Figure 3.4) and can still interact with DNAAF2 (Figure 3.12), it is possible that it can retain

some dynein arm assembly function and partially compensate for the loss of full-length SPAG1, which

could explain the presence of some ODAs and ciliary motility in probands UNC-68 III-2 and UNC-1231

II-1.
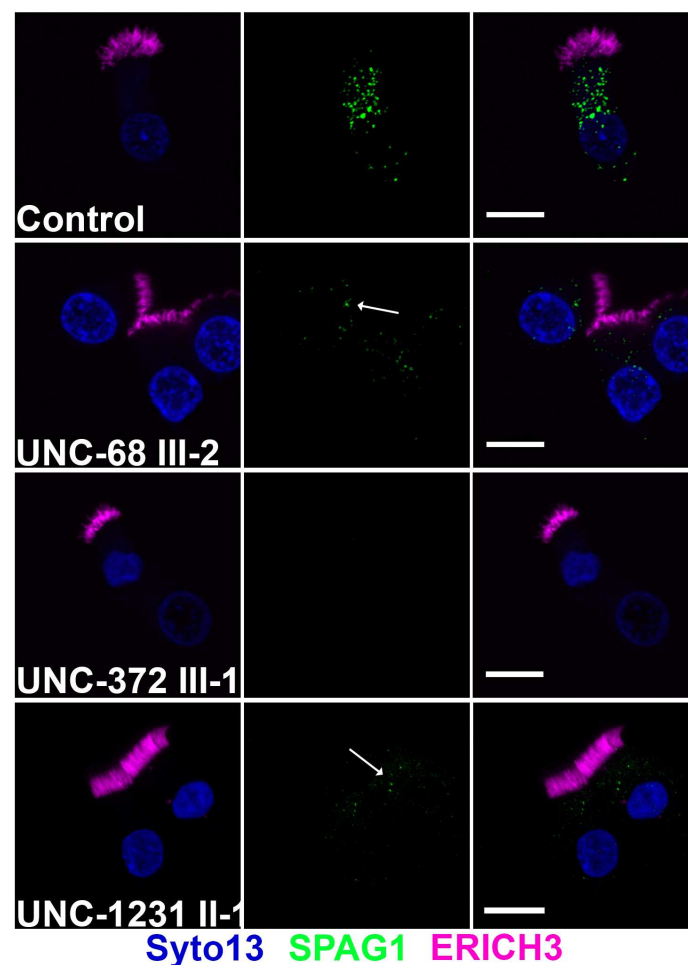
Figure 4.9 Presence and localization of SPAG1 in *SPAG1*-mutant PCD subjects

Representative isolated cell immunofluorescence images of multiciliated HNECs from normal and

PCD-affected individuals stained for SPAG1 (green) and ERICH3 (magenta), a cilia marker. SYTO13

(blue) was used to stain the nuclei. Arrows indicate SPAG1 puncta that is still present in UNC-68 III-2

and UNC-1231 II-1 nasal cells. Images are representative of a total of 20 fields across *n* = 2 independent experiments. Scale bars: 10 μm.

4.3.6 Dynein heavy chain protein levels are reduced in SPAG1-deficient PCD samples

To determine which step(s) in dynein arm assembly requires functional SPAG1, we investigated the axonemal dynein defects found in the three PCD subjects with *SPAG1* genetic variants. Immunofluorescent staining for DNAI1, DNAH5, DNALI1 and a cilia marker were performed on HNECs from control and PCD subjects. DNAI1 and DNAH5, ODA dynein chains, were present along the full length of the ciliary axoneme in subjects UNC-68 III-2 and UNC-1231 II-1, albeit at a reduced level compared to control cilia. There was no signal for DNAI1 or DNAH5 in ciliary axonemes of UNC-372 III-1 (Figure 4.10A - 4.10B). Additionally, there was no signal for DNALI1, an IDA dynein chain, in ciliary axonemes in all three subjects (Figure 4.10C). These results corroborate the TEM findings from nasal cilia from these *SPAG1*-deficient PCD individuals.
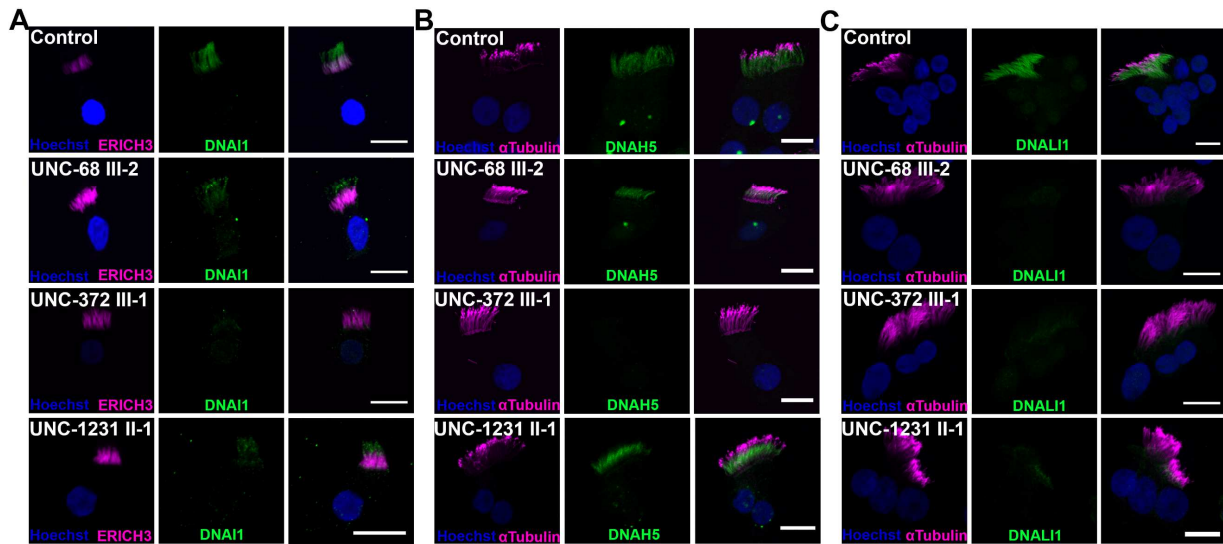
Figure 4.10 UNC-372 III-1 has defective ODAs + IDAs, whereas UNC-68 III-2 and UNC-1231 II-1 have normal ODAs

Single-cell immunofluorescence images of cultured multiciliated nasal cells from normal and PCD-affected individuals stained for ODAs, using DNAI1 **(A)** and DNAH5 **(B)**, and IDAs, using DNALI1 **(C)** (green). Cilia were labeled with either ERICH3 or acetylated α-tubulin (magenta), and nuclei (blue) were stained using Hoechst 33342. Images are representative of a total of 20 fields of view across n = 2 independent experiments. Scale bars: 10 μm.

Owing to the instability of dynein chains when not incorporated into dynein arm complexes, we hypothesized that distinct dynein chains in *SPAG1*-deficient PCD individuals would be reduced at a greater level than others based on the stage of dynein assembly SPAG1 is involved in (Omran et al., 2008; Mitchison et al., 2012). We performed immunoblots for various dynein chains present in ODAs or IDAs on whole-cell lysates of cultured HNECs. ODA dynein chains examined were the intermediate chains DNAI1 and DNAI2, the heavy chains DNAH5, DNAH9, and DNAH11, and the light chain DNAL1. IDA dynein chain examined was the light intermediate chain DNALI1. RSPH1 was used as a cilia control. All dynein chains examined were reduced in all three subjects (Figure 4.11A). Intriguingly, in subject UNC-372 III-1, in which no SPAG1 isoforms are expressed, there was a substantial reduction in the levels of ODA DHCs, DNAH5, DNAH9 and DNAH11, with DNAH5 being statistically significant (Figure 4.11B). This suggests, in combination with the data showing that SPAG1 interacts with various DHCs and chaperone complexes, that SPAG1 facilitates the formation and folding of the DHCs.
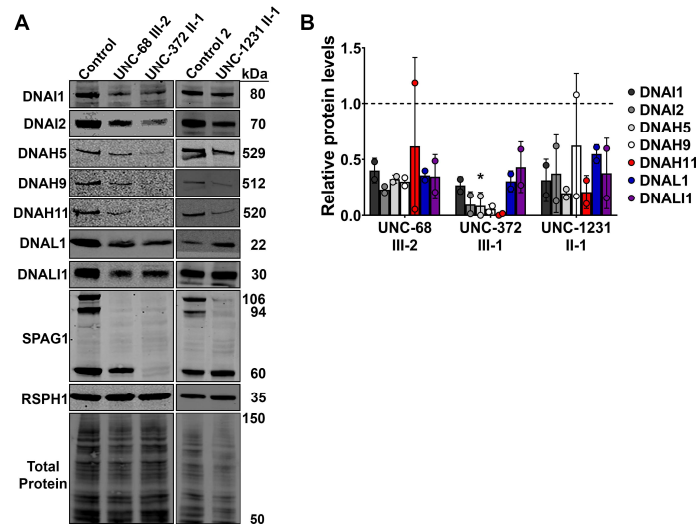
Figure 4.11 DHC protein levels are reduced in *SPAG1*-deficient PCD samples

(A) Immunoblots for various ODA dyneins (DNAI1, DNAI2, DNAH5, DNAH9, DNAH11 and DNAL1), an IDA dynein chain (DNALI1), SPAG1, a cilia control (RSPH1) and total protein stain on the HNEC lysates of normal and PCD-affected individuals. **(B)** Quantification of dynein chain protein expression measured by immunoblots on the HNEC lysates of normal and PCD-affected individuals. Background-subtracted signals were normalized to a total protein stain and RSPH1, then a separate differentiated HBEC lysate to control for variations across immunoblots, and then the nasal control means were set to 1.0 (dashed line). $n = 2$ independent experiments; *$P<0.05$ compared to control (Kruskal–Wallis test with uncorrected Dunn's multiple comparisons).

4.3.7 Dynein heavy chain interaction with dynein intermediate chain complex is inhibited in SPAG1-deficient PCD samples

To determine whether SPAG1 is required for the interaction of the ODA DHCs to the intermediate chains, co-immunoprecipitations for DNAI1 were performed on whole-cell lysates from differentiated cultures of control and PCD nasal cells. Immunoblots for DNAI1, DNAI2, DNAH5 and RSPH1 (a cilia control) were performed and quantified to determine the level of interaction between DNAI2 and DNAH5 with DNAI1. The interaction between DNAI1 with DNAI2 or DNAH5 was reduced

in subjects UNC-68 III-2 and UNC-1231 II-1 according to quantification of the immunoblots. Surprisingly, there was no further reduction in the levels of interaction between DNAI1 with DNAI2 in UNC-372 III-1, suggesting that SPAG1 facilitates the interaction, formation and/or stabilization of the ODA DIC complex but is not absolutely required. However, there was no interaction detected between DNAI1 with DNAH5 in subject UNC-372 III-1 (Figure 4.12). This result suggests that SPAG1 is necessary for the interaction between DNAI1-DNAI2 and DNAH5. Thus, SPAG1 could be involved in associating and binding the DHCs to the previously formed DIC-DLC complexes. It is also possible that this reduction in the DNAH5 and DNAI1-DNAI2 interaction was due to reduced DNAH5 formation, as shown by reduced total protein levels (Figure 4.11), suggesting SPAG1 could be involved in the formation and folding of DHCs. In conclusion, SPAG1 plays a crucial role in cytoplasmic axonemal dynein arm assembly by facilitating the formation of DIC-DHC complexes, either by forming and folding the DHCs, and/or binding the DHCs to previously formed DIC/LC complexes. Further studies are required to determine whether SPAG1 is necessary for one function, the other, or both.
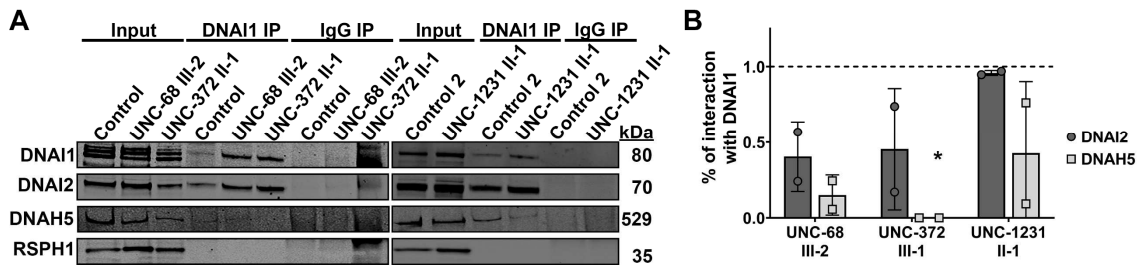
Figure 4.12 SPAG1 facilitates the addition of DHCs to the intermediate chain complex

(A) Immunoblots for DNAI1, DNAI2, DNAH5 and a cilia control (RSPH1) analyzing co-immunoprecipitation for endogenous DNAI1 samples from HNECs of normal and PCD-affected individuals. **(B)** Quantification of the amount of DNAI2 and DNAH5 protein interacting with DNAI1 in PCD-affected HNECs compared to normal HNECs. Background-subtracted DNAI2 or DNAH5 signals were divided by DNAI1 signals per each eluate sample, and then the control means were set to 1.0 (dashed line). Data shown are mean ± s.d.; $n = 2$ independent experiments; *$P<0.05$ compared to control

(Kruskal–Wallis test with uncorrected Dunn's multiple comparisons). IP, immunoprecipitation. Input lanes represent 5% of the total lysate.

4.4 Discussion

4.4.1 60-kDa SPAG1 can compensate for the lack of full-length SPAG1 and assemble ODA

Three individuals with clinical features compatible with PCD were identified as having genetic variants in *SPAG1*. Subject UNC-68 III-2 had heterozygous compound mutations in *SPAG1*: a novel deletion in exon 11 [c.1119delC (p.Ala374Argfs*16)] and a previously identified nonsense mutation in exon 16 [c.2014C>T (p.Gln672*)]. Subject UNC-372 III-1 had homozygous nonsense mutations in exon 16 [c.2014C>T (p.Gln672*)]. Subject UNC-1231 II-1 had a previously identified 11,971 bp deletion that encompassed exons 1 and 2, and thus the initiation codon, of *SPAG1*. Additionally, a VUS was identified [c.1097-11C>G (p. Lys366Ilefs*119)] in IVS10 of *SPAG1* that was shown to be pathogenic by RT-PCR and sanger sequencing due to shifting the splice donor site of the start of exon 11 upstream 10 nucleotides, which caused a frame shift and an eventual premature stop codon in the beginning of exon 12. Interestingly, individuals UNC-68 III-2 and UNC-1231 II-1 showed presence of normal ODA on TEM of nasal cilia cross-sections and some ciliary motility, while only subject UNC-372 III-1 had a typical mutated *SPAG1* PCD presentation with completely defective ODAs + IDAs and completely immotile cilia (Knowles et al., 2013). The TEM findings were confirmed by immunofluorescence for DNAI1, DNAH5, and DNALI1, and it was further proven that the normal ODA present in subjects UNC-68 III-2 and UNC-1231 II-1 was not due to a specific ODA type 1 (DNAH5 and DNAH11) or type 2 (DNAH5 and DNAH9) defect. Here, we show that this atypical PCD ciliary phenotype is due to expression of the previously uncharacterized 60-kDa isoform of SPAG1 in subjects UNC-68 III-2 and UNC-1231 II-1. Thus, the 60-kDa isoform of SPAG1 can partially compensate for the lack of full-length SPAG1 and assemble some normal ODAs.

To further test the ability for 60-kDa SPAG1 to participate in dynein arm assembly, co-immunoprecipitations for FLAG-tagged full-length or 60-kDa SPAG1 showed that both were able to co-

precipitate HA-tagged DNAAF2 in HEK293T cells, although the interaction with the 60-kDa SPAG1 was not as strong as with full-length SPAG1 (from Chapter 3). This suggests that 60-kDa isoform of SPAG1 can play a role in dynein arm assembly when full-length SPAG1 is absent, possibly through a pathway involving DNAAF2. The PIH1D2-binding motif (amino acids #732 - #784) is still present in the 60-kDa isoform, and it is possible that PIH1D2 can also play a role in assembling ODAs in the absence of full-length SPAG1 (Maurizy et al., 2018). However, one of the binding sites for HSP70 and/or HSP90, the first TPR domain, is absent in this isoform. Interestingly, these subjects only assemble a large number of normal ODAs, but not IDAs. This difference could be due to limitations of TEMs to detect IDA; however, DNALI1 immunofluorescence studies suggest normal IDA are not present in ciliary axonemes of multiciliated HNECs from these individuals. This demonstrates that assembly of the IDAs that involve SPAG1, such as IDA *g* and *f* based on interaction data, requires the first TPR domain, possibly due to facilitating the recruitment and interaction of multiple chaperones, or exchange of clients between chaperones, specifically HSP70 and HSP90 (zur Lage et al., 2018; Chagot et al., 2019; Dermouche et al., 2021). Interestingly, the homolog of SPAG1 (CG18472/Spag1) has recently been shown to have a conserved function in dynein arm assembly in *Drosophila*, but only contains the first TPR domain, suggesting this first TPR domain is required for efficient dynein arm assembly of ODAs and IDAs (zur Lage et al., 2018). Further studies are needed to corroborate that the 60-kDa SPAG1 can compensate for the lack of full-length SPAG1, specifically determining which interactors and/or complexes are involved in dynein assembly, and why only ODAs are assembled.

4.4.2 SPAG1 is required for the folding and/or interaction of DHCs to the DIC complex

To determine which step in the axonemal dynein assembly process requires SPAG1, we performed immunoblots to measure levels of dynein chains and immunoprecipitations for DNAI1 on normal and mutant HNECs. In HNEC lysates in which all isoforms of SPAG1 were absent (UNC-372 III-1), there was a drastic reduction in the levels of ODA intermediate chain DNAI2 and heavy chains DNAH5, DNAH9, and DNAH11. Surprisingly, co-immunoprecipitations for DNAI1 in HNECs showed that the interaction between DNAI1 and DNAI2 occurs in the absence of SPAG1 (UNC-372 III-1).

However, DNAH5 was not determined to be present in this dynein complex in the absence of SPAG1, which suggests that SPAG1 is required for the binding of DHCs to the DIC complex (DNAI1-DNAI2). However, the significant reduction in this interaction could be due to the decrease in total DHC protein levels, which suggests that SPAG1 plays a role in DHC formation and folding. If SPAG1 is instead only required for the binding of DHCs to the DIC complex, the reduction in total protein DHC levels could be attributed to the degradation of unstable DHCs not present in dynein complexes. Further new evidence for this was obtained by co-immunoprecipitations of endogenous SPAG1 indicating interactions between SPAG1 and ODA DHCs, including DNAH5 and DNAH9, IDA g heavy chain, DNAH6, and ODA DIC, DNAI1. Additionally, SPAG1 has also been shown to interact with IDA f heavy chain, DNAH10, in *Drosophila* testes and ODA intermediate chain, DNAI2, in Xenopus (zur Lage et al., 2018; Lee et al., 2020).

It was previously suggested that DNAAF1 (LRRC50) and DNAAF2 (KTU) were involved in the folding of DHCs, with DNAAF3 facilitating the stabilization or interaction of DHCs to the DIC complex based on immunoblots, immunoprecipitations, and trypsin sensitivity assays on dynein chains in *Chlamydomonas reinhardtii* (Fowkes and Mitchell, 1998; Omran et al., 2008; Mitchison et al., 2012). Additionally, quantitative proteomic analysis of $Wdr92^{-/-}$ mutant whole spermatocyte extracts in *Drosophila* showed reduction of six dynein heavy chains and several IDA intermediate chains compared to control extracts (zur Lage et al., 2018). In *Wdr92* mutant strains of *Chlamydomonas reinhardtii*, a significant decrease in the presence of ODA and IDA heavy chains was observed, while the intermediate chain complex was still present (IC1-IC2) (Liu et al., 2019; Patel-King et al., 2019). As SPAG1 has potential interactions with DNAAF1, DNAAF2, DNAAF3, and WDR92, these previous studies provide additional evidence that SPAG1 does play a role in the formation, folding, stabilization, and/or interaction of DHCs to the DIC complex.

Thus, in normal airway epithelium, full-length SPAG1 likely forms R2TP-like complexes with RUVBL1-RUVBL2, PIH1D2 and/or DNAAF2. SPAG1 can then interact further with chaperones, HSP70, HSP90, and possibly CCT through WDR92, and other DNAAFs, such as DNAAF1, DNAAF3,

and WDR92, as client adaptors. This co-chaperone complex then facilitates the formation of axonemal

dynein arms by folding, stabilizing, and/or binding DHCs to the DIC complex (Figure 5.1A). In PCD

cells lacking the full-length and 94-kDa isoforms of SPAG1, the expressed 60-kDa isoform partially

compensates for the lack of full-length SPAG1 and results in the assembly of some ODAs (Figure 5.1B).

In PCD cells that lack all isoforms of SPAG1, the R2TP-like complexes cannot form. This leads to

inhibition of the folding and interaction of DHCs causing impairment of dynein arm assembly, and the

unstable dynein complex components are eventually degraded (Figure 5.1C). Further studies are required

to determine if this mechanism is the same when applied to IDA, as the majority of these studies focused

on ODA assembly. Correlating the distinct dynein arm subtypes that are assembled with the

corresponding R2TP-like complexes, DNAAFs, and chaperones remains to be explored.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

Cilia are complex structures, composed of over 500 proteins that are precisely assembled into different substructures, each playing unique roles in ciliary function (Blackburn et al., 2017; Wallmeier et al., 2020). Defects in many of these substructures (e.g., ODAs, IDAs, radial spokes and the central pair) have been shown to affect ciliary function and cause PCD (Shapiro et al., 2016; Davis et al., 2019; Wallmeier et al., 2020). PCD is a genetically heterogenous disorder, with mutations currently identified in >50 different genes that result in PCD, but many more are expected to be revealed (Zariwala et al., 2007; Shapiro et al., 2016; Davis et al., 2019). Ciliary ultrastructural defects, the presence of laterality defects and the severity and progression of lung disease can vary in PCD based on the specific dysfunctional gene (Leigh et al., 2019; Davis et al., 2019). Here, we show that distinct genetic variants within the same gene can lead to variations in axonemal ultrastructural defects, with impacts on ciliary function and possibly disease severity, which can further complicate the diagnosis of PCD.

