

Functional and expression analysis of Androglobin in different branches of the metazoan tree

Dissertation

Zur Erlangung des Grades
„Doktor der Naturwissenschaften“

am Fachbereich Biologie
der Johannes-Gutenberg-Universität Mainz

Carina Osterhof

geb. am 30.07.1992 in Dülmen

Mainz, den 22.09.2022

Dean: Prof. Dr. Eckhard Thines

First examiner: [REDACTED]

Second examiner: [REDACTED]

Date of defense: 09.11.2022

„Es ist ein Gedankengebäude“, seufzte Kolibril. „Meine ewige Baustelle. Halbgare Theorien, Ideenruinen. Ich bezweifle, dass ich mit dieser Doktorarbeit zu Lebzeiten jemals zu Rande kommen werde“.

aus Walter Moers: Rumo und die Wunder im Dunkeln

Table of content

Table of figures.....	II
Table of abbreviations.....	III
1 Introduction	1
1.1 Globins: a diverse gene family	1
1.2 Globin phylogeny: ancestral and vertebrate-specific lineages	4
1.3 Functions of conventional globins in health and disease	9
1.4 Androglobin – an ancient globin with a peculiar structure	12
1.5 Aim of this study	16
2 Results.....	19
2.1 Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis	19
2.2 Androglobin, a chimeric mammalian globin, is required for male fertility.....	45
2.3 A role of Androglobin in cancer? – insights from an overexpression system and transcriptome data mining	73
2.4 Androglobin expression pattern in basal metazoans confirms its conserved functional association with cilia	113
3 Discussion and Outlook	145
3.1 Androglobin as part of the ciliary proteome	147
3.2 Loss of Androglobin is an indicator of ciliary or flagellar specialization in metazoans	149
3.3 Adgb beyond metazoans.....	151
3.4 Androglobin: the maverick of the protein family?	154
4 Summary	159
5 Zusammenfassung	161
6 References	163
7 Appendix	177
7.1 Supplementary data.....	177
7.2 Index of electronic supplement	178
7.3 Acknowledgements.....	179
7.4 Eidesstattliche Erklärung.....	180
7.5 Curriculum vitae.....	181

Table of figures

Figure 1-1: Structural models of the globin fold with its heme group (black).	2
Figure 1-2: Phylogeny of vertebrate globins.....	4
Figure 1-3: Model of the evolution of GbX in vertebrate species.	6
Figure 1-4: Occurrence of globins among vertebrates.	8
Figure 1-5: Structure of Androglobin.....	13
Figure 1-6: mRNA expression analysis of Adgb in mice.	14
Figure 1-7: Phylogenetic tree of Androglobin homologues	15
Figure 3-1: Occurrence of putative Adgb sequences across the eucaryotic tree.....	153
Figure 3-2: Expression data on Sept10 and Adgb from the Human protein atlas.....	157

Table of abbreviations

AA	amino acid
acc.	accession
Adgb	Androglobin
BLAST	Basic Local Alignment Search Tool
CaM	Calmodulin
CLS	ciliary localisation signal
CPC	ciliary pore complex
Cygb	Cytoglobin
DHC5	dynein heavy chain 5
DHC6	dynein heavy chain 6
EZH2	Enhancer of zeste homolog 2
GbE	Globin E
GbX	Globin X
GbY	Globin Y
glob2	globin 2
glob3	globin 3
Hb	Hemoglobin
KD	knockdown
kDa	kilo Dalton
KIF17	Kinesin Family Member 17
KO	knockout
Mb	Myoglobin
mRNA	messenger ribonucleic acid
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
N. vectensis	<i>Nematostella vectensis</i>
NCBI	National Center for Biotechnology Information
Ngb	Neuroglobin
NLS	nuclear localisation signal
NO	Nitric oxide
NPC	nuclear pore complex

O ₂	oxygen
PC1	polycystin-1
qPCR	quantitative polymerase chain reaction
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
RNA-Seq	RNA-Sequencing
RNS	reactive nitrogen species
ROS	reactive oxygen species
<i>S. purpuratus</i>	<i>Strongylocentrotus purpuratus</i>
<i>S. rosetta</i>	<i>Salpingocea rosetta</i>
shRNA	short hairpin ribonucleic acid
STXBP5-AS1	Syntaxin Binding Protein 5 Antisense RNA 1

1 Introduction

1.1 Globins: a diverse gene family

Globins are a family of small, globular proteins which are widely distributed over the phylogenetic tree of life. Their most famous members in vertebrates, Myoglobin (Mb) and Hemoglobin (Hb), are known for their vital role in oxygen storage and transport. Globin proteins are characterized by a highly conserved three-dimensional structure, where eight α -Helices (A-H) fold in a so-called 3-over-3 sandwich, the “globin”-fold (Figure 1-1, A). The resulting hydrophobic pocket allows binding of a heme prosthetic group at the first phenylalanine between helices C and D. At this heme, a variety of gaseous ligands such as O₂, CO₂ and NO may bind reversibly. The iron atom at the centre of the heme group is coordinated at four positions via nitrogen from the protoporphyrin and by the histidine at position F8 (proximal histidine) of the protein. Depending on the state of the sixth coordination site, globins may be distinguished into two groups: in penta-coordinated globins, such as the classical Mb, the distal binding site is free, whereas in hexa-coordinated globins, the iron atom is additionally coordinated by the side chain of histidine (or glutamine) at position 7 of the E-Helix (distal histidine E7) (Figure 1-1, B and C). This side chain competes with ligand binding to the iron atom, resulting in a structural shift which enables globins to act as sensors or signalling proteins (Reeder et al., 2011).

For nearly a century, Hb and Mb were regarded as the only members of this group of proteins. However, with the discovery of Neuroglobin (Ngb) (Burmester et al., 2000) and Cytochrome b5 (Cygb) (Kawada et al., 2001; Burmester et al., 2002; Trent & Hargrove, 2002), it became evident that vertebrates harbour a surprisingly diverse set of globins:

Neuroglobin was initially described in brain tissue from humans and rodents, hence its name (Burmester et al., 2000). It is highly expressed in the hypothalamus (Fabrizius et al., 2016), but lower amounts can also be found in several other central and peripheral nervous system regions. A few non-neuronal, endocrine tissues such as the testis, adrenal and pituitary gland do also express Ngb (Reuss et al., 2002; Fabrizio et al., 2016).