Through a multitude of studies using healthy and PCD human airway epithelia, we have determined the role that SPAG1 plays in axonemal dynein arm assembly in the cytoplasm of multiciliated cells. Through our expression studies, using immunoblots, targeted proteomics, and 5' RACE, we identified a previously uncharacterized 60-kDa SPAG1 isoform that is transcribed from an alternative start site and contains the second and third TPR domains and the RPAP3 C-terminal domain of full-length SPAG1. Using ddPCR, immunoblots, and immunofluorescence, we determined that the full-length and 94-kDa isoforms' expression corresponds with ciliogenesis, while the 60-kDa isoform is constitutively expressed and possibly has a cilia-independent function. Through co-immunoprecipitation studies, we demonstrated that SPAG1 interacts with other DNAAFs, dynein chains, and canonical components of the R2TP/Prefoldin-like complex, where the evidence for three of these interactions, DNAAF2, PIH1D2, and DNAAF1, were strengthened by reverse co-immunoprecipitations and co-expression studies using ddPCR

and immunoblots. Our studies combined with previous literature suggests that SPAG1 scaffolds R2TP-like complexes with other DNAAFs. We also discovered that the 60-kDa isoform of SPAG1 potentially interacts with different proteins from the full-length isoform, and these interactions suggest a role for this isoform in regulation of the actin-microtubule cytoskeleton. Co-immunoprecipitation studies of epitope-tagged wild-type or mutant SPAG1 and DNAAF2 in HEK293T cells demonstrated that DNAAF2 requires its CS domain to interact with SPAG1, while SPAG1 potentially interacts with DNAAF2 through multiple C-terminal domains.

We studied three individuals with genetic variants in *SPAG1* that have clinical features compatible with PCD. One of these individuals had a VUS in intron IVS10, c.1097-11C>G, that was confirmed to be pathogenic through RT-PCR studies. Interestingly, two of these three PCD-affected individuals had a ciliary phenotype that was atypical for those with *SPAG1* mutations with the presence of normal ODAs and some ciliary motility, as demonstrated by TEM and CBF measurements. This atypical ciliary phenotype was attributed to the expression of the 60-kDa isoform of SPAG1 in these two subjects as shown by immunoblots and immunofluorescence of HNECs. The 60-kDa isoform of SPAG1 was also shown to still be able to interact with DNAAF2, the most abundant DNAAF interactor of SPAG1 identified in our studies. Taken together, this suggests that the 60-kDa isoform could partially compensate for the loss of full-length SPAG1 and play a role in axonemal dynein arm assembly to assemble some ODAs. Thus, this study highlights the importance of performing functional studies of specific mutations associated with atypical PCD phenotypes, as there can be heterogeneity in phenotypes of PCD patients with distinct genetic variants within the same gene, in addition to the different genes associated with PCD. Additionally, immunoblots measuring the abundance of dynein chains in healthy control and *SPAG1*-deficient PCD HNECs revealed a significant reduction in dynein heavy chain levels in the complete absence of SPAG1. Co-immunoprecipitation studies for endogenous DNAI1 in healthy control and *SPAG1*-deficient PCD HNECs demonstrated an inhibition of the interaction between DNAH5 and the DIC complex, but not for the components of the DIC complex itself, DNAI1-DNAI2. These studies

together suggest that SPAG1 is required for the folding of the DHCs and/or the association of the DHCs to the DIC/LC complex in dynein arm assembly.

In summary, endogenous SPAG1 plays an essential role in axonemal dynein arm assembly in humans by forming R2TP-like complexes with other DNAAF and chaperones that likely participate in the folding and binding of the dynein heavy chains to the dynein intermediate chains (Figure 5.1A). In PCD-affected individuals that lack full-length SPAG1 but express the 60-kDa SPAG1 isoform, the 60-kDa isoform compensates for full-length SPAG1 and assembles some ODA, which leads to some ciliary motility (Figure 5.1B). In PCD-affected individuals that completely lack SPAG1, the R2TP-like complexes do not form, and the DHCs are not folded and are not associated with the DIC/LC complex, which leads to inhibition of functional dynein arms being fully assembled. Partially assembled dynein arms are not transported into the ciliary axonemes, and these complexes and dynein chains are eventually degraded. This results in immotile cilia, which leads to PCD (Figure 5.1C).

However, there are a multitude of questions about SPAG1 and the process of dynein arm assembly that remain to be answered. For instance, what is the function of the 60-kDa isoform of SPAG1? Previous literature and studies in mice and human cell lines have suggested SPAG1 plays a role in cell proliferation and cell motility, possibly as an indirect regulator of the cell cycle through the ERK pathway or the PI3K/AKT pathway through the microRNA miR-638 (Liu et al., 2006; Neesse et al., 2007; Hu et al., 2017; Lin et al., 2021). Interestingly, the 59-kDa SPAG1 isoform in mouse was found to be involved in spindle morphogenesis and cell cycle control in mouse oocytes, where actin expression was abruptly decreased and the formation of the cortical granule-free domains, actin caps, and contractile ring were disrupted by knockdown of SPAG1 expression (Huang et al., 2016). SPAG1 was also associated with spindle morphogenesis in Large White boar Sertoli cells where it co-localized with microtubules in metaphase and telophase (Hu et al., 2017). Our studies show that this 60-kDa SPAG1 isoform is constitutively expressed, and anecdotal evidence of PCD HNEC cultures that completely lacked all SPAG1 isoforms were slower to grow and become confluent compared to PCD HNEC cultures
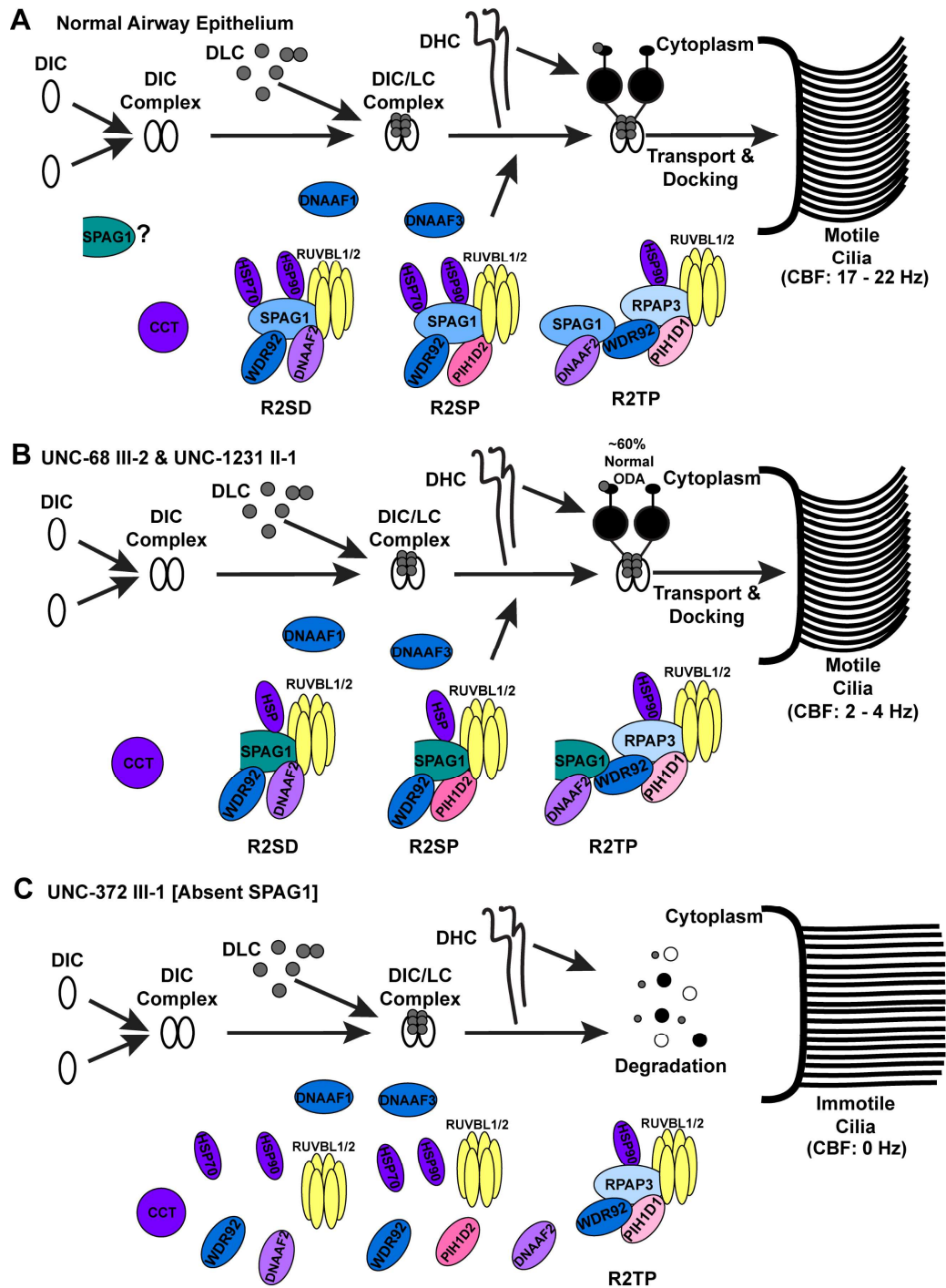
Figure 5.1 Function of full-length SPAG1 and compensation of 60-kDa SPAG1 in axonemal dynein arm assembly

(A) In normal human airway multiciliated cells, full-length SPAG1 (blue) interacts with RUVBL1/2, DNAAF2 or PIH1D2, and WDR92, to form potential R2TP-like complexes that facilitate the folding and

139

binding of the DHCs to the DIC-DLC (DIC/LC) complex. It is also possible that SPAG1 interacts with the canonical R2TP complex through WDR92 to facilitate dynein arm assembly. These R2TP-like complexes can further interact with different chaperones (purple), such as HSP70, HSP90 and CCT, and other DNAAFs (dark blue), including DNAAF1 and DNAAF3, to recruit dynein chains to chaperones for formation into dynein arms. The dynein arms are then transported into the ciliary axoneme by intraflagellar transport, leading to motile cilia (CBF, 17 - 22 Hz). The 60-kDa SPAG1 isoform (teal) could play a role in dynein arm assembly to assemble a subset of ODAs, or it has a cilia-independent function. **(B)** In cells that are lacking full-length SPAG1 but still express the 60-kDa isoform of SPAG1 (UNC-68 III-2 and UNC-1231 II-1), some assembly of ODAs occurs even though SPAG1 is lacking one of its HSP binding sites. About 60% of normal ODAs leads to some ciliary motility (CBF, 2-4 Hz). **(C)** If SPAG1 is completely absent in airway epithelium (UNC-372 III-1), these R2TP-like complexes are unlikely to form, and thus, the folding and binding of DHCs to the DIC/LC complex is impaired. The cytoplasmic assembly process of dynein arms is inhibited, and the partially assembled dynein complex is degraded, instead of being transported into the ciliary axoneme. With this, cilia are completely immotile (CBF, 0 Hz).

that only expressed the 60-kDa SPAG1 suggests that this isoform might be involved in cell proliferation or cell motility in HAECs as well. Additionally, co-immunoprecipitations of the 60-kDa isoform of SPAG1 in undifferentiated HBECs showed potential interactions with other proteins involved in actin-binding and the cell cortex, suggesting that the 60-kDa isoform of SPAG1 might play a role in the actin-microtubule cytoskeleton to regulate spindle morphogenesis and cell division in somatic cells as well. Further studies are required to confirm this hypothesis.

Additionally, in PCD subjects that only express the 60-kDa SPAG1 isoform, why is it that only normal ODAs are assembled, instead of both ODAs and IDAs? The absence of IDAs detected in the ciliary axonemes of these PCD-affected individuals could be due to the limitations of the transmission electron microscopy method. However, since the 60-kDa SPAG1 is missing the first TPR domain that is

found in full-length SPAG1, this domain is possibly required for the association of HSPs or other chaperones with IDA dynein chain clients (Chagot et al., 2019; Dermouche et al., 2021). Interestingly, the *Drosophila* ortholog of SPAG1 only contains one TPR domain that has similar identity to the first TPR domain in full-length SPAG1 in humans, and this domain in *Drosophila* is sufficient for assembling both IDAs and ODAs (zur Lage et al., 2018). Here, our studies focused on the assembly of ODAs and less so the assembly of IDAs. Further studies need to be performed to determine the mechanism behind the 60-kDa isoform assembling ODAs and what are the differences in the mechanisms of assembly of ODAs and the assembly of IDAs.

We have found evidence that demonstrates SPAG1 is required for the folding of the DHCs and/or the association of the DHCs to the DIC/LC complex in dynein arm assembly. However, further studies are necessary to determine if SPAG1 is required for one function over the other or is involved in both processes. Since SPAG1 has been shown to form a R2TP-like co-chaperone complex and interacts with both HSP70 and HSP90, it is very likely that SPAG1 is involved in the folding and formation of the DHCs (Maurizy et al., 2018; Chagot et al., 2019). Additionally, two of the most abundant of SPAG1 interactors are DNAAF1 and DNAAF2, which are two other DNAAFs that have been implicated in the folding and stabilization of the DHCs, which further suggests that SPAG1 plays a role in DHC folding as well (Mitchison et al., 2012). It is also possible that the inhibition of the interaction between the DHCs and the DIC/LC complex could be due to the reduced amount of DHCs caused by impaired folding due to the absence of SPAG1. However, in our co-immunoprecipitation proteomic studies, a DIC, DNAI1, also precipitated with SPAG1, which may suggest that SPAG1 is also involved in associating the DHCs with the DIC/LC complex.

Axonemal dynein assembly is a very complex process that involves multiple proteins, including DNAAFs and chaperones, that interact together to fold and assemble dynein chain clients into a fully assembled axonemal dynein arm. The interactions and mechanisms behind this process are slowly being uncovered. While SPAG1 has been shown to interact with canonical R2TP complex components to form R2TP-like complexes, the interaction sites and mechanisms behind how this association occurs is not

fully understood. Due to SPAG1 being a paralog of RPAP3, it is possible that SPAG1 interacts with its complex components in an analogous manner and uses similar protein domains. For example, the CS-domain in DNAAF2 is required for DNAAF2 to interact with SPAG1, comparable to how the CS-domain in PIH1D1 is required for its interaction with RPAP3. PIH1D1 interacts with RPAP3 through a motif between the last TPR domain in RPAP3 and its C-terminal domain, which is equivalent to how PIH1D2 interacts with SPAG1 through a motif between the last TPR domain in SPAG1 and its RPAP3-like C-terminal domain (Maurizy et al., 2018; Lynham and Houry, 2018). Further interaction site mapping studies are necessary to determine how these complex components associate together.

Additionally, it remains largely unknown how the DNAAFs recruit and associate the dynein chain clients with the R2TP-like co-chaperone complexes and the chaperones themselves. Do the other DNAAFs that are not considered components of R2TP-like co-chaperone complexes act as adaptor proteins for the recruitment and association of dynein chain clients to the co-chaperone complexes? With the large number of the different dynein chains, specifically dynein heavy chains, that make up the distinct ODA and IDA types, it also brings to question which DNAAFs, chaperones, and co-chaperone complexes are used to assemble which ODA and IDA subtypes. How do chaperone and chaperone complexes create specificity for which dynein chain client they are required to fold and assemble? The sequence of the PHI1 domain in PIH1D1 has been implicated in substrate specificity of the canonical R2TP-like co-chaperone complex (Hořejší et al., 2014; von Morgen et al., 2015). SPAG1 has been proposed to form two different R2TP-like complexes with two different PIH1 domain containing proteins, PIH1D2 and DNAAF2. The other two PIH1 domain containing proteins found in humans are PIH1D1, a canonical R2TP complex component, and PIH1D3, another known PCD causative gene that has been suggested to form a R2TP-like complex as well (Maurizy et al., 2108). All of these proteins' orthologs have been implicated in axonemal dynein arm assembly in *Drosophila*, *Danio rerio,* and/or *Chlamydomonas*, where they demonstrated defects in distinct dynein arm subtypes dependent on the PIH1 domain-containing protein that was mutant (zur Lage et al., 2018; Yamaguchi et al., 2018; Yamamoto et al., 2020). Taken together, these data support the hypothesis that the specificity of the

R2TP-like co-chaperone complexes for folding and formation of specific dynein arm subtypes relies on the specific PIH1 domain containing protein present in the complex. Further studies are necessary to confirm this theory, as well as to determine if the sequences of the PIH1 domain and/or phosphorylation status of the client protein also helps to determine the substrate specificity. Additionally, it is also possible that different isoforms of the other components of the R2TP-like complexes also adds to regulating substrate specificity. However, this is dependent on if these other isoforms can still interact with the other complex components to form a co-chaperone complex. For example, the 94-kDa isoform of SPAG1 has been specifically understudied, especially in terms of its function. Its expression pattern is the same as the full-length SPAG1, indicating it is also likely be involved in ciliogenesis. However, in performing reverse co-immunoprecipitations for DNAAF1, only the full-length SPAG1 co-precipitated with DNAAF1, suggesting that the 94-kDa isoform, unlike full-length SPAG1, does not interact with DNAAF1. If DNAAF1 is used as an adaptor protein or used to recruit specific dynein chain clients to the co-chaperone complex, the inability for one R2TP-like complex with the 94-kDa SPAG1 isoform to interact with DNAAF1, while the other R2TP-like complex with the full-length SPAG1 isoform can, could change the clients that are able to associate with these respective co-chaperone complexes. All in all, axonemal dynein arm assembly that occurs in the cytoplasm of multiciliated cells is still a complex and largely unexplored process, where many additional studies are necessary to start to comprehend the mechanisms that allow for the exquisite, intricate assembly of these multifaceted structures that play a fundamental part in a healthy life and well-being.

APPENDIX 1: COMPOSITION OF COMMONLY USED MEDIA

Additive	Final Concentration in Media		
	BEGM[1]	ALI[1]	CRC
Base Media	50:50 LHC Basal: DMEM-H[2]	100% LHC Basal	3:1 DMEM-H: F12 Nutrient Mix
Bovine serum albumin	0.5 mg/mL	0.5 mg/mL	N/A
Bovine pituitary extract	10 μg/mL	10 μg/mL	N/A
Insulin	0.87 μM	0.87 μM	5 μg/mL
Transferrin	0.125 μM	0.125 μM	N/A
Hydrocortisone	0.21 μM	0.21 μM	0.025 μg/mL
Triiodothyronine	0.01 μM	0.01 μM	N/A
Epinephrine	2.7 μM	2.7 μM	N/A
Epidermal growth factor	25 ng/mL	0.50 ng/mL	0.125 ng/mL
Retinoic acid	5×10^{-8} M	5×10^{-8} M	N/A
Phosphorylethanolamine	0.5 μM	0.5 μM	N/A
Ethanolamine	0.5 μM	0.5 μM	N/A
Zinc sulfate	3.0 μM	3.0 μM	N/A
Penicillin G sulfate	100 U/mL	100 U/mL	100 U/mL
Streptomycin sulfate	100 μg/mL	100 μg/mL	100 μg/mL
Gentamicin	50 μg/mL	N/A	10 μg/mL
Amphotericin B	0.25 μg/mL	N/A	0.25 μg/mL
Ferrous sulfate	1.5×10^{-6} M	1.5×10^{-6} M	N/A
Magnesium chloride	6×10^{-4} M	6×10^{-4} M	N/A
Calcium chloride	1.0 mM	0.11 mM	N/A
Selenium	30 nM	30 nM	N/A
Manganese	1 nM	1 nM	N/A
Silicone	500 nM	500 nM	N/A
Molybdenum	1 nM	1 nM	N/A
Vanadium	5 nM	5 nM	N/A
Nickel sulfate	1 nM	1 nM	N/A
Tin	0.5 nM	0.5 nM	N/A
Fluconazole	N/A	N/A	0.25 μg/mL
Cholera Toxin	N/A	N/A	0.1 nM
Fetal bovine serum	N/A	N/A	10%

Abbreviations are as follows: BEGM, Bronchial epithelial growth medium; ALI, air/liquid interface; CRC, Conditionally reprogrammed cell; DMEM-H, Dulbecco's modified Eagle medium – High glucose; N/A, Not applicable.

[1]BEGM and ALI medium composition is listed as previously described (Fulcher and Randell, 2013).
[2]DMEM-H is 1X DMEM with 4.5 g/L glucose, L-glutamine, and sodium pyruvate.

APPENDIX 2: ANTIBODIES

Primary Antibodies

Target	Host Species	Clonality	Source	Catalog Number	Immunoblot Dilution	Immunofluorescence Dilution
SPAG1 (N-terminus)	Rabbit	Polyclonal	Atlas Antibodies	HPA053682	1:1,000	1:500
SPAG1 (C-terminus)	Mouse	Monoclonal (10G1/2)	Novus Biologicals	NBP2-50301	1:1,000	1:100
DNAAF1	Mouse	Monoclonal (4E4)	Lifespan Biosciences	LS-C173838	1:2,000	N/A
DNAAF2	Rabbit	Polyclonal	Atlas Antibodies	HPA002894	1:1,000	N/A
PIH1D2	Mouse	Monoclonal (OTI1A2)	Thermo Fisher	MA5-25402	1:250	N/A
HEATR2	Rabbit	Polyclonal	Atlas Antibodies	HPA020243	1:1,000	N/A
DNAI1[1]	Mouse	Monoclonal	N/A	N/A	1:3,000	1:1,000
DNAI2	Rabbit	Polyclonal	Atlas Antibodies	HPA050565	1:1,000	N/A
DNAH5	Rabbit	Polyclonal	Atlas Antibodies	HPA037470	1:250	1:100
DNAH9	Rabbit	Polyclonal	Atlas Antibodies	HPA052641	1:250	N/A
DNAH11	Rabbit	Polyclonal	Atlas Antibodies	HPA045880	1:250	N/A
DNAL1	Rabbit	Polyclonal	Atlas Antibodies	HPA053129	1:500	N/A
DNALI1	Mouse	Monoclonal	Abmart	5598	1:500	N/A
DNALI1	Rabbit	Polyclonal	Atlas Antibodies	HPA028305	1:1,000	1:250
FOXJ1	Mouse	Monoclonal [2A5]	Thermo Fisher	14-9965-82	1:250	N/A
RSPH1	Rabbit	Polyclonal	Novus Biologicals	NBP1-89839	1:1,000	N/A
ERICH3	Rabbit	Polyclonal	Atlas Antibodies	HPA028655	N/A	1:100
ERICH3	Goat	Polyclonal	Santa Cruz	SC-245591	N/A	1:100
γTubulin	Mouse	Monoclonal [GTU-88]	Sigma-Aldrich	T6557	N/A	1:100
Acetylated αTubulin	Mouse	Monoclonal [6-11B-1]	Sigma-Aldrich	T7451	N/A	1:500
FLAG-tag	Rabbit	Polyclonal	Sigma-Aldrich	F7425	1:1,000	N/A
HA-tag	Rabbit	Polyclonal	Sigma-Aldrich	H6908	1:1,000	N/A

Secondary Antibodies

Target	Host Species	Clonality	Source	Catalog Number	Immunoblot Dilution	Immunofluorescence Dilution
αRabbit HRP	Donkey	Polyclonal	Jackson ImmunoResearch	711-036-152	1:60,000	N/A
αMouse HRP	Goat	Polyclonal	Jackson ImmunoResearch	115-035-164	1:40,000	N/A
αRabbit IRDye 800CW	Donkey	Polyclonal	LI-COR	925-32213	1:5,000 – 1:15,000	N/A
αRabbit IRDye 680RD	Donkey	Polyclonal	LI-COR	925-68073	1:5,000 – 1:15,000	N/A
αMouse IRDye 800CW	Goat	Polyclonal	LI-COR	925-32210	1:5,000 – 1:15,000	N/A
αMouse IRDye 680RD	Goat	Polyclonal	LI-COR	925-68070	1:5,000 – 1:15,000	N/A
αRabbit Alexa Fluor 555	Donkey	Polyclonal	Thermo Fisher	A-31572	N/A	1:500
αMouse Cy3	Donkey	Polyclonal	Jackson ImmunoResearch	715-165-151	N/A	1:500
αMouse Alexa Fluor 488	Donkey	Polyclonal	Thermo Fisher	A-21207	N/A	1:500
αRabbit Alexa Fluor 488	Donkey	Polyclonal	Thermo Fisher	A-21206	N/A	1:500
αRabbit Rhodamine Red X	Donkey	Polyclonal	Jackson ImmunoResearch	711296152	N/A	1:500
αRabbit Alexa Fluor 647	Donkey	Polyclonal	Thermo Fisher	A-31573	N/A	1:500
αGoat Alexa Fluor 647	Donkey	Polyclonal	Thermo Fisher	A-21447		1:500
SYTO13 (DNA Dye)	N/A	N/A	Thermo Fisher	S7575	N/A	1:1,000
Hoechst33342 (DNA Dye)	N/A	N/A	Thermo Fisher	H1399	N/A	1:250

[1]Anti-DNAI1 antibody was previously produced by Ostrowski et al (Ostrowski et al., 2010).

APPENDIX 3: PRIMERS

Primer Name	Sequence (5' → 3')	Amplicon Length (bp)
Large Deletion Confirmation in UNC-1231 II-1		
POLR2K_SPAG1_del_F	TGTAAAACGACGGCCAGT-CGGATGGACTTGTGAGAAGC	503
POLR2K_SPAG1_del_R	CAGGAAACAGCTATGACC-CATCTGGCTGTGGACCAAC	
SPAG1_Exon2_F	TGTAAAACGACGGCCAGT-ATTGCTTGAGGCCAGAAGTT	536
SPAG1_Exon2_R	CAGGAAACAGCTATGACC- TCTCATGCTGGCAGTAGACC	
DNAI1_Exon3_F	TGAATGAAGGGGCTTTCTGT	266
DNAI1_Exon3_R	TCTGGGCAAGCTCCTAATCA	
Pathogenicity of Intronic Variant in UNC-1231 II-1		
hSPAG1_Ex8/9_F	TGTAAAACGACGGCCAGTAACGTAAAGGCTCTTCTGCG	875
hSPAG1_Ex13/14_R	CAGGAAACAGCTATGACCCTTGATAGCCTGTTAACACTGTC	
hSPAG1_intEx10_F	CTGAAGCTGCATCTGAGAC	550
hSPAG1_intEx12_R	TCTTGAATGCAGCCACTG	
M13_F	GTAAAACGACGGCCAG	875
M13_R	CAGGAAACAGCTATGAC	
5' RACE to Determine 60-kDa SPAG1 Transcript		
GSP1_hSPAG1_Ex18_F	GATTACGCCAAGCTTAGGAGCCTGGAAGACCTGCAGGGGA	428
GSP1_hSPAG1_Ex19_R	GATTACGCCAAGCTTTGGTGTGTCCGAAAGGTCCTCAAACAGC	
GSP2_hSPAG1_Ex14_F	GATTACGCCAAGCTTTTCCTGCTGTGCCTGCTTCTGTGCC	200
GSP2_hSPAG1_Ex15_R	GATTACGCCAAGCTTGCATTCGCTGTATTTACTGAGGGCGTCT	
Universal Primer Mix	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	Variable
hSPAG1_Ex14_F	CTTGGCATCCGGCAAAAGAG	163
hSPAG1_Ex15_R	TCGCTGTATTTACTGAGGGCG	
pRACE_M13_F	GTAAAACGACGGCCAGT	Variable
pRACE_M13_R	CAGGAAACAGCTATGAC	
ddPCR mRNA Expression		
ddPCR hSPAG1 Gene Expression Probe (Bio-Rad Assay ID: dHsaCPE5033992)[1]	GATTATGAAGAAGCAGTGATGTATTATACCAGGAGCATATCAGCGCTTCCCACTGTAGTTGCCT ATAACAATCGAGCTCAAGCAGAAATCAAATTACAGAACTGGAATAGTGCTTTTCAGGA	76
hSPAG1_Ex14_F[2]	AAGCTGTCACCTATTCCTGC	127
hSPAG1_Ex15_R[2]	TTTCATCTGTGATGCCCTGC	
ddPCR hDNAAF1 Gene Expression Probe (Bio-Rad Assay ID: dHsaCPE5044988)[1]	TTTGATCGCATTGAGAACCTGGAAGAGTACACAGGGCTGCGCTGTCTCTGGCTGCAGAGCAAT GGAATACAGAAAATCGAAAACCTGGAGGCCCAAACTGAGTTGCGTTGCCTCTTCTTGCAA	117
ddPCR hDNAAF2 Gene Expression Probe (Bio-Rad Assay ID: dHsaCPE5030054)[1]	ATCATATAGCAAAGAATTCTGAAAACTGAATGCACAATTGGTCACATGATCTTTAATGACCTGC ACACTACCATCCTGCATATTGGTTTCTTTTATGCTGTCGAATCCAGGTATATTGTGAAC	79
ddPCR PIH1D2 Gene Expression Probe (Bio-Rad Assay ID: dHsaCPE5038344)[1]	GAGAGAGACAGAATTAATACCAGGTAATTCAACTTTCAACTCAATTTTCAGAGGTTTCTCACTG TGATCATGCACAATTTTTAGTTCATAGGCTGGCATCTTCATCTCCACCTGGATTTCAGT	101
ddPCR FOXJ1 Gene Expression Probe (Bio-Rad Assay ID: dHsaCPE5054850)[1]	GCCTGGTTCGTCCTTCTCCCGAGGCACTTTGATGAAGCACTTGTTCAGAGACAGGTTGTGGCGG ATTGAATTCTGCCAGGTGGGATCTGCGTGGCGGAAGTAGCAGAAGTTGTCCGTGATCCA	96
ddPCR DNAI1 Gene Expression Probe (Bio-Rad Assay ID: dHsaCPE5035646)[1]	CCCTGACCAGCTGGAGTTGACCGATGCGGAGTTAAAGGAGGAGTTCACTCGGATTTTGACAGC CAACAACCCACACGCACCCCAGAACATTGTCAGGTACAGCTTCAAAGAAGGCACATATAA	71
ddPCR CALR Gene Expression Probe (Bio-Rad Assay ID: dHsaCPE5055017)[1]	AGTATTCTCCCGATCCCAGTATCTATGCCTATGATAACTTTGGCGTGCTGGGCCTGGACCTCTG GCAGGTCAAGTCTGGCACCATCTTTGACAACTTCCTCATCACCAACGATGAGGCATAC	131

Creation of pcDNA3.1(+)-hSPAG1-FLAG and pcDNA3.1(+)-hDNAAF2-HA Plasmids

pcDNA3.1+hSPAG1FLAG_60kDa_F	ATGGGCAACATCCAGAAGAAGC	7,028
pcDNA3.1+hSPAG1FLAG_60kDa_R	GGTGGCGGATCCGAGCTC	
pcDNA3.1+hSPAG1FLAG_-TPR1_F	GATTTGGCCAAGAAAACC	7,868
pcDNA3.1+hSPAG1FLAG_-TPR1_R	AAGAAAATCCTTTTCTTTCTCAG	
pcDNA3.1+hSPAG1FLAG_-TPR2_F	CAGCTAGCAAATGACAGTGTTAACAGGC	7,841
pcDNA3.1+hSPAG1FLAG_-TPR2_R	GTTCTCCGCGCCCTGCCC	
pcDNA3.1+hSPAG1FLAG_-TPR3_F	ATTGAGGCAAAGATGGAAC	7,865
pcDNA3.1+hSPAG1FLAG_-TPR3_R	TGTTTTTTCATCTGTGATGC	
pcDNA3.1+hSPAG1FLAG_-PIHIS_F	GGAGGCAAAAGCAGCAGG	8,012
pcDNA3.1+hSPAG1FLAG_-PIHIS_R	CAGTTCCATCTTTGCCTCAATAATAC	
pcDNA3.1+hSPAG1FLAG_-RPAP3_F	GGCCAAAAGGAGCTAATTG	7,901
pcDNA3.1+hSPAG1FLAG_-RPAP3_R	GGCTATCGGAAGTTTCTC	
pcDNA3.1+hDNAAF2HA_-PIH1_F	AGGAATGCCAAGACCCTGAAGGCC	7,439
pcDNA3.1+hDNAAF2HA_-PIH1_R	CGACGAGGAGGCCGCCGC	
pcDNA3.1+hDNAAF2HA_-CS_F	CTGCCGGCCGCGCGCCGG	7,670
pcDNA3.1+hDNAAF2HA_-CS_R	AGGCTCGGTGGGGGCGGGC	
pcDNA3.1+_CMVP_F	TGGGAGGTCTATATAAGCAGAG	Variable
pcDNA3.1+_BGHP_R	TAGAAGGCACAGTCGAGG	
pCMV3_T7P_F	TAATACGACTCACTATAGGG	
hSPAG1_Ex14_F	CTTGGCATCCGGCAAAAGAG	163
hSPAG1_Ex15_R	TCGCTGTATTTACTGAGGGCG	
hDNAAF2_Ex1_F	CGGAGATAGCAGTGTGGAAA	116
hDNAAF2_Ex1_R	GTAACGGAGGACACAAAGG	

SPAG1 Knockout by CRISPR/Cas9

hSPAG1_sgRNA_Exon4	GATAATTTGCCTCCAGTTCG	
hSPAG1_Exon4_F	TTTTCTCTATTCTCAGAGTTGGGT	178
hSPAG1_Exon4_R	CCTGCTACCACCCACAAAGA	

[1]ddPCR gene expression probe assays were acquired from Bio-Rad. Sequences of the primers and fluorescent internal oligonucleotide are proprietary, so the region containing the amplicon sequence is provided instead.

[2]Custom ordered ddPCR gene expression probe assay from Bio-Rad to target carboxyl-terminal half of SPAG1 (i.e., 60-kDa SPAG1 isoform).

APPENDIX 4: IDENTIFIED INTERACTORS OF SPAG1 IN DIFFERENTIATED HUMAN BRONCHIAL EPITHEALIAL CELLS[1]

Identified Proteins	Gene ID	*P*-value[2]	Fold Change[3]	Replicate #1				Replicate #2			
				Peptides		Quantitative Value		Peptides		Quantitative Value	
				IgG	SPAG1	IgG	SPAG1	IgG	SPAG1	IgG	SPAG1
Protein kintoun	DNAAF2	< 0.00010	INF	0	45	0	88	0	48	0	117
Protein arginine N-methyltransferase 5	PRMT5	< 0.00010	INF	0	20	0	33	0	20	0	29
Protein ERGIC-53	LMAN1	< 0.00010	INF	0	20	0	54	0	19	0	76
Ubiquitin carboxyl-terminal hydrolase 19	USP19	< 0.00010	INF	0	14	0	14	0	23	0	28
Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	MCCC2	< 0.00010	INF	0	11	0	11	0	11	0	15
PIH1 domain-containing protein 2	PIH1D2	< 0.00010	INF	0	10	0	14	0	15	0	29
Glutamine-dependent NAD(+) synthetase	NADSYN1	< 0.00010	INF	0	6	0	8	0	14	0	21
Dynein assembly factor 1, axonemal	DNAAF1	< 0.00010	INF	0	6	0	6	0	12	0	12
L-xylulose reductase	DCXR	< 0.00010	INF	0	6	0	9	0	10	0	14
COP9 signalosome complex subunit 2	COPS2	0.00072	INF	0	4	0	5	0	6	0	7
Nischarin	NISCH	0.0013	INF	0	4	0	4	0	6	0	7
COP9 signalosome complex subunit 8	COPS8	0.027	INF	0	4	0	4	0	2	0	2
DNA-directed RNA polymerases I, II, and III subunit RPABC1	POLR2E	0.027	INF	0	4	0	4	0	2	0	2
S-phase kinase-associated protein 1	SKP1	0.027	INF	0	4	0	4	0	2	0	2
Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	ECH1	0.0024	INF	0	3	0	3	0	6	0	7
Protein DPCD	DPCD	0.0044	INF	0	3	0	3	0	5	0	6
Cysteine protease ATG4B	ATG4B	0.008	INF	0	3	0	3	0	4	0	5
RNA polymerase II-associated protein 3	RPAP3	0.015	INF	0	3	0	3	0	3	0	4
Dynein heavy chain 6, axonemal	DNAH6	< 0.00010	INF	0	2	0	2	0	25	0	26
COP9 signalosome complex subunit 5	COPS5	0.015	INF	0	2	0	2	0	5	0	5
Acyl-CoA dehydrogenase family member 9, mitochondrial	ACAD9	0.027	INF	0	2	0	2	0	4	0	4
Enoyl-CoA delta isomerase 2, mitochondrial	ECI2	0.027	INF	0	2	0	2	0	4	0	4
NHL repeat-containing protein 2	NHLRC2	0.015	INF	0	2	0	2	0	4	0	5
Disabled homolog 2-interacting protein	DAB2IP	0.049	INF	0	2	0	3	0	2	0	2
IgGFc-binding protein	FCGBP	< 0.00010	52	0	3	0	3	1	38	1	49
Ubiquitin-conjugating enzyme E2 variant 3	UEVLD	< 0.00010	33	0	10	0	12	1	18	1	21

Protein	Gene	P-value	Fold								
Nuclear pore complex protein Nup50	NUP50	0.00018	18	1	8	1	9	0	8	0	9
COP9 signalosome complex subunit 4	COPS4	0.00053	16	0	5	0	8	1	8	1	8
Methylosome protein 50	WDR77	< 0.00010	15	2	9	2	10	0	11	0	19
WD repeat-containing protein 92	WDR92	< 0.00010	14	2	9	2	10	0	13	0	17
COP9 signalosome complex subunit 1	GPS1	0.0046	12	1	5	1	6	0	6	0	6
Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	MCCC1	0.0046	12	1	4	1	4	0	7	0	8
Thioredoxin reductase 1, cytoplasmic	TXNRD1	0.00011	11	2	6	2	8	0	13	0	14
Annexin A11	ANXA11	0.022	9	1	4	1	4	0	5	0	5
Dynein assembly factor 3, axonemal	DNAAF3	0.037	8	0	3	0	4	1	3	1	4
Vesicle transport through interaction with t-SNAREs homolog 1B	VTI1B	0.037	8	1	4	1	4	0	4	0	4
Estradiol 17-beta-dehydrogenase 8	HSD17B8	0.0062	7	1	4	1	5	1	6	1	9
3-hydroxyacyl-CoA dehydrogenase type-2	HSD17B10	< 0.00010	6.4	4	11	4	14	1	10	1	18
Peripheral plasma membrane protein CASK	CASK	0.0029	6	3	6	3	7	0	9	0	11
Carbonyl reductase family member 4	CBR4	0.026	5.5	1	2	1	2	1	6	1	9
RuvB-like 2	RUVBL2	< 0.00010	5	15	28	17	58	8	31	9	73
Dynein heavy chain 5, axonemal	DNAH5	0.04	5	2	4	2	4	0	6	0	6
RuvB-like 1	RUVBL1	< 0.00010	4.5	13	20	16	45	6	21	8	62
Hydroxysteroid dehydrogenase-like protein 2	HSDL2	0.027	4.3	2	5	2	5	1	8	1	8
Heat shock protein 105 kDa	HSPH1	< 0.00010	3.6	10	19	13	25	3	24	3	32
Mitotic checkpoint protein BUB3	BUB3	0.025	3.4	4	9	4	10	1	5	1	7
Cell cycle and apoptosis regulator protein 2	CCAR2	0.011	3.3	6	6	6	7	1	15	1	16
Complement C3	C3	0.0045	3	5	9	6	10	5	22	5	23
E3 ubiquitin-protein ligase TRIM21	TRIM21	< 0.00010	2.9	6	19	7	31	16	25	26	64
Intraflagellar transport protein 80 homolog	IFT80	0.026	2.7	7	11	9	16	0	7	0	8
Neuroblast differentiation-associated protein AHNAK	AHNAK	< 0.00010	2.6	34	53	38	57	3	58	10	67
Tripartite motif-containing protein 29	TRIM29	0.011	2.2	15	13	20	14	1	26	1	33

[1]This table shows the complete list of 52 positive hits from endogenous SPAG1 co-immunoprecipitations in differentiated HBECs. Results from DAVID gene ontology analysis of this list of positive hits are shown in Table 3.4.

[2]P-values were calculated by Fisher's exact test between the SPAG1 and IgG samples. Positive hits were narrowed down using $P<0.05$.

[3]Fold changes were calculated using the spectral counts of the SPAG1 and IgG samples, averaged between the two replicates. Positive hits were narrowed down using a fold change of at least two. INF, infinity or undefined due to zero being denominator.

APPENDIX 5: IDENTIFIED INTERACTORS OF SPAG1 DURING DIFFERENTIATION OF HUMAN BRONCHIAL EPITHELIAL CELLS[1]

		REPLICATE #1						REPLICATE #2					
		Fold Change[2]		Peptides[3]				Fold Change[2]		Peptides[3]			
		Day 0	Day 13	Day 0		Day 13		Day 0	Day 18	Day 0		Day 18	
Identified Proteins	Alternate ID			IgG	SPAG1	IgG	SPAG1			IgG	SPAG1	IgG	SPAG1
Sperm-associated antigen 1	SPAG1	8.18	INF	1	8	0	16	INF	INF	0	8	0	35
E3 ubiquitin-protein ligase TRIM21	TRIM21	INF	INF	0	1	0	10	2.00	4.43	8	11	7	17
Glutamine-dependent NAD(+) synthetase	NADSYN1	INF	INF	0	3	0	4	1.00	INF	1	1	0	6
Small nuclear ribonucleoprotein Sm D2	SNRPD2	INF	2.51	0	1	1	2	0.67	2.00	2	1	1	2
Myeloid-associated differentiation marker	MYADM	INF	0.00	0	2	2	0	0.50	2.00	2	1	1	2
40S ribosomal protein S25	RPS25	INF	INF	0	1	0	3						
Hornerin	HRNR	INF	-	0	2	0	0						
Eukaryotic initiation factor 4A-III	EIF4A3	INF	3.76	0	1	1	5						
60S ribosomal protein L34	RPL34	INF	2.51	0	1	2	4						
Elongation factor 1-delta	EEF1D	INF	2.51	0	1	2	4						
Polymerase I and transcript release factor	PTRF	INF	0.94	0	3	3	3						
Cluster of Plastin-3	PLS3	INF	0.49	0	2	18	7						
MAGUK p55 subfamily member 7	MPP7	-	3.76	0	0	1	3	INF	1.50	0	2	1	3
Protein kintoun	DNAAF2	-	INF	0	0	0	5	-	INF	0	0	0	16
WD repeat-containing protein 92	WDR92	-	INF	0	0	0	5	-	INF	0	0	0	11
ATP-dependent DNA helicase Q1	RECQL	-	INF	0	0	0	2	2.50	1.00	1	5	2	1
Phenylalanine--tRNA ligase alpha subunit	FARSA	-	2.51	0	0	1	2	2.50	INF	2	4	0	1
Microsomal glutathione S-transferase 1	MGST1	-	2.51	0	0	1	2	2.33	1.00	3	4	1	1
Asparagine--tRNA ligase, cytoplasmic	NARS	-	2.51	0	0	2	4	2.33	1.50	7	11	2	3
SH3 domain-binding glutamic acid-rich-like protein 3	SH3BGRL3	-	2.51	0	0	1	2	2.00	1.00	1	2	2	2
T-complex protein 1 subunit beta	CCT2	-	2.51	0	0	1	2	1.80	INF	9	21	0	2
Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADVL	-	4.38	0	0	2	7	1.40	2.33	11	15	3	6
26S proteasome non-ATPase regulatory subunit 7	PSMD7	-	2.51	0	0	1	2	1.33	INF	3	4	0	4
Bifunctional glutamate/proline--tRNA ligase	EPRS	-	2.51	0	0	1	2	1.24	2.50	18	20	2	4
Cathepsin B	CTSB	-	2.51	0	0	1	2	1.20	3.00	4	5	1	3
Annexin A7	ANXA7	-	2.51	0	0	1	2	1.11	3.00	10	9	1	3
Glutamine synthetase	GLUL	-	4.38	0	0	2	7	1.00	3.00	6	8	1	3
Phospholipid hydroperoxide glutathione peroxidase,	GPX4	-	2.51	0	0	1	2	1.00	INF	1	1	0	4

mitochondrial

Calcyphosin	CAPS	-	INF	0	0	0	3	0.67	2.00	3	2	1	2
Mucin-1	MUC1	-	INF	0	0	0	2	0.33	3.00	3	1	1	3
Glutathione peroxidase 1	GPX1	-	2.51	0	0	1	2	0.20	2.50	5	1	2	4
Magnesium transporter protein 1	MAGT1	-	INF	0	0	0	2	0.00	INF	1	0	0	2
Adenosylhomocysteinase	AHCY	-	INF	0	0	0	2						
Annexin A4	ANXA4	-	INF	0	0	0	6						
Cluster of Alpha-enolase	ENO1	-	INF	0	0	0	4						
AP-3 complex subunit beta-1	AP3B1	-	INF	0	0	0	3						
Cluster of Aldo-keto reductase family 1 member C2	AKR1C2	-	INF	0	0	0	2						
N-alpha-acetyltransferase 50	NAA50	-	INF	0	0	0	2						
E3 ubiquitin-protein ligase TRIM56	TRIM56	-	INF	0	0	0	2						
Calpain-2 catalytic subunit	CAPN2	-	INF	0	0	0	2						
Plakophilin-1	PKP1	-	INF	0	0	0	2						
10 kDa heat shock protein, mitochondrial	HSPE1	-	INF	0	0	0	2						
Elongation factor 1-beta	EEF1B2	-	INF	0	0	0	2						
Protein AHNAK2	AHNAK2	-	12.11	0	0	3	28						
Phenylalanine--tRNA ligase beta subunit	FARSB	-	7.52	0	0	1	6						
Transketolase	TKT	-	6.26	0	0	1	5						
Reticulon-4	RTN4	-	6.26	0	0	1	4						
Cluster of Protein arginine N-methyltransferase 1	PRMT1	-	5.01	0	0	1	4						
Thioredoxin	TXN	-	5.01	0	0	1	4						
28S ribosomal protein S22, mitochondrial	MRPS22	-	3.76	0	0	1	3						
Cluster of Interferon-induced GTP-binding protein Mx2	MX2	-	3.76	0	0	1	3						
Triosephosphate isomerase	TPI1	-	3.76	0	0	1	3						
Toll-interacting protein	TOLLIP	-	3.76	0	0	1	2						
Actin-related protein 2/3 complex subunit 1B	ARPC1B	-	3.13	0	0	2	5						
Cluster of Copine-2	CPNE2	-	3.13	0	0	2	5						
Activating signal cointegrator 1	TRIP4	-	2.51	0	0	1	2						
Carnitine O-palmitoyltransferase 1, liver isoform	CPT1A	-	2.51	0	0	1	2						
Sphingosine-1-phosphate lyase 1	SGPL1	-	2.51	0	0	1	2						
FERM, RhoGEF and pleckstrin domain-containing protein 1	FARP1	-	2.51	0	0	1	2						
Proto-oncogene tyrosine-protein kinase Src	SRC	-	2.51	0	0	1	2						
GTP-binding protein 1	GTPBP1	-	2.51	0	0	1	2						

Protein	Gene												
Fatty aldehyde dehydrogenase	ALDH3A2	-	2.51	0	0	1	2						
Destrin	DSTN	-	2.51	0	0	1	2						
FACT complex subunit SSRP1	SSRP1	-	2.51	0	0	1	2						
Adenylyl cyclase-associated protein 1	CAP1	-	2.51	0	0	1	2						
Nucleoside diphosphate kinase B	NME2	-	2.51	0	0	1	2						
Lupus La protein	SSB	-	2.51	0	0	1	2						
Arginine--tRNA ligase, cytoplasmic	RARS	-	2.51	0	0	1	2						
Regulator of nonsense transcripts 2	UPF2	-	2.51	0	0	1	2						
Carcinoembryonic antigen-related cell adhesion molecule 5	CEACAM5	-	2.51	0	0	1	2						
Mitotic checkpoint protein BUB3	BUB3	-	2.51	0	0	2	4						
BPI fold-containing family B member 1	BPIFB1	-	2.51	0	0	2	4						
Actin-related protein 2/3 complex subunit 2	ARPC2	-	2.51	0	0	2	4						
Cluster of Serpin B3	SERPINB3	-	2.51	0	0	1	4						
Brain acid soluble protein 1	BASP1	-	2.51	0	0	2	4						
40S ribosomal protein S21	RPS21	-	2.51	0	0	1	3						
Cluster of Anterior gradient protein 2 homolog	AGR2	-	2.51	0	0	1	3						
Calpain-1 catalytic subunit	CAPN1	-	2.19	0	0	4	7						
Protein S100-A14	S100A14	-	2.09	0	0	3	4						
Glutathione S-transferase P	GSTP1	-	2.09	0	0	2	4						
Galectin-7	LGALS7	5.11	1.02	1	5	6	6	1.25	2.00	4	5	1	2
Kinesin-like protein KIF11	KIF11	5.11	-	1	5	0	0						
Cluster of Eukaryotic initiation factor 4A-I	EIF4A1	4.09	1.18	1	4	15	15						
Cluster of Myosin regulatory light chain 12A	MYL12A	4.09	0.87	1	4	10	9						
Junction plakoglobin	JUP	3.41	0.88	2	10	16	12						
Mitotic interactor and substrate of PLK1	MISP	3.07	1.25	1	3	1	1	2.50	0.78	2	5	9	5
Kinesin-1 heavy chain	KIF5B	3.07	1.88	1	3	2	3	1.68	INF	20	34	0	25
Heat shock 70 kDa protein 4	HSPA4	3.07	1.57	1	3	4	5	1.52	11.00	17	24	1	9
Protein-L-isoaspartate(D-aspartate) O-methyltransferase	PCMT1	3.07	INF	1	3	0	1	0.67	2.00	2	1	1	2
Cluster of Pyruvate kinase PKM	PKM	3.07	2.19	1	3	8	13						
Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	3.07	1.25	1	3	2	3						
Cluster of Casein kinase I isoform alpha	CSNK1A1	3.07	1.11	1	3	9	7						
ATP synthase subunit beta, mitochondrial	ATP5B	3.07	1.10	1	3	6	7						
Cluster of Guanine nucleotide-binding protein G(i) subunit alpha-2	GNAI2	3.07	1.01	1	3	15	15						
Serine/arginine-rich splicing factor 1	SRSF1	3.07	0.54	1	3	5	3						

Protein	Gene												
Filamin-A	FLNA	3.07	0.97	3	9	78	67						
Annexin A2	ANXA2	2.66	2.76	5	12	19	26						
Polyadenylate-binding protein 4	PABPC4	2.56	0.88	2	5	15	13						
Tight junction protein ZO-1	TJP1	2.56	0.48	4	10	27	12						
tRNA-splicing ligase RtcB homolog	RTCB	2.39	0.80	3	7	9	7						
Drebrin	DBN1	2.04	0.54	3	5	5	3	2.00	0.63	1	2	8	4
Inosine-5'-monophosphate dehydrogenase 2	IMPDH2	2.04	1.25	1	2	1	1	2.00	INF	5	10	0	2
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	RPN2	2.04	0.42	1	2	2	1	2.00	0.25	6	8	4	1
Supervillin	SVIL	2.04	0.00	1	2	3	0	2.00	0.62	2	4	28	16
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	2.04	2.00	2	3	5	6	0.69	2.25	11	12	3	7
60S ribosomal protein L8	RPL8	2.04	1.88	1	2	4	5						
MICOS complex subunit MIC19	CHCHD3	2.04	1.88	1	2	2	3						
Poly(U)-binding-splicing factor PUF60	PUF60	2.04	1.57	1	2	4	4						
60S ribosomal protein L12	RPL12	2.04	1.25	2	4	6	6						
X-ray repair cross-complementing protein 5	XRCC5	2.04	1.25	1	2	5	6						
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	2.04	1.25	1	2	4	4						
Cleavage and polyadenylation specificity factor subunit 6	CPSF6	2.04	1.25	1	2	2	2						
Eukaryotic translation initiation factor 3 subunit C	EIF3C	2.04	1.16	1	2	12	12						
Cluster of Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	2.04	1.11	2	4	10	11						
Phosphate carrier protein, mitochondrial	SLC25A3	2.04	1.07	1	2	6	4						
40S ribosomal protein S13	RPS13	2.04	1.04	1	2	5	5						
Spectrin beta chain, non-erythrocytic 4	SPTBN4	2.04	1.04	1	2	3	2						
60S ribosomal protein L7a	RPL7A	2.04	0.98	1	2	11	10						
40S ribosomal protein S11	RPS11	2.04	0.97	1	2	6	6						
40S ribosomal protein S5	RPS5	2.04	0.87	1	1	7	7						
Eukaryotic translation initiation factor 3 subunit F	EIF3F	2.04	0.84	1	2	9	8						
Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	2.04	0.79	2	4	12	11						
Brain-specific angiogenesis inhibitor 1-associated protein 2	BAIAP2	2.04	0.74	1	2	14	10						
Paraspeckle component 1	PSPC1	2.04	0.57	1	2	10	5						
26S proteasome non-ATPase regulatory subunit 2	PSMD2	1.02	2.51	1	1	1	2	1.65	INF	16	23	0	26
T-complex protein 1 subunit theta	CCT8	1.02	5.01	2	2	1	4	0.97	2.25	23	22	4	8
T-complex protein 1 subunit epsilon	CCT5	1.02	3.76	1	1	1	3						
Cluster of L-lactate dehydrogenase B chain	LDHB	1.02	3.76	1	1	1	3						

Protein	Gene												
N-acylneuraminate cytidylyltransferase	CMAS	1.02	2.51	1	1	4	7						
Nascent polypeptide-associated complex subunit alpha, muscle-specific form	NACA	1.02	2.51	1	1	3	4						
40S ribosomal protein S20	RPS20	1.02	2.51	2	2	1	2						
Cluster of T-complex protein 1 subunit zeta	CCT6A	1.02	2.51	1	1	1	2						
Septin-2	SEPT2	1.02	2.51	2	2	2	4						
Histo-blood group ABO system transferase	ABO	1.02	2.51	3	3	1	2	INF	0.00	0	2	1	0
Cluster of Aldehyde dehydrogenase family 1 member A3	ALDH1A3	0.77	7.52	4	3	1	6						
60S ribosomal protein L13a	RPL13A	0.51	2.19	2	1	4	5						
Splicing factor 3B subunit 2	SF3B2	0.00	2.51	1	0	1	2	INF	2.00	0	2	1	2
26S proteasome non-ATPase regulatory subunit 1	PSMD1	0.00	INF	1	0	0	3	2.07	INF	15	25	0	25
Glutamine--tRNA ligase	QARS	0.00	INF	1	0	0	3						
Clathrin light chain A	CLTA	0.00	3.76	1	0	1	3						
Putative RNA-binding protein Luc7-like 1	LUC7L	0.00	3.76	1	0	1	3						
Cofilin-1	CFL1	0.00	3.76	1	0	1	3						
40S ribosomal protein S15	RPS15	0.00	3.76	1	0	1	2						
Ribosome-binding protein 1	RRBP1	0.00	3.13	1	0	3	9						
Cluster of Metallothionein-1G	MT1G	0.00	2.51	2	0	1	2						
Actin-related protein 2/3 complex subunit 4	ARPC4	0.00	2.51	1	0	1	2						
Single-stranded DNA-binding protein, mitochondrial	SSBP1	0.00	2.51	1	0	1	2						
Protein LSM12 homolog	LSM12	0.00	2.51	1	0	1	2						
Dynein heavy chain 6, axonemal	DNAH6							INF	INF	0	1	0	2
Aminoacyl tRNA synthase complex-interacting multifunctional protein 1	AIMP1							INF	INF	0	3	0	1
Cluster of BTB/POZ domain-containing adapter for CUL3-mediated RhoA degradation protein 3	KCTD10							INF	INF	0	7	0	6
COP9 signalosome complex subunit 6	COPS6							INF	-	0	2	0	0
WD repeat-containing protein 61	WDR61							INF	-	0	2	0	0
39S ribosomal protein L38, mitochondrial	MRPL38							INF	-	0	2	0	0
Beta-galactosidase	GLB1							INF	-	0	4	0	0
Matrix metalloproteinase-14	MMP14							INF	-	0	4	0	0
Amidophosphoribosyltransferase	PPAT							INF	-	0	4	0	0
Dynactin subunit 4	DCTN4							INF	-	0	3	0	0
Cluster of Glutathione S-transferase Mu 1	GSTM1							INF	-	0	3	0	0
NudC domain-containing protein 1	NUDCD1							INF	-	0	3	0	0
Bifunctional coenzyme A synthase	COASY							INF	-	0	3	0	0
Protein disulfide-isomerase A5	PDIA5							INF	-	0	3	0	0

Microtubule-associated protein 2	MAP2	INF	-	0	3	0	0
Interleukin-18	IL18	INF	-	0	3	0	0
Caspase-7	CASP7	INF	-	0	3	0	0
Heme-binding protein 1	HEBP1	INF	-	0	3	0	0
Retinoic acid-induced protein 1	RAI1	INF	-	0	2	0	0
Uridine 5'-monophosphate synthase	UMPS	INF	-	0	2	0	0
Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma	PIP4K2C	INF	-	0	2	0	0
N-alpha-acetyltransferase 25, NatB auxiliary subunit	NAA25	INF	-	0	2	0	0
Ephrin-B1	EFNB1	INF	-	0	2	0	0
Retinol-binding protein 1	RBP1	INF	-	0	2	0	0
All-trans-retinol 13,14-reductase	RETSAT	INF	-	0	2	0	0
BTB/POZ domain-containing adapter for CUL3-mediated RhoA degradation protein 2	TNFAIP1	INF	-	0	2	0	0
Serine/threonine-protein phosphatase 5	PPP5C	INF	-	0	2	0	0
1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-1	PLCD1	INF	-	0	2	0	0
Osteoclast-stimulating factor 1	OSTF1	INF	-	0	2	0	0
Signal transducer and activator of transcription 6	STAT6	INF	-	0	2	0	0
Cytochrome c-type heme lyase	HCCS	INF	-	0	2	0	0
Protein phosphatase methylesterase 1	PPME1	INF	-	0	2	0	0
N-acetylglucosamine-6-sulfatase	GNS	INF	-	0	2	0	0
Vacuolar protein sorting-associated protein 45	VPS45	INF	-	0	2	0	0
A disintegrin and metalloproteinase with thrombospondin motifs 4	ADAMTS4	INF	-	0	2	0	0
Nectin-1	NECTIN1	INF	-	0	2	0	0
Synaptotagmin-like protein 1	SYTL1	INF	-	0	2	0	0
ATPase family AAA domain-containing protein 1	ATAD1	INF	-	0	2	0	0
Ceramide synthase 2	CERS2	INF	-	0	2	0	0
Casein kinase II subunit alpha'	CSNK2A2	INF	-	0	2	0	0
Tetratricopeptide repeat protein 38	TTC38	INF	-	0	2	0	0
ELMO domain-containing protein 2	ELMOD2	INF	-	0	2	0	0
Myotrophin	MTPN	INF	-	0	2	0	0
DnaJ homolog subfamily B member 4	DNAJB4	INF	-	0	2	0	0
Cluster of 5'-AMP-activated protein kinase subunit gamma-1	PRKAG1	INF	2.00	0	2	1	2
Toll-like receptor 5	TLR5	INF	2.00	0	1	5	9
Bifunctional polynucleotide phosphatase/kinase	PNKP	INF	2.00	0	1	1	2
DAZ-associated protein 1	DAZAP1	INF	1.00	0	3	2	2

Cyclin-G-associated kinase	GAK	INF	1.00	0	2	18	18
La-related protein 4B	LARP4B	INF	1.00	0	2	5	4
Cluster of Tubulin gamma-1 chain	TUBG1	INF	1.00	0	2	2	2
Focal adhesion kinase 1	PTK2	INF	1.00	0	2	2	1
Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	INF	1.00	0	2	1	1
Ras GTPase-activating protein-binding protein 1	G3BP1	INF	0.95	0	7	12	10
Retinoic acid-induced protein 3	GPRC5A	INF	0.50	0	2	2	1
Dolichol-phosphate mannosyltransferase subunit 1	DPM1	INF	0.50	0	2	2	1
COP9 signalosome complex subunit 5	COPS5	INF	0.00	0	2	1	0
Cation-independent mannose-6-phosphate receptor	IGF2R	INF	0.00	0	4	7	0
ATP synthase subunit delta, mitochondrial	ATP5D	INF	0.00	0	2	1	0
Glycogen synthase kinase-3 beta	GSK3B	INF	0.00	0	2	2	0
Ubiquitin carboxyl-terminal hydrolase 19	USP19	-	INF	0	0	0	3
Dynein assembly factor 1, axonemal	DNAAF1	-	INF	0	0	0	10
Protein DPCD	DPCD	-	INF	0	0	0	2
RNA polymerase II-associated protein 3	RPAP3	-	INF	0	0	0	2
Dynein heavy chain 9, axonemal	DNAH9	-	INF	0	0	0	2
WD repeat-containing protein 24	WDR24	-	INF	0	0	0	3
26S proteasome non-ATPase regulatory subunit 10	PSMD10	-	INF	0	0	0	5
Thioredoxin domain-containing protein 6	NME9	-	INF	0	0	0	3
Glutamate-rich protein 3	ERICH3	-	INF	0	0	0	2
Transcriptional regulator ATRX	ATRX	-	INF	0	0	0	2
Protein SMG7	SMG7	-	INF	0	0	0	2
Pleckstrin homology domain-containing family A member 7	PLEKHA7	-	INF	0	0	0	2
Cilia- and flagella-associated protein 57	CFAP57	-	INF	0	0	0	2
Pre-mRNA-processing factor 19	PRPF19	-	INF	0	0	0	2
Cartilage oligomeric matrix protein	COMP	-	INF	0	0	0	2
Dynein heavy chain 5, axonemal	DNAH5	-	5.00	0	0	1	3
Dynein intermediate chain 1, axonemal	DNAI1	-	4.00	0	0	2	7
Nucleoside diphosphate kinase 7	NME7	-	3.00	0	0	1	3
RNA-binding protein 33	RBM33	-	3.00	0	0	1	3
TRMT1-like protein	TRMT1L	-	2.67	0	0	3	7
OTU domain-containing protein 4	OTUD4	-	2.50	0	0	2	4
Palmdelphin	PALMD	-	2.50	0	0	2	4
SLIT-ROBO Rho GTPase-activating protein 1	SRGAP1	-	2.00	0	0	1	4

Protein	Gene						
USP6 N-terminal-like protein	USP6NL	-	2.00	0	0	1	2
Structural maintenance of chromosomes protein 3	SMC3	-	2.00	0	0	1	2
Ral GTPase-activating protein subunit beta	RALGAPB	-	2.00	0	0	3	5
Aftiphilin	AFTPH	-	2.00	0	0	1	2
DNA dC->dU-editing enzyme APOBEC-3G	APOBEC3G	-	2.00	0	0	0	2
Protein LSM14 homolog B	LSM14B	-	2.00	0	0	1	2
RNA-binding protein 8A	RBM8A	-	2.00	0	0	1	2
Leucine-rich repeat-containing protein 49	LRRC49	-	2.00	0	0	1	2
Regulator of nonsense transcripts 3B	UPF3B	-	2.00	0	0	1	2
Tripeptidyl-peptidase 1	TPP1	-	2.00	0	0	1	2
Triokinase/FMN cyclase	TKFC	7.00	-	1	6	0	0
26S proteasome non-ATPase regulatory subunit 14	PSMD14	5.00	INF	1	3	0	4
NAD-dependent malic enzyme, mitochondrial	ME2	5.00	-	1	5	0	0
Serine/threonine-protein kinase MRCK gamma	CDC42BPG	5.00	-	1	4	0	0
Translationally-controlled tumor protein	TPT1	5.00	-	1	4	0	0
Eukaryotic translation initiation factor 5	EIF5	5.00	2.50	1	5	2	4
Tumor susceptibility gene 101 protein	TSG101	5.00	1.67	1	5	3	4
26S protease regulatory subunit 6B	PSMC4	4.33	INF	3	13	0	16
Cluster of NACHT, LRR and PYD domains-containing protein 2	NLRP2	4.00	INF	1	4	0	1
26S proteasome non-ATPase regulatory subunit 8	PSMD8	4.00	INF	1	3	0	3
Elongation factor G, mitochondrial	GFM1	4.00	-	1	4	0	0
Probable E3 ubiquitin-protein ligase HERC4	HERC4	4.00	-	1	4	0	0
Cytosolic purine 5'-nucleotidase	NT5C2	4.00	-	1	3	0	0
Septin-10	SEPT10	4.00	5.00	1	4	1	4
Casein kinase II subunit alpha 3	CSNK2A3	4.00	1.25	1	3	3	4
40S ribosomal protein S17	RPS17	4.00	0.50	1	4	4	3
26S proteasome non-ATPase regulatory subunit 4	PSMD4	3.67	14.00	3	9	1	8
Lon protease homolog, mitochondrial	LONP1	3.50	-	2	8	0	0
60S ribosomal protein L10a	RPL10A	3.50	0.88	2	5	7	6
26S proteasome non-ATPase regulatory subunit 3	PSMD3	3.20	INF	6	16	0	15
Importin subunit alpha-3	KPNA4	3.00	INF	1	2	0	0
Calcyclin-binding protein	CACYBP	3.00	INF	1	3	0	1
Importin subunit alpha-5	KPNA1	3.00	INF	0	2	0	1
Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	MCCC2	3.00	-	1	3	0	0

Protein	Gene					
Suppressor of tumorigenicity 14 protein	ST14	3.00	-	1	3	0	0
Clustered mitochondria protein homolog	CLUH	3.00	-	2	6	0	0
Acylamino-acid-releasing enzyme	APEH	3.00	-	2	5	0	0
Growth factor receptor-bound protein 2	GRB2	3.00	-	1	3	0	0
Integrin-linked protein kinase	ILK	3.00	-	1	3	0	0
ER membrane protein complex subunit 1	EMC1	3.00	-	1	3	0	0
Peptidyl-prolyl cis-trans isomerase FKBP5	FKBP5	3.00	-	1	3	0	0
Serine palmitoyltransferase 2	SPTLC2	3.00	-	1	3	0	0
Coiled-coil domain-containing protein 47	CCDC47	3.00	-	1	3	0	0
PEST proteolytic signal-containing nuclear protein	PCNP	3.00	-	1	3	0	0
Vesicle-associated membrane protein 8	VAMP8	3.00	-	1	3	0	0
Squalene monooxygenase	SQLE	3.00	-	1	3	0	0
Protein SGT1 homolog	SUGT1	3.00	-	1	3	0	0
Cluster of Adenylate kinase isoenzyme 1	AK1	3.00	-	1	3	0	0
NEDD8-activating enzyme E1 regulatory subunit	NAE1	3.00	-	1	3	0	0
Superoxide dismutase [Mn], mitochondrial	SOD2	3.00	-	1	3	0	0
Oligoribonuclease, mitochondrial	REXO2	3.00	-	1	3	0	0
1,4-alpha-glucan-branching enzyme	GBE1	3.00	-	1	3	0	0
Ubiquitin-protein ligase E3A	UBE3A	3.00	-	1	3	0	0
Adapter molecule crk	CRK	3.00	-	1	3	0	0
Catechol O-methyltransferase	COMT	3.00	-	1	3	0	0
D-tyrosyl-tRNA(Tyr) deacylase 1	DTD1	3.00	-	1	3	0	0
Secretory carrier-associated membrane protein 3	SCAMP3	3.00	-	1	3	0	0
EGF domain-specific O-linked N-acetylglucosamine transferase	EOGT	3.00	-	1	3	0	0
Vacuolar protein sorting-associated protein 53 homolog	VPS53	3.00	-	1	2	0	0
Homer protein homolog 3	HOMER3	3.00	-	1	2	0	0
Vesicle-associated membrane protein-associated protein A	VAPA	3.00	2.00	1	5	1	1
Galectin-3	LGALS3	3.00	2.00	1	2	1	2
Cluster of KH domain-containing, RNA-binding, signal transduction-associated protein 1	KHDRBS1	3.00	1.33	1	3	4	5
Oxysterol-binding protein-related protein 8	OSBPL8	3.00	1.00	1	3	1	1
Cluster of Histone deacetylase 1	HDAC1	3.00	1.00	1	3	1	1
Profilin-2	PFN2	3.00	0.50	1	3	2	1
Chromobox protein homolog 3	CBX3	3.00	0.00	2	5	1	
C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	2.77	15.50	13	32	2	20

Prolyl endopeptidase	PREP	2.67	-	3	9	0	0
Coronin-1C	CORO1C	2.67	0.67	3	9	8	5
ELKS/Rab6-interacting/CAST family member 1	ERC1	2.67	0.00	3	8	2	0
Apoptosis inhibitor 5	API5	2.50	-	2	5	0	0
Prenylcysteine oxidase 1	PCYOX1	2.50	-	2	5	0	0
Pyridoxal kinase	PDXK	2.50	-	2	4	0	0
Proteasome subunit alpha type-2	PSMA2	2.50	-	2	4	0	0
Endophilin-B2	SH3GLB2	2.50	-	2	4	0	0
Cluster of Ras-related protein Rap-2a	RAP2A	2.50	-	1	3	0	0
Protein LSM14 homolog A	LSM14A	2.50	1.25	2	5	4	4
Splicing factor 3B subunit 1	SF3B1	2.50	1.00	2	5	1	1
Mitotic interactor and substrate of PLK1	MISP	2.50	0.78	2	5	9	5
Arf-GAP domain and FG repeat-containing protein 1	AGFG1	2.50	0.33	2	4	3	1
Thioredoxin reductase 1, cytoplasmic	TXNRD1	2.33	-	3	6	0	0
Epidermal growth factor receptor	EGFR	2.33	-	3	7	0	0
Cluster of Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	PPP3CA	2.33	1.00	3	7	1	1
Cluster of Kinesin light chain 1	KLC1	2.27	INF	12	20	0	12
CAD protein	CAD	2.20	-	5	12	0	0
26S protease regulatory subunit 10B	PSMC6	2.14	INF	7	12	0	14
WD repeat-containing protein 1	WDR1	2.11	0.67	9	16	3	2
26S proteasome non-ATPase regulatory subunit 6	PSMD6	2.00	INF	7	11	0	16
Cysteine-rich protein 2	CRIP2	2.00	INF	1	2	0	2
Monofunctional C1-tetrahydrofolate synthase, mitochondrial	MTHFD1L	2.00	INF	3	6	0	7
Alpha-centractin	ACTR1A	2.00	INF	2	4	0	1
FAS-associated factor 2	FAF2	2.00	INF	2	4	0	1
Glutathione S-transferase kappa 1	GSTK1	2.00	INF	2	4	0	1
NHL repeat-containing protein 2	NHLRC2	2.00	-	1	2	0	0
28S ribosomal protein S9, mitochondrial	MRPS9	2.00	-	1	2	0	0
Translocating chain-associated membrane protein 1	TRAM1	2.00	-	1	2	0	0
Actin-related protein 2/3 complex subunit 5	ARPC5	2.00	-	1	2	0	0
D-beta-hydroxybutyrate dehydrogenase, mitochondrial	BDH1	2.00	-	1	2	0	0
Trifunctional purine biosynthetic protein adenosine-3	GART	2.00	-	5	11	0	0
ERO1-like protein alpha	ERO1A	2.00	-	6	10	0	0
Ribosome maturation protein SBDS	SBDS	2.00	-	4	8	0	0
Importin subunit alpha-1	KPNA2	2.00	-	3	6	0	0

Hepatocyte growth factor-regulated tyrosine kinase substrate	HGS	2.00	-	3	6	0	0
Copine-1	CPNE1	2.00	-	3	6	0	0
Redox-regulatory protein FAM213A	FAM213A	2.00	-	3	5	0	0
Translocon-associated protein subunit delta	SSR4	2.00	-	3	5	0	0
Elongator complex protein 3	ELP3	2.00	-	2	4	0	0
Dedicator of cytokinesis protein 9	DOCK9	2.00	-	2	4	0	0
Signal recognition particle 54 kDa protein	SRP54	2.00	-	2	4	0	0
Protein bicaudal D homolog 2	BICD2	2.00	-	2	4	0	0
ADP-ribosylation factor GTPase-activating protein 3	ARFGAP3	2.00	-	2	4	0	0
Serine hydroxymethyltransferase, cytosolic	SHMT1	2.00	-	3	4	0	0
Glycylpeptide N-tetradecanoyltransferase 1	NMT1	2.00	-	2	4	0	0
Retinol dehydrogenase 10	RDH10	2.00	-	2	4	0	0
Acetyl-CoA acetyltransferase, cytosolic	ACAT2	2.00	-	2	4	0	0
DNA-(apurinic or apyrimidinic site) lyase	APEX1	2.00	-	2	4	0	0
Crk-like protein	CRKL	2.00	-	2	4	0	0
Alpha-2-macroglobulin receptor-associated protein	LRPAP1	2.00	-	1	2	0	0
MOB kinase activator 1B	MOB1B	2.00	-	1	2	0	0
Lysine-specific demethylase 3B	KDM3B	2.00	-	1	2	0	0
ADP-ribosylation factor GTPase-activating protein 2	ARFGAP2	2.00	-	1	2	0	0
Thymidylate kinase	DTYMK	2.00	-	1	2	0	0
Eukaryotic translation initiation factor 1A, X-chromosomal	EIF1AX	2.00	-	1	2	0	0
Translocation protein SEC63 homolog	SEC63	2.00	-	1	2	0	0
Ancient ubiquitous protein 1	AUP1	2.00	-	1	2	0	0
2',5'-phosphodiesterase 12	PDE12	2.00	-	1	2	0	0
SUMO-conjugating enzyme UBC9	UBE2I	2.00	-	1	2	0	0
Lysophospholipid acyltransferase LPCAT4	LPCAT4	2.00	-	1	2	0	0
Lysophosphatidylcholine acyltransferase 1	LPCAT1	2.00	-	1	2	0	0
Branched-chain-amino-acid aminotransferase, mitochondrial	BCAT2	2.00	-	1	2	0	0
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial	NDUFA10	2.00	-	1	2	0	0
DNA replication licensing factor MCM2	MCM2	2.00	-	1	2	0	0
Histone-arginine methyltransferase CARM1	CARM1	2.00	-	1	2	0	0
WASH complex subunit 5	WASHC5	2.00	-	1	2	0	0
Acyl-coenzyme A thioesterase 1	ACOT1	2.00	-	1	2	0	0
Zinc transporter 7	SLC30A7	2.00	-	1	2	0	0

Mixed lineage kinase domain-like protein	MLKL	2.00	-	1	2	0	0
cAMP-dependent protein kinase type I-alpha regulatory subunit	PRKAR1A	2.00	-	1	2	0	0
Fructose-2,6-bisphosphatase TIGAR	TIGAR	2.00	-	1	2	0	0
Tissue alpha-L-fucosidase	FUCA1	2.00	-	1	2	0	0
RNA-binding protein 12	RBM12	2.00	-	1	2	0	0
Leukocyte surface antigen CD47	CD47	2.00	-	1	2	0	0
Glutamate--cysteine ligase regulatory subunit	GCLM	2.00	-	1	2	0	0
Ankyrin repeat domain-containing protein SOWAHC	SOWAHC	2.00	-	1	2	0	0
Uroporphyrinogen decarboxylase	UROD	2.00	-	1	2	0	0
Normal mucosa of esophagus-specific gene 1 protein	NMES1	2.00	-	1	2	0	0
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	PLOD2	2.00	-	1	2	0	0
Cathepsin Z	CTSZ	2.00	-	1	2	0	0
Prothymosin alpha	PTMA	2.00	-	1	2	0	0
Protein NOXP20	FAM114A1	2.00	-	1	2	0	0
Amyloid beta A4 protein	APP	2.00	-	1	2	0	0
Protein NDRG2	NDRG2	2.00	-	1	2	0	0
Adenylosuccinate lyase	ADSL	2.00	-	1	2	0	0
Long-chain fatty acid transport protein 4	SLC27A4	2.00	-	1	2	0	0
Density-regulated protein	DENR	2.00	-	1	2	0	0
Transcription factor BTF3	BTF3	2.00	-	1	2	0	0
Elongation factor Ts, mitochondrial	TSFM	2.00	-	1	2	0	0
Peptidyl-prolyl cis-trans isomerase FKBP2	FKBP2	2.00	-	1	2	0	0
Glucosamine 6-phosphate N-acetyltransferase	GNPNAT1	2.00	-	1	2	0	0
Proteasome inhibitor PI31 subunit	PSMF1	2.00	-	1	2	0	0
Eukaryotic translation initiation factor 1b	EIF1B	2.00	-	1	2	0	0
Esterase OVCA2	OVCA2	2.00	-	1	2	0	0
Proline synthase co-transcribed bacterial homolog protein	PROSC	2.00	-	1	2	0	0
26S proteasome non-ATPase regulatory subunit 11	PSMD11	2.00	38.00	9	16	1	21
Pre-mRNA-processing-splicing factor 8	PRPF8	2.00	3.00	2	4	1	3
UPF0568 protein C14orf166	C14orf166	2.00	2.67	1	2	3	5
Cluster of cAMP-dependent protein kinase type II-alpha regulatory subunit	PRKAR2A	2.00	2.00	3	6	1	2
60S ribosomal protein L27a	RPL27A	2.00	2.00	2	4	3	5
CD109 antigen	CD109	2.00	2.00	2	4	1	2
Vimentin	VIM	2.00	2.00	0	2	0	0

Cluster of RNA-binding protein 4	RBM4	2.00	1.50	1	2	4	5
Heterogeneous nuclear ribonucleoprotein A3	HNRNPA3	2.00	1.38	1	2	6	5
ATP-binding cassette sub-family D member 3	ABCD3	2.00	1.29	1	2	6	7
Cluster of Lamina-associated polypeptide 2, isoform alpha	TMPO	2.00	1.22	2	4	9	9
Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	BAIAP2L1	2.00	1.20	1	2	5	5
60S ribosomal protein L21	RPL21	2.00	1.14	2	4	5	4
Alpha-soluble NSF attachment protein	NAPA	2.00	1.00	2	4	2	2
Aflatoxin B1 aldehyde reductase member 2	AKR7A2	2.00	1.00	1	2	1	1
Chloride intracellular channel protein 6	CLIC6	2.00	1.00	2	4	5	5
Far upstream element-binding protein 3	FUBP3	2.00	1.00	1	3	5	4
ADP-ribosylation factor 6	ARF6	2.00	1.00	2	3	1	1
Leucine-rich repeat flightless-interacting protein 2	LRRFIP2	2.00	1.00	1	2	8	9
Cluster of Mitochondrial glutamate carrier 1	SLC25A22	2.00	1.00	1	2	1	1
Eukaryotic translation elongation factor 1 epsilon-1	EEF1E1	2.00	1.00	1	2	1	1
Cluster of ATPase family AAA domain-containing protein 3A	ATAD3A	2.00	0.94	4	9	24	25
Serine/threonine-protein kinase MRCK beta	CDC42BPB	2.00	0.90	1	2	8	8
AP-2 complex subunit alpha-1	AP2A1	2.00	0.89	5	7	28	26
Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	2.00	0.89	1	3	4	4
SPATS2-like protein	SPATS2L	2.00	0.85	1	2	11	9
F-actin-capping protein subunit alpha-2	CAPZA2	2.00	0.75	1	3	2	2
Elongin-C	ELOC	2.00	0.67	2	4	3	2
Alpha-taxilin	TXLNA	2.00	0.50	2	4	2	1
ELAV-like protein 1	ELAVL1	2.00	0.46	1	2	6	5
BUB3-interacting and GLEBS motif-containing protein ZNF207	ZNF207	2.00	0.33	1	2	2	1
Dynamin-like 120 kDa protein, mitochondrial	OPA1	2.00	0.00	3	6	3	0
Coatomer subunit epsilon	COPE	2.00	0.00	3	5	1	0
Splicing factor 3A subunit 1	SF3A1	2.00	0.00	1	2	1	0
Caspase-3	CASP3	2.00	0.00	1	2	1	0
C-terminal-binding protein 2	CTBP2	2.00	0.00	0	2	2	0
Ras-related protein Ral-B	RALB	2.00	0.00	1	2	0	0
26S protease regulatory subunit 8	PSMC5	1.90	INF	7	14	0	16
Cluster of Copine-3	CPNE3	1.83	3.50	7	12	2	6
Prohibitin-2	PHB2	1.83	2.50	7	11	2	4
Cadherin-3	CDH3	1.67	2.00	3	4	1	2

26S protease regulatory subunit 4	PSMC1	1.62	INF	12	18	0	18
26S proteasome non-ATPase regulatory subunit 12	PSMD12	1.57	INF	8	10	0	14
Proteasomal ubiquitin receptor ADRM1	ADRM1	1.50	INF	2	3	0	4
Sorbitol dehydrogenase	SORD	1.50	INF	2	3	0	2
26S protease regulatory subunit 7	PSMC2	1.50	9.67	11	13	3	16
Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	HACD3	1.50	5.00	2	3	1	4
Splicing factor 3B subunit 3	SF3B3	1.50	3.00	4	6	1	3
Cluster of 60S ribosomal protein L36a-like	RPL36AL	1.50	2.50	2	3	2	4
Heat shock protein 105 kDa	HSPH1	1.42	3.00	9	13	4	11
26S proteasome non-ATPase regulatory subunit 5	PSMD5	1.40	INF	6	7	0	3
Cluster of Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform	PPP2CB	1.40	INF	5	7	0	2
Catalase	CAT	1.40	3.33	5	8	3	9
DnaJ homolog subfamily C member 7	DNAJC7	1.40	2.00	5	7	1	2
Four and a half LIM domains protein 2	FHL2	1.40	2.00	6	6	1	2
26S proteasome non-ATPase regulatory subunit 13	PSMD13	1.38	INF	8	10	0	11
26S protease regulatory subunit 6A	PSMC3	1.29	31.00	6	8	1	19
Thioredoxin-like protein 1	TXNL1	1.25	INF	3	5	0	3
Aspartyl/asparaginyl beta-hydroxylase	ASPH	1.25	2.20	4	5	5	8
Nuclear migration protein nudC	NUDC	1.20	2.50	6	5	2	4
General transcription factor II-I	GTF2I	1.20	2.00	6	5	5	9
Calcium-binding mitochondrial carrier protein Aralar1	SLC25A12	1.17	2.00	3	3	2	3
Transitional endoplasmic reticulum ATPase	VCP	1.05	4.50	25	28	2	7
Cluster of Ras-related protein Rab-10	RAB10	1.04	2.33	23	23	2	3
Twinfilin-1	TWF1	1.00	INF	4	5	0	2
Transcription intermediary factor 1-beta	TRIM28	1.00	6.00	5	5	1	5
Exocyst complex component 4	EXOC4	1.00	5.00	3	3	1	4
Anterior gradient protein 3 homolog	AGR3	1.00	3.00	0	0	1	2
NAD(P)H dehydrogenase [quinone] 1	NQO1	1.00	3.00	3	4	1	3
Tetratricopeptide repeat protein 37	TTC37	1.00	3.00	2	2	1	3
Cluster of Ras-related protein Rab-5C	RAB5C	1.00	2.50	8	8	2	3
DnaJ homolog subfamily C member 10	DNAJC10	1.00	2.50	2	2	2	4
Adaptin ear-binding coat-associated protein 2	NECAP2	1.00	2.50	1	1	2	4
Cytochrome c oxidase subunit NDUFA4	NDUFA4	1.00	2.00	1	1	1	2
ATP-dependent zinc metalloprotease YME1L1	YME1L1	1.00	2.00	1	1	1	2

Creatine kinase B-type	CKB	1.00	2.00	7	6	1	2
Poly [ADP-ribose] polymerase 4	PARP4	1.00	2.00	5	5	1	2
Galectin-3-binding protein	LGALS3BP	1.00	2.00	4	4	3	5
Acyl-coenzyme A thioesterase 9, mitochondrial	ACOT9	1.00	2.00	3	3	3	5
ATP synthase F(0) complex subunit B1, mitochondrial	ATP5F1	1.00	2.00	3	3	1	2
U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	1.00	2.00	2	1	1	2
60S ribosomal protein L35	RPL35	1.00	2.00	1	1	1	2
Coatomer subunit gamma-1	COPG1	0.94	INF	14	10	0	2
Cluster of Talin-1	TLN1	0.91	INF	44	42	0	2
Vacuolar protein sorting-associated protein 35	VPS35	0.91	2.00	11	11	1	2
Alcohol dehydrogenase class 4 mu/sigma chain	ADH7	0.86	INF	8	6	0	2
Stromal interaction molecule 1	STIM1	0.80	INF	5	4	0	2
Cluster of Mitogen-activated protein kinase 1	MAPK1	0.80	3.00	3	4	2	4
LIM and SH3 domain protein 1	LASP1	0.80	2.00	9	9	1	2
ATP synthase subunit g, mitochondrial	ATP5L	0.75	INF	4	3	0	2
Eukaryotic translation initiation factor 3 subunit J	EIF3J	0.75	2.50	4	3	2	4
Band 4.1-like protein 1	EPB41L1	0.75	2.00	4	3	3	6
Mitochondrial ribonuclease P protein 1	TRMT10C	0.75	2.00	4	3	1	2
Cluster of Serine/threonine-protein kinase 24	STK24	0.70	2.00	10	6	1	2
Ferrochelatase, mitochondrial	FECH	0.67	INF	3	2	0	2
AP2-associated protein kinase 1	AAK1	0.67	2.25	3	2	4	7
Coatomer subunit delta	ARCN1	0.64	5.00	12	8	1	4
Cluster of Guanine nucleotide-binding protein G(s) subunit alpha isoforms XLas	GNAS	0.50	4.50	3	1	1	4
GRB10-interacting GYF protein 2	GIGYF2	0.50	3.00	2	1	2	4
Eukaryotic translation initiation factor 4E	EIF4E	0.50	2.00	3	2	3	4
Polyribonucleotide nucleotidyltransferase 1, mitochondrial	PNPT1	0.50	2.00	2	1	1	2
Exocyst complex component 3	EXOC3	0.33	INF	3	1	0	2
AFG3-like protein 2	AFG3L2	0.33	2.50	3	1	2	4
Methylosome protein 50	WDR77	0.33	2.00	3	1	1	2
Cleavage and polyadenylation specificity factor subunit 5	NUDT21	0.25	2.00	4	1	4	6
Syndecan-1	SDC1	0.25	2.00	3	1	1	2
Cysteine--tRNA ligase, cytoplasmic	CARS	0.14	INF	8	1	0	2
von Willebrand factor A domain-containing protein 8	VWA8	0.00	INF	1	0	0	3
Cluster of cAMP-dependent protein kinase catalytic subunit beta	PRKACB	0.00	INF	1	0	0	2

Small nuclear ribonucleoprotein E	SNRPE	0.00	3.00	2	0	1	2
SUN domain-containing protein 2	SUN2	0.00	2.00	1	0	1	2
Vacuolar protein sorting-associated protein 29	VPS29	0.00	2.00	3	0	1	2

[1]This table shows the complete list of 506 positive hits from endogenous SPAG1 co-immunoprecipitations on ALI day 0, 13, and 18 in differentiating HBECs. Results from DAVID gene ontology analysis of this list of positive hits are shown in Appendix 6.

[2]Fold changes were calculated using the spectral counts of the SPAG1 and IgG samples. Positive hits were narrowed down using a fold change of at least two for at least one time point. INF, infinity or undefined due to zero being denominator.

[3]Positive hits were narrowed down based on the proteins having at least two unique peptides for a protein for at least one time point.

APPENDIX 6: DAVID GENE ONTOLOGY ANALYSIS ON SPAG1 INTERACTORS IN DIFFERNTIATING HUMAN BRONCHIAL EPITHELIAL CELLS

Term	Count	%	P-Value	Benjamini
ALI Day 0				
poly(A) RNA binding	65	21.5	1.40E-17	8.70E-15
protein binding	208	68.9	7.60E-11	2.30E-08
regulation of cellular amino acid metabolic process	11	3.6	1.60E-08	2.40E-05
translational initiation	16	5.3	1.80E-08	1.40E-05
regulation of mRNA stability	14	4.6	2.90E-08	1.50E-05
cadherin binding involved in cell-cell adhesion	22	7.3	4.00E-08	8.20E-06
antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	11	3.6	1.30E-07	5.10E-05
NIK/NF-kappaB signaling	11	3.6	2.10E-07	6.40E-05
negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	11	3.6	4.20E-07	1.10E-04
Wnt signaling pathway, planar cell polarity pathway	12	4	6.00E-07	1.30E-04
positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	11	3.6	8.00E-07	1.50E-04
negative regulation of canonical Wnt signaling pathway	15	5	1.10E-06	1.90E-04
anaphase-promoting complex-dependent catabolic process	11	3.6	1.20E-06	1.80E-04
positive regulation of canonical Wnt signaling pathway	13	4.3	1.30E-06	1.80E-04
cell-cell adhesion	19	6.3	1.40E-06	1.90E-04
proteasome-mediated ubiquitin-dependent protein catabolic process	16	5.3	3.00E-06	3.60E-04
SRP-dependent cotranslational protein targeting to membrane	11	3.6	5.70E-06	6.30E-04
stimulatory C-type lectin receptor signaling pathway	11	3.6	1.50E-05	1.60E-03
MAPK cascade	17	5.6	1.60E-05	1.60E-03
RNA binding	25	8.3	3.40E-05	5.20E-03
nucleotide binding	19	6.3	4.10E-05	5.10E-03
tumor necrosis factor-mediated signaling pathway	11	3.6	4.30E-05	3.90E-03
ATP binding	48	15.9	5.10E-05	5.20E-03
mRNA binding	11	3.6	6.70E-05	5.90E-03
Fc-epsilon receptor signaling pathway	13	4.3	7.20E-05	6.20E-03
translation initiation factor activity	8	2.6	8.80E-05	6.80E-03
translation	15	5	1.60E-04	1.30E-02
transferase activity	9	3	2.70E-04	1.80E-02
T cell receptor signaling pathway	11	3.6	2.80E-04	2.20E-02
liver development	8	2.6	3.10E-04	2.30E-02
oxidation-reduction process	24	7.9	3.20E-04	2.20E-02
mRNA processing	12	4	3.20E-04	2.10E-02
protein polyubiquitination	12	4	4.10E-04	2.60E-02
actin crosslink formation	4	1.3	7.90E-04	4.80E-02
viral transcription	9	3	7.90E-04	4.60E-02
purine nucleotide biosynthetic process	4	1.3	1.00E-03	5.80E-02
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	9	3	1.20E-03	6.30E-02

purine ribonucleoside monophosphate biosynthetic process	4	1.3	1.30E-03	6.90E-02
structural constituent of ribosome	12	4	1.80E-03	1.10E-01
mRNA splicing, via spliceosome	12	4	1.90E-03	9.30E-02
ADP binding	5	1.7	2.50E-03	1.30E-01
glutathione metabolic process	6	2	2.90E-03	1.40E-01
purine nucleobase biosynthetic process	3	1	3.00E-03	1.30E-01
glutathione transferase activity	5	1.7	3.10E-03	1.50E-01
retrograde vesicle-mediated transport, Golgi to ER	7	2.3	3.10E-03	1.40E-01
negative regulation of epidermal growth factor receptor signaling pathway	5	1.7	3.50E-03	1.50E-01
regulation of translational initiation	5	1.7	3.50E-03	1.50E-01
'de novo' IMP biosynthetic process	3	1	4.40E-03	1.80E-01
cAMP-dependent protein kinase regulator activity	3	1	6.00E-03	2.50E-01
modulation by virus of host process	3	1	6.10E-03	2.30E-01
viral process	13	4.3	6.60E-03	2.40E-01
cellular response to drug	6	2	7.20E-03	2.50E-01
formation of translation preinitiation complex	4	1.3	7.30E-03	2.50E-01
response to drug	13	4.3	7.50E-03	2.50E-01
positive regulation of actin filament polymerization	5	1.7	7.80E-03	2.60E-01
pre-mRNA intronic binding	3	1	7.90E-03	2.90E-01
protein disulfide oxidoreductase activity	4	1.3	8.10E-03	2.80E-01
aging	9	3	8.70E-03	2.70E-01
signal transducer activity	10	3.3	9.60E-03	3.10E-01
microtubule binding	10	3.3	1.10E-02	3.30E-01
cell redox homeostasis	6	2	1.10E-02	3.30E-01
identical protein binding	23	7.6	1.20E-02	3.40E-01
tetrahydrofolate interconversion	3	1	1.30E-02	3.60E-01
rRNA processing	10	3.3	1.30E-02	3.70E-01
NF-kappaB binding	4	1.3	1.50E-02	3.90E-01
ubiquitin-dependent protein catabolic process	9	3	1.50E-02	4.00E-01
embryonic viscerocranium morphogenesis	3	1	1.50E-02	3.90E-01
positive regulation of protein kinase B signaling	6	2	1.60E-02	4.00E-01
pyridoxal phosphate binding	5	1.7	1.70E-02	4.20E-01
ephrin receptor signaling pathway	6	2	1.80E-02	4.30E-01
proteasome assembly	3	1	1.80E-02	4.30E-01
positive regulation of substrate adhesion-dependent cell spreading	4	1.3	1.80E-02	4.30E-01
neuron apoptotic process	4	1.3	2.00E-02	4.50E-01
ER-associated ubiquitin-dependent protein catabolic process	5	1.7	2.10E-02	4.60E-01
protein folding in endoplasmic reticulum	3	1	2.10E-02	4.60E-01
ER to Golgi vesicle-mediated transport	8	2.6	2.30E-02	4.80E-01
protein kinase A catalytic subunit binding	3	1	2.40E-02	5.10E-01
protein phosphatase binding	5	1.7	2.40E-02	4.90E-01
regulation of protein catabolic process	3	1	2.40E-02	4.90E-01
kinase activity	10	3.3	2.60E-02	5.00E-01

actin filament binding	7	2.3	2.80E-02	5.20E-01
ubiquitin protein ligase binding	11	3.6	2.90E-02	5.20E-01
protein heterooligomerization	5	1.7	3.00E-02	5.60E-01
transferase activity, transferring acyl groups	4	1.3	3.00E-02	5.10E-01
rRNA binding	4	1.3	3.20E-02	5.30E-01
tRNA aminoacylation for protein translation	4	1.3	3.30E-02	5.90E-01
cellular oxidant detoxification	5	1.7	3.40E-02	6.00E-01
stress granule disassembly	2	0.7	3.50E-02	6.00E-01
response to magnetism	2	0.7	3.50E-02	6.00E-01
regulation of actin filament depolymerization	2	0.7	3.50E-02	6.00E-01
protein transporter activity	5	1.7	3.70E-02	5.60E-01
folic acid metabolic process	3	1	3.90E-02	6.30E-01
protein folding	8	2.6	3.90E-02	6.30E-01
response to nutrient	5	1.7	4.10E-02	6.40E-01
activation of MAPKK activity	4	1.3	4.60E-02	6.80E-01
enzyme regulator activity	3	1	4.70E-02	6.40E-01
actin cytoskeleton reorganization	4	1.3	4.90E-02	7.00E-01
organ regeneration	4	1.3	4.90E-02	7.00E-01
regulation of actin cytoskeleton organization	4	1.3	4.90E-02	7.00E-01

ALI Day 13

poly(A) RNA binding	24	23.3	1.60E-07	4.00E-05
epithelial cell differentiation	8	7.8	2.00E-07	1.40E-04
tRNA aminoacylation for protein translation	6	5.8	4.00E-06	1.40E-03
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	8	7.8	7.50E-06	1.80E-03
cadherin binding involved in cell-cell adhesion	11	10.7	9.30E-06	1.20E-03
positive regulation of protein localization to Cajal body	4	3.9	1.20E-05	2.00E-03
positive regulation of establishment of protein localization to telomere	4	3.9	1.70E-05	2.40E-03
cell-cell adhesion	10	9.7	3.70E-05	4.40E-03
positive regulation of telomerase RNA localization to Cajal body	4	3.9	9.10E-05	9.20E-03
protein binding	72	69.9	1.40E-04	1.10E-02
translation	9	8.7	1.40E-04	1.30E-02
glutathione peroxidase activity	4	3.9	2.60E-04	1.60E-02
SRP-dependent cotranslational protein targeting to membrane	6	5.8	2.60E-04	2.00E-02
canonical glycolysis	4	3.9	5.00E-04	3.40E-02
viral transcription	6	5.8	5.70E-04	3.60E-02
cellular oxidant detoxification	5	4.9	8.30E-04	4.80E-02
positive regulation of telomere maintenance via telomerase	4	3.9	9.20E-04	4.90E-02
toxin transport	4	3.9	1.30E-03	6.40E-02
translational initiation	6	5.8	1.40E-03	6.50E-02
rRNA processing	7	6.8	1.90E-03	7.90E-02
skeletal muscle tissue regeneration	3	2.9	3.60E-03	1.40E-01
unfolded protein binding	5	4.9	4.30E-03	1.90E-01
response to virus	5	4.9	4.40E-03	1.60E-01

protein folding	6	5.8	4.60E-03	1.60E-01
negative regulation of translation	4	3.9	5.10E-03	1.70E-01
translational elongation	3	2.9	5.10E-03	1.60E-01
oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	3	2.9	5.70E-03	2.10E-01
enzyme regulator activity	3	2.9	6.30E-03	2.00E-01
ATP binding	18	17.5	7.20E-03	2.00E-01
Arp2/3 complex-mediated actin nucleation	3	2.9	7.70E-03	2.20E-01
actin filament binding	5	4.9	8.10E-03	2.00E-01
cellular response to fatty acid	3	2.9	8.30E-03	2.30E-01
negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	4	3.9	9.00E-03	2.30E-01
protein stabilization	5	4.9	9.20E-03	2.30E-01
structural constituent of ribosome	6	5.8	1.10E-02	2.40E-01
positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	4	3.9	1.10E-02	2.60E-01
glyceraldehyde-3-phosphate biosynthetic process	2	1.9	1.20E-02	2.70E-01
anaphase-promoting complex-dependent catabolic process	4	3.9	1.20E-02	2.60E-01
positive regulation of protein kinase B signaling	4	3.9	1.40E-02	2.90E-01
ephrin receptor signaling pathway	4	3.9	1.50E-02	3.00E-01
RNA binding	9	8.7	1.70E-02	3.20E-01
translation elongation factor activity	3	2.9	1.80E-02	3.10E-01
glycolytic process	3	2.9	1.80E-02	3.30E-01
binding of sperm to zona pellucida	3	2.9	1.90E-02	3.40E-01
NAD binding	3	2.9	2.20E-02	3.40E-01
response to reactive oxygen species	3	2.9	2.30E-02	3.90E-01
stimulatory C-type lectin receptor signaling pathway	4	3.9	2.50E-02	4.10E-01
protein heterotetramerization	3	2.9	2.60E-02	4.20E-01
gluconeogenesis	3	2.9	2.90E-02	4.40E-01
phenylalanine-tRNA ligase activity	2	1.9	3.00E-02	4.20E-01
osteoclast development	2	1.9	3.00E-02	4.40E-01
phenylalanyl-tRNA aminoacylation	2	1.9	3.00E-02	4.40E-01
organ regeneration	3	2.9	3.20E-02	4.60E-01
cytoskeletal protein binding	3	2.9	3.30E-02	4.30E-01
response to organic cyclic compound	3	2.9	3.50E-02	4.80E-01
identical protein binding	10	9.7	3.50E-02	4.30E-01
aldehyde dehydrogenase [NAD(P)+] activity	2	1.9	3.50E-02	4.10E-01
positive regulation of canonical Wnt signaling pathway	4	3.9	3.60E-02	4.70E-01
tRNA binding	3	2.9	3.70E-02	4.10E-01
regulation of cellular amino acid metabolic process	3	2.9	3.80E-02	4.80E-01
mRNA binding	4	3.9	3.90E-02	4.10E-01
Fc-gamma receptor signaling pathway involved in phagocytosis	4	3.9	4.10E-02	5.10E-01
glutathione metabolic process	3	2.9	4.50E-02	5.30E-01
cellular response to interferon-gamma	3	2.9	4.60E-02	5.30E-01
RNA stem-loop binding	2	1.9	4.70E-02	4.50E-01
actin filament depolymerization	2	1.9	4.70E-02	5.30E-01

calcium-dependent phospholipid binding	3	2.9	4.70E-02	4.40E-01
calcium-dependent protein binding	3	2.9	4.70E-02	4.40E-01

ALI Day 18

regulation of cellular amino acid metabolic process	20	0.1	2.60E-26	2.50E-23
antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	19	0.1	1.50E-22	7.20E-20
NIK/NF-kappaB signaling	19	0.1	4.00E-22	1.20E-19
negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	19	0.1	1.70E-21	4.10E-19
positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	19	0.1	6.70E-21	1.30E-18
anaphase-promoting complex-dependent catabolic process	19	0.1	1.40E-20	2.20E-18
stimulatory C-type lectin receptor signaling pathway	20	0.1	1.30E-19	1.70E-17
Wnt signaling pathway, planar cell polarity pathway	19	0.1	2.70E-19	3.20E-17
positive regulation of canonical Wnt signaling pathway	20	0.1	1.80E-18	1.90E-16
regulation of mRNA stability	19	0.1	2.30E-18	2.20E-16
tumor necrosis factor-mediated signaling pathway	19	0.1	2.90E-17	2.50E-15
T cell receptor signaling pathway	20	0.1	1.00E-16	8.70E-15
proteasome-mediated ubiquitin-dependent protein catabolic process	22	0.1	2.00E-16	1.60E-14
protein polyubiquitination	21	0.1	4.20E-16	3.00E-14
Fc-epsilon receptor signaling pathway	20	0.1	3.50E-15	2.20E-13
negative regulation of canonical Wnt signaling pathway	19	0.1	1.10E-14	6.30E-13
MAPK cascade	20	0.1	4.00E-12	2.20E-10
proteasome-activating ATPase activity	6	0	3.50E-10	1.10E-07
positive regulation of RNA polymerase II transcriptional preinitiation complex assembly	6	0	7.70E-09	4.00E-07
ATPase activity	14	0	1.30E-08	2.10E-06
protein binding	115	0.4	1.60E-08	1.80E-06
proteasome assembly	6	0	4.70E-08	2.30E-06
ER-associated ubiquitin-dependent protein catabolic process	9	0	7.50E-08	3.50E-06
ATP binding	36	0.1	1.50E-07	1.20E-05
positive regulation of proteasomal protein catabolic process	6	0	2.50E-07	1.10E-05
TBP-class protein binding	6	0	1.10E-06	6.90E-05
protein catabolic process	7	0	1.90E-06	8.30E-05
poly(A) RNA binding	28	0.1	3.50E-06	1.90E-04
hydrolase activity	11	0	1.00E-05	4.70E-04
cadherin binding involved in cell-cell adhesion	13	0	1.40E-05	5.80E-04
cell-cell adhesion	12	0	4.40E-05	1.80E-03
ubiquitin-dependent protein catabolic process	10	0	4.90E-05	1.90E-03
mRNA splicing, via spliceosome	10	0	2.20E-04	8.30E-03
epithelial cell differentiation	6	0	4.70E-04	1.70E-02
identical protein binding	18	0.1	4.90E-04	1.80E-02
microtubule motor activity	6	0	8.20E-04	2.60E-02
response to hydrogen peroxide	5	0	1.20E-03	4.30E-02
positive regulation of telomere maintenance via telomerase	4	0	3.10E-03	1.00E-01
axonemal dynein complex assembly	3	0	3.60E-03	1.10E-01
GTP biosynthetic process	3	0	6.20E-03	1.80E-01

viral process	9	0	6.60E-03	1.80E-01
microtubule-based movement	5	0	6.70E-03	1.80E-01
outer dynein arm assembly	3	0	8.20E-03	2.10E-01
protein stabilization	6	0	8.60E-03	2.10E-01
RNA binding	12	0	1.20E-02	2.90E-01
sperm motility	4	0	1.40E-02	3.10E-01
enzyme regulator activity	3	0	1.40E-02	3.20E-01
response to ethanol	5	0	1.60E-02	3.50E-01
negative regulation of translation	4	0	1.60E-02	3.40E-01
stress granule disassembly	2	0	1.80E-02	3.70E-01
mitochondrial RNA 5'-end processing	2	0	1.80E-02	3.70E-01
RNA splicing	6	0	1.90E-02	3.70E-01
termination of RNA polymerase II transcription	4	0	2.10E-02	4.00E-01
spliceosomal complex assembly	3	0	2.40E-02	4.30E-01
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	5	0	2.40E-02	4.30E-01
mRNA processing	6	0	2.50E-02	4.30E-01
cilium movement	3	0	2.60E-02	4.30E-01
regulation of protein stability	4	0	2.70E-02	4.40E-01
cellular oxidant detoxification	4	0	2.70E-02	4.40E-01
spliceosomal snRNP assembly	3	0	2.70E-02	4.30E-01
proteasome regulatory particle assembly	2	0	2.70E-02	4.30E-01
positive regulation of telomerase activity	3	0	2.90E-02	4.40E-01
cell redox homeostasis	4	0	3.40E-02	4.90E-01
renal water homeostasis	3	0	3.50E-02	4.90E-01
formate-tetrahydrofolate ligase activity	2	0	3.60E-02	6.00E-01
methenyltetrahydrofolate cyclohydrolase activity	2	0	3.60E-02	6.00E-01
methylenetetrahydrofolate dehydrogenase (NADP+) activity	2	0	3.60E-02	6.00E-01
positive regulation of inclusion body assembly	2	0	3.60E-02	5.00E-01
translational initiation	5	0	3.80E-02	5.10E-01
ubiquitin protein ligase binding	7	0	4.80E-02	6.80E-01

[1]DAVID gene ontology analysis on proteins identified as positive hits in endogenous SPAG1 co-immunoprecipitations on ALI day 0, 13, and 18 in differentiating human bronchial epithelial cells, listed in Appendix 5.
[2]A functional annotation chart was created using the categories, GOTERM_BP_DIRECT and GOTERM_MF_DIRECT, with the background set to *Homo Sapiens*.

REFERENCES

Alapati, D., and Morrisey, E. E. (2017). Gene Editing and Genetic Lung Disease. Basic Research Meets Therapeutic Application. *Am. J. Respir. Cell Mol. Biol.* **56**, 283–290. https://doi.org/10.1165/rcmb.2016-0301PS

Antony, D., Brunner, H. G., and Schmidts, M. (2021). Ciliary Dyneins and Dynein Related Ciliopathies. *Cells* **10**, 1885. https://doi.org/10.3390/cells10081885

Austin-Tse, C., Halbritter, J., Zariwala, M. A., Gilberti, R. M., Gee, H. Y., Hellman, N., Pathak, N., Liu, Y., Panizzi, J. R., Patel-King, R. S. et al. (2013). Zebrafish ciliopathy screen plus human mutational analysis identifies C21orf59 and CCDC65 defects as causing primary ciliary dyskinesia. *Am. J. Hum. Genet.* **93**, 672–686. https://doi.org/10.1016/j.ajhg.2013.08.015

Avasthi, P., and Marshall, W. F. (2012). Stages of Ciliogenesis and Regulations of Ciliary Length. *Differrentiation* **83**, S30-S42. https://doi.org/10.1016/j.diff.2011.11.015

Basten, S. G., Davis, E. E., Gillis, A. J., van Rooijen, E., Stoop, H., Babala, N., Logister, I., Heath, Z. G., Jonges, T. N., Katsanis, N., Voest, E. E., van Eeden, F. J., Medema, R. H., Ketting, R. F., Schulte-Merker, S., Looijenga, L. H., and Giles, R. H. (2013). Mutations in LRRC50 predispose zebrafish and humans to seminomas. *PLoS Genetics* **9**, e1003384. https://doi.org/10.1371/journal.pgen.1003384

Benbahouche, N. E. H., Illiopoulos, I., Török, I., Marhold, J., Henri, J., Kajava, A. V., Farkaš, R., Kempf, T., Schnölzer, M., Meyer, P. et al. (2014). Drosophila Spag is the homolog of DNA polymerase II-associated protein 3 (RPAP3) and recruits the heat shock proteins 70 and 90 (Hsp70 and Hsp90) during the assembly of cellular machineries. *J. Biol. Chem*. **289**, 6236-6247. https://doi.org/10.1074/jbc.M113.499608

Biermann, K., Heukamp, L. C., Steger, K., Zhou, H., Franke, F. E., Sonnack, V., Brehm, R., Berg, J., Bastian, P. J. and Müller, S. C. (2007). Genome-wide expression profiling reveals new insights into pathogenesis and progression of testicular germ cell tumors. *Cancer Genomics and Proteomics* **4**, 359–367.

Bio-Rad. (2018). Droplet Digital™ PCR Droplet Digital™ PCR Applications Guide. Bulletin **6407**, 1-145.

Blackburn, K., Bustamante, X., Yin, W., Goshe, M. B., and Ostrowski, L. E. (2017). Quantitative proteomic analysis of human airway cilia identifies previously uncharacterized proteins of high abundance. *J. Proteome Res*. **16**, 1579 – 1592. https://doi.org/10.1021/acs.jproteome.6b00972

Bonnefoy, S., Watson, C. M., Kernohan, K. D., Lemos, M., Hutchinson, S., Poulter, J. A., Crinnion, L. A., Berry, I., Simmonds, J., Vasudevan, P., et al. (2018). Biallelic Mutations in LRRC56, Encoding a Protein Associated with Intraflagellar Transport, Cause Mucociliary Clearance and Laterality Defects. *Am. J. Hum. Genet.* **103**, 727–739. https://doi.org/10.1016/j.ajhg.2018.10.003

Braschi, B., Omran, H., Witman, G. B., Pazour, G. J., Pfister, K. K., Bruford, E. A., and King, S. M. (2022). Consensus nomenclature for dyneins and associated assembly factors. *J. Cell Biol.* **221**, e202109014. https://doi.org/10.1083/jcb.202109014

Brennan, S. K., Ferkol, T. W., and Davis, S. D. (2021). Emerging Genotype-Phenotype Relationships in Primary Ciliary Dyskinesia. *Int. J. Mol. Sci.* **22**, 8272. https://doi.org/10.3390/ijms22158272

Bustamante-Marin, X. M. and Ostrowski, L. E. (2017). Cilia and Mucociliary Clearance. *Cold Spring Harb. Perspect. Biol.* **9**, a028241. https://doi.org/10.1101/cshperspect.a028241

Bustamante-Marin, X. M., Yin, W., Sears, P. R., Werner, M. E., Brotslaw, E. J., Mitchell, B. J., Jania, C. M., Zeman, K. L., Rogers, T. D., Herring, L. E. et al. (2019). Lack of GAS2L2 causes PCD by impairing cilia orientation and mucociliary clearance. *Am. J. Hum. Genet.* **104**, 229-245. https://doi.org/10.1016/j.ajhg.2018.12.009

Bustamante-Marin, X. M., Horani, A., Stoyanova, M., Charng, W-L., Bottier, M., Sears, P. R., et al. (2020) Mutation of CFAP57, a protein required for the asymmetric targeting of a subset of inner dynein arms in Chlamydomonas, causes primary ciliary dyskinesia. *PLoS Genet.* **16**, e1008691. https://doi.org/10.1371/journal.pgen.1008691

Button, B., Cai, L. H., Ehre, C., Kesimer, M., Hill, D. B., Sheehan, J. K., Boucher, R. C., and Rubinstein, M. (2012). A periciliary brush promotes the lung health by separating the mucus layer from airway epithelia. *Science* **337**, 937–941. https://doi.org/10.1126/science.1223012

Canty, J. T., Tan, R., Kusakci, E., Fernandes, J., and Yildiz, A. (2021). Structure and Mechanics of Dynein Motors. *Annu. Rev. Biophys.* **50**, 549–574. https://doi.org/10.1146/annurev-biophys-111020-101511

Carson, J. L., Collier, A. M., Fernald, G. W., and Hu, S. C. (1994). Microtubular discontinuities as acquired ciliary defects in airway epithelium of patients with chronic respiratory diseases. *Ultrastruct. Pathol.* **18**, 327–332. https://doi.org/10.3109/01913129409023201

Carson, J. L., Hu, S. S., and Collier, A. M. (2000). Computer-assisted analysis of radial symmetry in human airway epithelial cilia: assessment of congenital ciliary defects in primary ciliary dyskinesia. *Ultrastruct. Pathol.* **24**, 169–174. https://doi.org/10.1080/01913120050132903

Chagot, M. E., Santos Morais, R. D., Dermouche, S., Lefebvre, D., Manival, X., Chipot, C., Dehez, F. and Quinternet, M. (2019). Binding properties of the quaternary assembly protein SPAG1. *Biochem. J.* **476**, 1679–1694. https://doi.org/10.1042/BCJ20190198

Cho, K. J., Noh, S. H., Han, S. M., Choi, W. I., Kim, H. Y., Yu, S., Lee, J. S., Rim, J. H., Lee, M. G., Hildebrandt, F., and Gee, H. Y. (2018). ZMYND10 stabilizes intermediate chain proteins in the cytoplasmic pre-assembly of dynein arms. *PLoS Genet.* 14, e1007316. https://doi.org/10.1371/journal.pgen.1007316

Cloutier, P., Poitras, C., Durand, M., Hekmat, O., Fiola-Masson, É., Bouchard, A., Faubert, D., Chabot, B., and Coulombe, B. (2017). R2TP/Prefoldin-like component RUVBL1/RUVBL2 directly interacts with ZNHIT2 to regulate assembly of U5 small nuclear ribonucleoprotein. *Nat. Commun.* **8**, 15615. https://doi.org/10.1038/ncomms15615

Dafinger, C., Rinschen, M. M., Borgal, L., Ehrenberg, C., Basten, S. G., Franke, M., Höhne, M., Rauh, M., Göbel, H., Bloch, W. et al. (2018). Targeted deletion of the AAA-ATPase Ruvbl1 in mice disrupts ciliary integrity and causes renal disease and hydrocephalus. *Exp. Mol. Med.* **50**, 1-17. https://doi.org/10.1038/s12276-018-0108-z

Davis, S. D., Margaret, R., Lee, H., Ferkol, T. W., Sagel, S. D., Dell, S. D., Milla, C., Pittman, J. E., Shapiro, A. J., Sullivan, K. M. et al. (2019). Primary Ciliary Dyskinesia: Longitudinal Study of Lung Disease by Ultrastructure Defect and Genotype. *Am. J. Respir. Crit. Care Med.* **199**, 190. https://doi.org/10.1164/rccm.201803-0548OC

De longh, R. U. and Rutland, J. (1995). Ciliary defects in healthy subjects, bronchiectasis, and primary ciliary dyskinesia. *Am. J. Respir. Crit. Care Med.***151**, 1559-1567. https://doi.org/10.1164/ajrccm.151.5.7735615

Dermouche, S., Chagot, M. E., Manival, X. and Quinternet, M. (2021). Optimizing the First TPR Domain of the Human SPAG1 Protein Provides Insight into the HSP70 and HSP90 Binding Properties. *Biochem* **60**, 2349-2363. https://doi.org/10.1021/acs.biochem.1c00052

Desai, P. B., Freshour, J. R., and Mitchell, D. R. (2015). Chlamydomonas axonemal dynein assembly locus ODA8 encodes a conserved flagellar protein needed for cytoplasmic maturation of outer dynein arm complexes. *Cytoskeleton (Hoboken)* **72**, 16–28. https://doi.org/10.1002/cm.21206

Desai, P. B., Dean, A. B. and Mitchell, D. R. (2017). Cytoplasmic preassembly and trafficking of axonemal dyneins. In *The Biology of Dynein Motors* (ed. S. King). Cambridge, MA: Elsevier, Inc.

Diggle, C. P., Moore, D. J., Mali, G., zur Lage, P., Ait-Lounis, A., Schmidts, M., Shoemark, A., Munoz, A. G., Halachev, M. R., Gautier, P. et al. (2014). HEATR2 Plays a Conserved Role in Assembly of the Ciliary Motile Apparatus. *PLoS Genet.* **10**, e1004577. https://doi.org/10.1371/journal.pgen.1004577

Dong, F., Shinohara, K., Botilde, Y., Nabeshima, R., Asai, Y., Fukumoto, A., Hasegawa, T., Matsuo, M., Takeda, H., Shiratori, H., Nakamura, T., and Hamada, H. (2014). Pih1d3 is required for cytoplasmic preassembly of axonemal dynein in mouse sperm. *J. Cell Biol.* **204**, 203–213. https://doi.org/10.1083/jcb.201304076

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. and Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137. https://doi.org/10.1093/hmg/ddu733

Duan, S., Huang, W., Liu, X., Liu, X., Chen, N., Xu, Q., Hu, Y., Song, W. and Zhou, J. (2018). IMPDH2 promotes colorectal cancer progression through activation of the PI3K/AKT/mTOR and PI3K/AKT/FOXO1 signaling pathways *J. Exp. Clin. Cancer Res.* **37**, 304. https://doi.org/10.1186/s13046-018-0980-3

Everman, J. L., Rios, C., and Seibold, M. A. (2018). Primary Airway Epithelial Cell Gene Editing Using CRISPR-Cas9. *Methods in Molecular Biology* (Clifton, N.J.), **1706**, 267–292. https://doi.org/10.1007/978-1-4939-7471-9_15

Fabczak, H. and Osinka, A. (2019). Role of the Novel Hsp90 Co-Chaperones in Dynein Arms' Preassembly. *Int. J. Mol. Sci.* **20**, 6174. https://doi.org/10.3390/ijms20246174

Fassad, M. R., Shoemark, A., le Borgne, P., Koll, F., Patel, M., Dixon, M., Hayward, J., Richardson, C., Frost, E., Jenkins, L., et al. (2018). C11orf70 Mutations Disrupting the Intraflagellar Transport-Dependent Assembly of Multiple Axonemal Dyneins Cause Primary Ciliary Dyskinesia. *Am. J. Hum. Genet.* **102**, 956–972. https://doi.org/10.1016/j.ajhg.2018.03.024

Fliegauf, M., Olbrich, H., Horvath, J., Wildhaber, J. H., Zariwala, M. A., Kennedy, M., Knowles, M. R. and Omran, H. (2005). Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. *Am. J. Respir. Crit. Care Med.* **171**, 1343–1349. https://doi.org/10.1164/rccm.200411-1583OC

Fliegauf, M., Benzing, T. and Omran, H. (2007). When cilia go bad: cilia defects and ciliopathies. *Nat. Rev. Mol. Cell. Biol.* **8**, 880–893. https://doi.org/10.1038/nrm2278

Fowkes, M. E. and Mitchell, D. R. (1998). The role of preassembled cytoplasmic complexes in assembly of flagellar dynein subunits. *Mol. Biol. Cell* **9**, 2337–2347. https://doi.org/10.1091/mbc.9.9.2337

Fulcher, M. L. and Randell, S. H. (2013). Human nasal and tracheo-bronchial respiratory epithelial cell culture. *Methods Mol. Biol.* **945**, 109-121. https://doi.org/10.1007/978-1-62703-125-7_8

Gu, Y., Chu, M. Q., Xu, Z. J., Yuan, Q., Zhang, T. J., Lin, J., and Zhou, J. D. (2022). Comprehensive analysis of SPAG1 expression as a prognostic and predictive biomarker in acute myeloid leukemia by integrative bioinformatics and clinical validation. *BMC Med. Genomics* **15**, 38. https://doi.org/10.1186/s12920-022-01193-0

Ha, S., Lindsay, A. M., Timms, A. E., and Beier, D. R. (2016). Mutations in Dnaaf1 and Lrrc48 Cause Hydrocephalus, Laterality Defects, and Sinusitis in Mice. *G3* (Bethesda, Md.) **6**, 2479–2487. https://doi.org/10.1534/g3.116.030791

Hannah, W. B., Seifert, B. A., Truty, R., Zariwala, M. A., Ameel, K., Zhao, Y., Nykamp, K., and Gaston, B. (2022). The global prevalence and ethnic heterogeneity of primary ciliary dyskinesia gene variants: a genetic database analysis. *Lancet Respir. Med.* **10**, 459-468. https://doi.org/10.1016/S2213-2600(21)00453-7

Hartill, V. L., van de Hoek, G., Patel, M. P., Little, R., Watson, C. M., Berry, I. R., Shoemark, A., Abdelmottaleb, D., Parkes, E., Bacchelli, C. et al. (2018). DNAAF1 links heart laterality with the AAA 1 ATPase RUVBL1 and ciliary intraflagellar transport. *Hum. Mol. Genet.* **27**, 529–545. https://doi.org/10.1093/hmg/ddx422

Heetvelde, M. V., Loocke, W. V., Trypsteen, W., Baert, A., Vanderheyden, K., Crombez, B., Vandesompele, J., Leeneer, K. D. and Claes, K. B. M. (2017). Evaluation of relative quantification of alternatively spliced transcripts using droplet digital PCR. *Biomol. Detect. Quantif.* **13**, 40-48. https://doi.org/10.1016/j.bdq.2017.09.001

Henri, J., Chagot, M. E., Bourguet, M., Abel, Y., Terral, G., Maurizy, C., Aigueperse, C., Georgescauld, F., Vandermoere, F., Saint-Fort, R., Behm-Ansmant, I., Charpentier, B., Pradet-Balade, B., Verheggen, C., Bertrand, E., Meyer, P., Cianférani, S., Manival, X., and Quinternet, M. (2018). Deep Structural Analysis of RPAP3 and PIH1D1, Two Components of the HSP90 Co-chaperone R2TP Complex. *Structure* **26**, 1196–1209.e8. https://doi.org/10.1016/j.str.2018.06.002

Höben, I. M., Hjeij, R., Olbrich, H., Dougherty, G. W., Nöthe-Menchen, T., Aprea, I., Frank, D., Pennekamp, P., Dworniczak, B., Wallmeier, J. et al. (2018). Mutations in C11orf70 Cause Primary Ciliary Dyskinesia with Randomization of Left/Right Body Asymmetry Due to Defects of Outer and Inner Dynein Arms. *Am. J. Hum. Genet.* **102**, 973–984. https://doi.org/10.1016/j.ajhg.2018.03.025

Hoque, M., Kim, E. N., Chen, D., Li, F., and Takemaru, K. (2022). Essential Roles of Efferent Duct Multicilia in Male Fertility. *Cells* **11**, 341. https://doi.org/10.3390/cells11030341

Horani, A., Druley, T. E., Zariwala, M. A., Patel, A. C., Levinson, B. T., Van Arendonk, L. G., Thornton, K. C., Giacalone, J. C., Albee, A. J., Wilson, K. S., et al. (2012). Whole-exome capture and sequencing identifies HEATR2 mutation as a cause of primary ciliary dyskinesia. *Am. J. Hum. Genet.* **91**, 685–693. https://doi.org/10.1016/j.ajhg.2012.08.022

Horani, A., Ferkol, T. W., Shoseyov, D., Wasserman, M. G., Oren, Y. S., Kerem, B., Amirav, I., Cohen-Cymberknoh, M., Dutcher, S. K., Brody, S. L., et al. (2013). LRRC6 mutation causes primary ciliary dyskinesia with dynein arm defects. *PLoS One* **8**, e59436. https://doi.org/10.1371/journal.pone.0059436

Horani, A., Ustione, A., Huang, T., Firth, A. L., Pan, J., Gunsten, S. P., Haspel, J. A., Piston, D. W. and Brody, S. L. (2018). Establishment of the early cilia preassembly protein complex during motile ciliogenesis. *Proc. Natl. Acad. Sci. USA* **115**, E1221-E1228. https://doi.org/10.1073/pnas.1715915115

Hořejší, Z., Stach, L., Flower, T. G., Joshi, D., Flynn, H., Skehel, J. M., O'Reilly, N. J., Ogrodowicz, R. W., Smerdon, S. J., and Boulton, S. J. (2014). Phosphorylation-dependent PIH1D1 interactions define substrate specificity of the R2TP cochaperone complex. *Cell Rep.* **7**, 19–26. https://doi.org/10.1016/j.celrep.2014.03.013

Houry, W. A., Bertrand, E., and Coulombe, B. (2018). The PAQosome, an R2TP-based chaperone for quaternary structure formation. *Trends Biochem. Sci*. **43**, 4–9. https://doi.org/10.1016/j.tibs.2017.11.001

Hu, P., Guan, K., Fen, Y., Ma, C., Song, H., Li, Y., Xia, X., Li, J. and Li, F. (2017). miR-638 Inhibits immature Sertoli cell growth by indirectly inactivating PI3K/AKT pathway via SPAG1 gene. *Cell Cycle* **16**, 2290–2300. https://doi.org/10.1080/15384101.2017.1380130

Huang, C., Wu, D., Khan, F. A., Jiao, X., Guan, K. and Huo, L. (2016). The GTPase SPAG-1 orchestrates meiotic program by dictating meiotic resumption and cytoskeleton architecture in mouse oocytes. *Mol. Biol. Cell* **27**, 1776–1785. https://doi.org/10.1091/mbc.E16-02-0132

Huizar, R. L., Lee, C., Boulgakov, A. A., Horani, A., Tu, F., Marcotte, E. M., Brody, S. L. and Wallingford, J. B. (2018). A liquid-like organelle at the root of motile ciliopathy. *Elife* **7**, e38497. https://doi.org/10.7554/eLife.38497

Inaba, Y., Shinohara, K., Botilde, Y., Nabeshima, R., Takaoka, K., Ajima, R., Lamri, L., Takeda, H., Saga, Y., Nakamura, T., and Hamada, H. (2016). Transport of the outer dynein arm complex to cilia requires a cytoplasmic protein Lrrc6. *Genes Cells* **21**, 728–739. https://doi.org/10.1111/gtc.12380

Ishikawa, T. (2017). Axoneme Structure form Motile Cilia. *Cold Spring Harb. Perspect. Biol.* **9**, a028076. https://doi.org/10.1101/cshperspect.a028076

Itsuki, Y., Saeki, M., Nakahara, H., Egusa, H., Irie, Y., Terao, Y., Kawabata, S., Yatani, H. and Kamisaki, Y. (2008). Molecular cloning of novel Monad binding protein containing tetratricopeptide repeat domains. *FEBS Lett.* **582**, 2365–2370. https://doi.org/10.1016/j.febslet.2008.05.041

Jain, R., Javidan-Nejad, C., Alexander-Brett, J., Horani, A., Cabellon, M. C., Walter, M. J., and Brody, S. L. (2012). Sensory functions of motile cilia and implication for bronchiectasis. *Front. Biosci.* **4**, 1088-1098. https://doi.org/10.2741/s320

Kamiya, R. (2002). Functional diversity of axonemal dyneins as studied in Chlamydomonas mutants. *Int. Rev. Cytol.* **219**, 115-155. https://doi.org/10.1016/s0074-7696(02)19012-7

Kartagener, M. (1933). Zur pathogenese der bronkiectasien. Bronkiectasien bei situs viscerum inversus. *Beitr Klin Tuberk* **82**, 489–501.

Kasioulis, I., Das, R. M. and Storey, K. G. (2017). Inter-dependent apical microtubule and actin dynamics orchestrate centrosome retention and neuronal delamination. *Elife* **6**, e26215. https://doi.org/10.7554/eLife.26215

King, S. M. (2016). Axonemal dynein arms. *Cold Spring Harb. Perspect. Biol.* **8**, a028100. https://doi.org/10.1101/cshperspect.a028100

King, S. M. (2021). Cytoplasmic factories for axonemal dynein assembly. *J. Cell Sci.* **134**, jcs258626. https://doi.org/10.1242/jcs.258626

Knowles, M. R., Ostrowski, L. E., Loges, N. T., Hurd, T., Leigh, M. W., Huang, L, Wolf, W. E., Carson, J. L., Hazucha, M. J., Yin, W. et al. (2013). Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms. *Am. J. Hum. Genet.* **93**, 711–720. https://doi.org/10.1016/j.ajhg.2013.07.025

Kobayashi, D., Asano-Hoshino, A., Nakakura, T., Nishimaki, T., Ansai, S., Kinoshita, M., Ogawa, M., Hagiwara, H., and Yokoyama, T. (2017). Loss of zinc finger MYND-type containing 10 (zmynd10) affects cilia integrity and axonemal localization of dynein arms, resulting in ciliary dysmotility, polycystic kidney and scoliosis in medaka (Oryzias latipes). *Dev. Biol.* **430**, 69–79. https://doi.org/10.1016/j.ydbio.2017.08.016

Kollmar, M. (2016). Fine-tuning motile cilia and flagella: Evolution of the dynein motor proteins from plants to humans at high resolution. *Mol. Biol. Evol.* **33**, 3249–3267. https://doi.org/10.1093/molbev/msw213

Kott, E., Duquesnoy, P., Copin, B., Legendre, M., Moal, F. D., Montantin, G., Jeanson, L., Tamalet, A., Papon, J. F., Siffroi, J. P. et al. (2012). Loss-of-function mutations in LRRC6, a gene essential for proper axonemal assembly of inner and outer dynein arms, cause primary ciliary dyskinesia. *Am. J. Hum. Genet.* **91**, 958–964. https://doi.org/10.1016/j.ajhg.2012.10.003

Krasteva, G., Canning, B. J., Hartmann, P., and Kummer, W. (2011). Cholinergic chemosensory cells in the trachea regulate breathing. *Proc. Natl. Acad. Sci.* **108**, 9478-9483. https://doi.org/10.1073/pnas.1019418108

Lee, C., Cox, R. M., Papoulas, O., Horani, A., Drew, K., Devitt, C. C., Brody, S. L., Marcotte, E. M. and Wallingford, J. B. (2020). Functional partitioning of a liquid-like organelle during assembly of axonemal dyneins. *Elife* **9**, e58662. https://doi.org/10.7554/eLife.58662

Lee, Y. T., Jacob, J., Michowski, W., Nowotny, M., Kuznicki, J., and Chazin, W. J. (2004). Human Sgt1 binds HSP90 through the CHORD-Sgt1 domain and not the tetratricopeptide repeat domain. *J. Bio. Chem.* **279**, 16511–16517. https://doi.org/10.1074/jbc.M400215200

Legendre, M., Zaragosi, L. E., and Mitchison, H. M. (2021). Motile cilia and airway disease. *Semin. Cell Dev. Biol.* **110**, 19-33. https://doi.org/10.1016/j.semcdb.2020.11.007

Leigh, M. W., Hazucha, M. J., Chawla, K. K., Baker, B. R., Shapiro, A. J., Brown, D. E., LaVange, L. M., Horton, B. J., Qaqish, B., Carson, J. L. et al. (2013). Standardizing nasal nitric oxide measurement as a test for primary ciliary dyskinesia. *Ann. Am. Thorac. Soc.* **10**, 574-581. https://doi.org/10.1513/AnnalsATS.201305-110OC

Leigh, M. W., Horani, A., Kinghorn, B., O'Connor, M. G., Zariwala, M. A. and Knowles, M. R. (2019). Primary ciliary dyskinesia (PCD): A genetic disorder of motile cilia. *Transl. Sci. Rare Dis.* **4**, 51-75. https://doi.org/10.3233/TRD-190036

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [Preprint]. arXiv:1303.3997

Li, Y., Zhao, L., Yuan, S., Zhang, J. and Sun, Z. (2017). Axonemal dynein assembly requires the R2TP complex component Pontin. *Development* **144**, 4684–4693. https://doi.org/10.1242/dev.152314

Lin, S., Lv, Y., Zheng, L., Mao, G. and Peng, F. (2021). Expression and Prognosis of Sperm-Associated Antigen 1 in Human Breast Cancer. *Onco. Targets. Ther.* **14**, 2689–2698. https://doi.org/10.2147/OTT.S288484

Lin, W., Zhou, X., Zhang, M., Li, Y., Miao, S., Wang, L., Zong, S. and Koide, S. S. (2001). Expression and function of the HSD-3.8 gene encoding a testis-specific protein. *Mol. Hum. Reprod.* **7**, 811–818. https://doi.org/10.1093/molehr/7.9.811

Little, R. B., and Norris, D. P. (2021). Right, left and cilia: How asymmetry is established. *Semin. Cell Dev. Biol.* **110**, 11-18. https://doi.org/10.1016/j.semcdb.2020.06.003

Liu, G., Wang, L., Pan, J. and Yao, X. (2019). Chlamydomonas WDR92 in association with R2TP-like complex and multiple DNAAFs to regulate ciliary dynein preassembly. *J. Mol. Cell Biol.* **11**, 770–780. https://doi.org/10.1093/jmcb/mjy067

Liu, N., Qiao, Y., Cai, C., Lin, W., Zhang, J., Miao, S., Zong, S., Koide, S. S., and Wang, L. (2006). A sperm component, HSD-3.8 (SPAG1), interacts with G-protein beta 1 subunit and activates extracellular signal-regulated kinases (ERK). *Front. Biosci.* **11**, 1679–1689. https://doi.org/10.2741/1913

Lodes, N., Seidensticker, K., Perniss, A., Nietzer, S., Oberwinkler, H., May, T., Walles, T., Hebestreit, H., Hackenberg, S., and Steinke, M. (2020). Investigation on Ciliary Functionality of Different Airway Epithelial Cell Lines in Three-Dimensional Cell Culture. *Tissue Eng. Part A* **26**, 432–440. https://doi.org/10.1089/ten.TEA.2019.0188

Loges, N. T., Olbrich, H., Becker-Heck, A., Häffner, K., Heer, A., Reinhard, C., Schmidts, M., Kispert, A., Zariwala, M. A., Leigh, M. W. et al. (2009). Deletions and Point Mutations of LRRC50 Cause Primary Ciliary Dyskinesia Due to Dynein Arm Defects. *Am. J. Hum. Genet.* **85**, 883–889. https://doi.org/10.1016/j.ajhg.2009.10.018

Loreng, T. D. and Smith, E. F. (2017). The Central Apparatus of Cilia and Eukaryotic Flagella. *Cold Spring Harb. Perspect. Biol.* **9**, a028118. https://doi.org/10.1101/cshperspect.a028118

Lynham, J., and Houry, W. A. (2018). The Multiple Functions of the PAQosome: An R2TP- and URI1 Prefoldin-Based Chaperone Complex. *Adv. Exp. Med. Biol*. **1106**, 37-72. https://doi.org/10.1007/978-3-030-00737-9_4

Lyons, R. A., Saidogan, E., and Djahanbackhch, O. (2006). The reproductive significance of human fallopian tube cilia. *Hum. Reprod. Update* **12**, 363-372. https://doi.org/10.1093/humupd/dml012

Maier, B., Kirsch, M., Anderhub, S., Zentgraf, H. and Krämer, A. (2013). The novel actin/focal adhesion-associated protein MISP is involved in mitotic spindle positioning in human cells. *Cell Cycle* **12**, 1457–1471. https://doi.org/10.4161/cc.24602

Mali, G. R., Yeyati, P. L., Mizuno, S., Dodd, D. O., Tennant, P. A., Keighren, M. A., Zur Lage, P., Shoemark, A., Garcia-Munoz, A., Shimada, A., et al. (2018). ZMYND10 functions in a chaperone relay during axonemal dynein assembly. *eLife* **7**, e34389. https://doi.org/10.7554/eLife.34389

Martino, F., Pal, M., Muñoz-Hernández, H., Rodríguez, C. F., Núñez-Ramírez, R., Gil-Carton, D., Degliesposti, G., Skehel, J. M., Roe, S. M., Prodromou, C. et al. (2018). RPAP3 provides a flexible scaffold for coupling HSP90 to the human R2TP co-chaperone complex. *Nat. Commun.* **9**, 1501. https://doi.org/10.1038/s41467-018-03942-1

Maurizy, C., Quinternet, M., Abel, Y., Verheggen, C., Santo, P. E., Bourguet, M., Paiva, A. C. F., Bragantini, B., Chagot, M. E., Robert, M. C. et al. (2018). The RPAP3-Cterminal domain identifies R2TP-like quaternary chaperones. *Nat. Commun.* **9**, 2093. https://doi.org/10.1038/s41467-018-04431-1

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. https://doi.org/10.1101/gr.107524.110

Meunier, A. and Azimzadeh, J. (2016). Multiciliated Cells in Animals. *Cold Spring Harb. Perspect. Biol.* **8**, a028233. https://doi.org/10.1101/cshperspect.a028233

Miao, C., Jiang, Q., Li, H., Zhang, Q., Bai, B., Bao, Y., and Zhang, T. (2016). Mutations in the Motile Cilia Gene DNAAF1 Are Associated with Neural Tube Defects in Humans. *G3 (Bethesda, Md.)* **6**, 3307–3316. https://doi.org/10.1534/g3.116.033696

Mitchison, H. M., Schmidts, M., Loges, N. T., Freshour, J., Dritsoula, A., Hirst, R. A., O'Callaghan, C., Blau, H., Dabbagh, M. A., Olbrich, H. et al. (2012). Mutations in axonemal dynein assembly factor DNAAF3 cause primary ciliary dyskinesia. *Nat. Genet.* **44**, 381–389. https://doi.org/10.1038/ng.1106

Mitchison, H.M. and Valente, E.M. (2017). Motile and non-motile cilia in human pathology: from function to phenotypes. *J. Pathol.* **241**, 294-309. https://doi.org/10.1002/path.4843

Moore, D. J., Onoufriadis, A., Shoemark, A., Simpson, M. A., zur Lage, P. I., de Castro, S. C., Bartoloni, L., Gallone, G., Petridi, S., Woollard, W. J. et al. (2013). Mutations in ZMYND10, a gene essential for proper axonemal assembly of inner and outer dynein arms in humans and flies, cause primary ciliary dyskinesia. *Am. J. Hum. Genet.* **93**, 346–356. https://doi.org/10.1016/j.ajhg.2013.07.009

Müller, L., Brighton, L. E., Carson, J. L., Fischer, W. A., and Jaspers, I. (2013). Culturing of human nasal epithelial cells at the air liquid interface. *J. Vis. Exp.* **80**, 50646. https://doi.org/10.3791/50646

Neesse, A., Gangeswaran, R., Luettges, J., Feakins, R., Weeks, M. E., Lemoine, N. R., and Crnogorac-Jurcevic, T. (2007). Sperm-associated antigen 1 is expressed early in pancreatic tumorigenesis and promotes motility of cancer cells. *Oncogene* **26**, 1533–1545. https://doi.org/10.1038/sj.onc.1209961

Nishikawa, T., Ota, T., and Isogai, T. (2000). Prediction whether a human cDNA sequence contains initiation codon by combining statistical information and similarity with protein sequences. *Bioinformatics* **16**, 960-967. https://doi.org/10.1093/bioinformatics/16.11.960

Niziolek, M., Bicka, M., Osinka, A., Samsel, Z., Sekretarska, J., Poprzeczko, M., Bazan, R., Fabczak, H., Joachimiak, E., and Wloga, D. (2022). PCD Genes-From Patients to Model Organisms and Back to Humans. *Int. J. Mol. Sci.* **23**, 1749. https://doi.org/10.3390/ijms23031749

Olcese, C., Patel, M. P., Shoemark, A., Kiviluoto, S., Legendre, M., Williams, H. J., Vaughan, C. K., Hayward, J., Goldenberg, A., Emes, R. D. et al. (2017). X-linked primary ciliary dyskinesia due to mutations in the cytoplasmic axonemal dynein assembly factor PIH1D3. *Nat. Commun.* **8**, 14279. https://doi.org/10.1038/ncomms14279

Olin, J. T., Burns, K., Carson, J. L., Metjian, H., Atkinson, J. J., Davis, S. D., Dell, S. D., Ferkol, T. W., Milla, C. E., Olivier, K. N., Rosenfeld, M., Baker, B., Leigh, M. W., Knowles, M. R., Sagel, S. D., and Genetic Disorders of Mucociliary Clearance Consortium. (2011). Diagnostic yield of nasal scrape biopsies in primary ciliary dyskinesia: a multicenter experience. *Pediatr. Pulmonol.* **46**, 483–488. https://doi.org/10.1002/ppul.21402

Omran, H., Kobayashi, D., Olbrich, H., Tsukahara, T., Loges, N. T., Hagiwara, H., Zhang, Q., Leblond, G., O'Toole, E., Hara, C. et al. (2008). Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. *Nature* **456**, 611–616. https://doi.org/10.1038/nature07471

Osinka, A., Poprzeczko, M., Zielinska, M. M., Fabczak, H., Joachimiak, E., and Wloga, D. (2019). Ciliary Proteins: Filling the Gaps. Recent Advances in Deciphering the Protein Composition of Motile Ciliary Complexes. *Cells* **8**, 730. https://doi.org/10.3390/cells8070730

Ostrowski, L. E., Blackburn, K., Radde, K. M., Moyer, M. B., Schlatzer, D. M., Moseley, A. and Boucher, R. C. (2002). A proteomic analysis of human cilia: identification of novel components. *Mol. Cell Proteomics* **1**, 451-465. https://doi.org/10.1074/mcp.m200037-mcp200

Ostrowski, L. E., Yin, W., Rogers, T. D., Busalacchi, K. B., Chua, M., O'Neal, W. K. and Grubb, B. R. (2010). Conditional Deletion of *Dnaic1* in a Murine Model of Primary Ciliary Dyskinesia Causes Chronic Rhinosinusitis. *Am. J. Respir. Cell Mol. Biol.* **43**, 55 – 63. https://doi.org/10.1165/rcmb.2009-0118OC

Panizzi, J. R., Becker-Heck, A., Castleman, V. H., Al-Mutairi, D. A., Liu, Y., Loges, N. T., Pathak, N., Austin-Tse, C., Sheridan, E., Schmidts, M., Olbrich, H., et al. (2012). CCDC103 mutations cause primary ciliary dyskinesia by disrupting assembly of ciliary dynein arms. *Nat. Genet.* **44**, 714-719. https://doi.org/10.1038/ng.2277

Patel-King, R. S., and King, S. M. (2016). A prefoldin-associated WD-repeat protein (WDR92) is required for the correct architectural assembly of motile cilia. *Mol. Biol. Cell* **27**, 1204–1209. https://doi.org/10.1091/mbc.E16-01-0040

Patel-King, R. S., Sakato-antoku, M., Yankova, M. and King, S. M. (2019). WDR92 Is Required for Axonemal Dynein Heavy Chain Stability in Cytoplasm. *Mol. Biol. Cell* **30**, 1834-1845. https://doi.org/10.1091/mbc.E19-03-0139

Pazour, G. J., Agrin, N., Walker, B. L. and Witman, G. B. (2006). Identification of predicted human outer dynein arm genes: Candidates for primary ciliary dyskinesia genes. *J. Med. Genet.* **43**, 62–73. https://doi.org/10.1136/jmg.2005.033001

Perez-Riba, A. and Itzhaki, L. S. (2019). The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. *Curr. Opin. Struct. Biol.* **54**, 43–49. https://doi.org/10.1016/j.sbi.2018.12.004

Pfister, K. K. and Witman, G. B. (1984). Subfractionation of Chlamydomonas 18 S dynein into two unique subunits containing ATPase activity. *J. Biol. Chem.* **259**, 12072–12080.

Plasschaert, L. W., Zilionis, R., Choo-Wing, R., Savova, V., Knehr, J., Roma, G., Klein, A. M., and Jaffe, A. B. (2018). A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377-381. https://doi.org/10.1038/s41586-018-0394-6

Rayner, R.E., Makena, P., Prasad, G.L., and Cormet-Boyaka, E. (2019). Optimization of Normal Human Bronchial Epithelial (NHBE) Cell 3D Cultures for in vitro Lung Model Studies. *Sci. Rep.* **9**, 500. https://doi.org/10.1038/s41598-018-36735-z

Ringers, C., Olstad, E. W., and Jurisch-Yaksi, N. (2020). The role of motile cilia in the development and physiology of the nervous system. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **375**, 20190156. https://doi.org/10.1098/rstb.2019.0156

Sears, P. R., Yin, W. and Ostrowski, L. E. (2015). Continuous mucociliary transport by primary human airway epithelial cells in vitro. *Am. J. Physiol. Lung Cell Mol. Physiol.* **309**, L99-L108. https://doi.org/10.1152/ajplung.00024.2015

Shamsara, E., and Shamsara, J. (2020). Bioinformatics analysis of the genes involved in the extension of prostate cancer to adjacent lymph nodes by supervised and unsupervised machine learning methods: The role of SPAG1 and PLEKHF2. *Genomics* **112**, 3871–3882. https://doi.org/10.1016/j.ygeno.2020.06.035

Shapiro, A. J. and Leigh, M. W. (2017). Value of transmission electron microscopy for primary ciliary dyskinesia diagnosis in the era of molecular medicine: Genetic defects with normal and non-diagnositc ciliary ultrastructure. *Ultrastruct. Pathol.* **41**, 373-385. https://doi.org/10.1080/01913123.2017.1362088

Shapiro, A. J., Zariwala, M. A., Ferkol, T., Davis, S. D., Sagel, S. D., Dell, S. D., Rosenfeld, M., Olivier, K. N., Milla, C., Daniel, S. J. et al. (2016). Diagnosis, monitoring, and treatment of primary ciliary dyskinesia: PCD foundation consensus recommendations based on state of the art review. *Pediatr. Pulmonol.* **51**, 115–132. https://doi.org/10.1002/ppul.23304

Siewert, A. K. (1904). Uber einem Fall von Bronchiectasie bei einem Patienten mit Situs inversus viscerum. *Berliner klinische Wochenschrift* **41**, 139–141.

Siliņa, K., Zayakin, P., Kalnina, Z., Ivanova, L., Meistere, I., Endzelins, E., Abols, A., Stengrēvics, A., Leja, M., Ducena, K. et al. (2011). Sperm-associated antigens as targets for cancer immunotherapy: Expression pattern and humoral immune response in cancer patients. *J. Immunother.* **34**, 28–44. https://doi.org/10.1097/CJI.0b013e3181fb64fa

Smith, T. C., Fang, Z. and Luna, E. J. (2010). Novel interactors and a role for Supervillin in early cytokinesis. *Cytoskeleton* **67**, 346–364. https://doi.org/10.1002/cm.20449

Spassky, N. and Meunier, A. (2017). The development and functions of multiciliated epithelia. *Nat. Rev. Mol. Cell Biol.*, **18**, 423–436. https://doi.org/10.1038/nrm.2017.21

Sun, C., Shou, P., Du, H., Hirabayashi, K., Chen, Y., Herring, L. E., Ahn, S., Xu, Y., Suzuki, K., Li, G. et al. (2020). THEMIS-SHP1 recruitment by 4-1BB tunes LCK-mediated priming of chimeric antigen receptor-redirected T cells. *Cancer Cell* **37**, 216-225. https://doi.org/10.1016/j.ccell.2019.12.014

Suprynowicz, F. A., Upadhyay, G., Krawczyk, E., Kramer, S.C., Hebert, J.D., Liu, X., Yuan, H., Cheluvaraju, C., Clapp, P.W., Boucher, R.C. et al. (2012). Conditionally reprogrammed cells represent a stem-like state of adult epithelial cells. *Proc. Natl. Acad. Sci. USA* **109**, 20035-20040. https://doi.org/10.1073/pnas.1213241109

Takaishi, M. and Huh, N. H. (1999). A tetratricopeptide repeat-containing protein gene, tpis, whose expression is induced with differentiation of spermatogenic cells. *Biochem. Biophys. Res. Commun.* **264**, 81–85. https://doi.org/10.1006/bbrc.1999.1477

Tang, W. J. Y., Bell, C. W., Sale, W. S. and Gibbons, I. R. (1982). Structure of the dynein-1 outer arm in sea urchin sperm flagella. I. Analysis by separation of subunits. *J. Biol. Chem.* **257**, 508-515.

Tarkar, A., Loges, N. T., Slagle, C. E., Francis, R., Dougherty, G. W., Tamayo, J. V., Shook, B., Cantino, M., Schwartz, D., Jahnke, C. et al. (2013). DYX1C1 is required for axonemal dynein assembly and ciliary motility. *Nat. Genet.* **45**, 995–1003. https://doi.org/10.1038/ng.2707

Thomas, L., Bouhouche, K., Whitfield, M., Thouvenin, G., Coste, A., Louis, B., Szymanski, C., Bequignon, E., Papon, J. F., Castelli, M., et al. (2020). TTC12 Loss-of-Function Mutations Cause Primary Ciliary Dyskinesia and Unveil Distinct Dynein Assembly Mechanisms in Motile Cilia Versus Flagella. *Am. J. Hum. Genet.* **106**, 153–169. https://doi.org/10.1016/j.ajhg.2019.12.010

Thomas, P. and Smart, T. G. (2005). HEK293 cell line: a vehicle for the expression of recombinant proteins. *J. Pharmacol. Toxicol. Methods* **51**, 187-200. https://doi.org/10.1016/j.vascn.2004.08.014

van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Angel, G. D., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J. et al. (2013). From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Vertii, A., Hung, H. F., Hehnly, H. and Doxsey, S. (2016). Human basal body basics. *Cilia* **5**, 13. https://doi.org/10.1186/s13630-016-0030-8

von Morgen, P., Hořejší, Z., and Macurek, L. (2015). Substrate recognition and function of the R2TP complex in response to cellular stress. *Front. Genet.* **6**, 69. https://doi.org/10.3389/fgene.2015.00069

Wallmeier, J., Nielsen, K. G., Kuehni, C. E., Lucas, J. S., Leigh, M. W., Zariwala, M. A. and Omran, H. (2020). Motile ciliopathies. *Nat. Rev. Dis. Prim.* **6**, 1–29. https://doi.org/10.1038/s41572-020-0209-6

Wang, K., Li, M. and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. **38,** e164. https://doi.org/10.1093/nar/gkq603

Wang, Y., Liu, Z., Yang, G., Gao, Q., Xiao, L., Li, J., Guo, C., Troutwine, B. R., Gray, R. S., Xie, L., and Zhang, H. (2020). Coding Variants Coupled With Rapid Modeling in Zebrafish Implicate Dynein Genes, dnaaf1 and zmynd10, as Adolescent Idiopathic Scoliosis Candidate Genes. *Front. Cell Dev. Biol.* **8**, 582255. https://doi.org/10.3389/fcell.2020.582255

Whitfield, M., Thomas, L., Bequignon, E., Schmitt, A., Stouvenel, L., Montantin, G., Tissier, S., Duquesnoy, P., Copin, B., Chantot, S., et al. (2019). Mutations in DNAH17, Encoding a Sperm-Specific Axonemal Outer Dynein Arm Heavy Chain, Cause Isolated Male Infertility Due to Asthenozoospermia. *Am. J. Hum. Genet.* **105**, 198–212. https://doi.org/10.1016/j.ajhg.2019.04.015

Xu, J., Yu, H., and Sun, H. (2020). Less is More: Rare Pulmonary Neuroendocrine Cells Function as Critical Sensors in Lung. *Dev. Cell* **55**, 123-132. https://doi.org/10.1016/j.devcel.2020.09.024

Yamaguchi, H., Oda, T., Kikkawa, M. and Takeda, H. (2018). Systematic studies of all PIH proteins in zebrafish reveal their distinct roles in axonemal dynein assembly. *Elife* **7**, e36979. https://doi.org/10.7554/eLife.36979

Yamamoto, R., Hirono, M., and Kamiya, R. (2010). Discrete PIH proteins function in the cytoplasmic preassembly of different subsets of axonemal dyneins. *J. Cell Biol.* **190**, 65–71. https://doi.org/10.1371/journal.pgen.1009126

Yamamoto, R., Yanagi, S., Nagao, M., Yamasaki, Y., Tanaka, Y., Sale, W. S., Yagi, T., and Kon, T. (2020). Mutations in PIH proteins MOT48, TWI1, PF13 define common and unique steps for preassembly of each, different ciliary dynein. *PLoS Genet.* **16**, e1009126. https://doi.org/10.1371%2Fjournal.pgen.1009126

Zariwala, M. A., Knowles, M.R. and Leigh, M. W. (2007 [Updated 2019]). Primary Ciliary Dyskinesia. In: *GeneReviews*. University of Washington.

Zariwala, M. A., Gee, H. Y., Kurkowiak, M., Al-Mutairi, D. A., Leigh, M. W., Hurd, T. W., Hjeij, R., Dell, S. D., Chaki, M., Dougherty, G. W., et al. (2013). ZMYND10 is mutated in primary ciliary dyskinesia and interacts with LRRC6. *Am. J. Hum. Genet.* **93**, 336–345. https://doi.org/10.1016/j.ajhg.2013.06.007

Zhang, M. L., Wang, L. F., Miao, S. Y., and Koide, S. S. (1992). Isolation and Sequencing of the cDNA Encoding the 75-kD Human Sperm Protein Related to Infertility. *Chin. Med. J.* **105**, 998-1003.

Zhao, L., Yuan, S., Cao, Y., Kallakuri, S., Li, Y., Kishimoto, N., DiBella, L. and Sun, Z. (2013). Reptin/Ruvbl2 is a Lrrc6/Seahorse interactor essential for cilia motility. *Proc. Natl. Acad. Sci. USA.* **110**, 12697–12702. https://doi.org/10.1073/pnas.1300968110

Zhu, X., Liu, Y., and Yang, P. (2017). Radial Spokes-A Snapshot of the Motility Regulation, Assembly, and Evolution of Cilia and Flagella. *Cold Spring Harb. Perspect. Biol. 9*, a028126. https://doi.org/10.1101/cshperspect.a028126

zur Lage, P., Stefanopoulou, P., Styczynska-Soczka, K., Quinn, N., Mali, G., von Kriegsheim, A., Mill, P. and Jarman, A. P. (2018). Ciliary dynein motor preassembly is regulated by Wdr92 in association with HSP90 co-chaperone, R2TP. *J. Cell Biol.* **217**, 2583–2598. https://doi.org/10.1083/jcb.201709026

zur Lage, P., Xi, Z., Lennon, J., Hunter, I., Chan, W. K., Bolado Carrancio, A., von Kriegsheim, A., and Jarman, A. P. (2021). The Drosophila orthologue of the primary ciliary dyskinesia-associated gene, DNAAF3, is required for axonemal dynein assembly. *Biol. Open* **10**, bio058812. https://doi.org/10.1242/bio.058812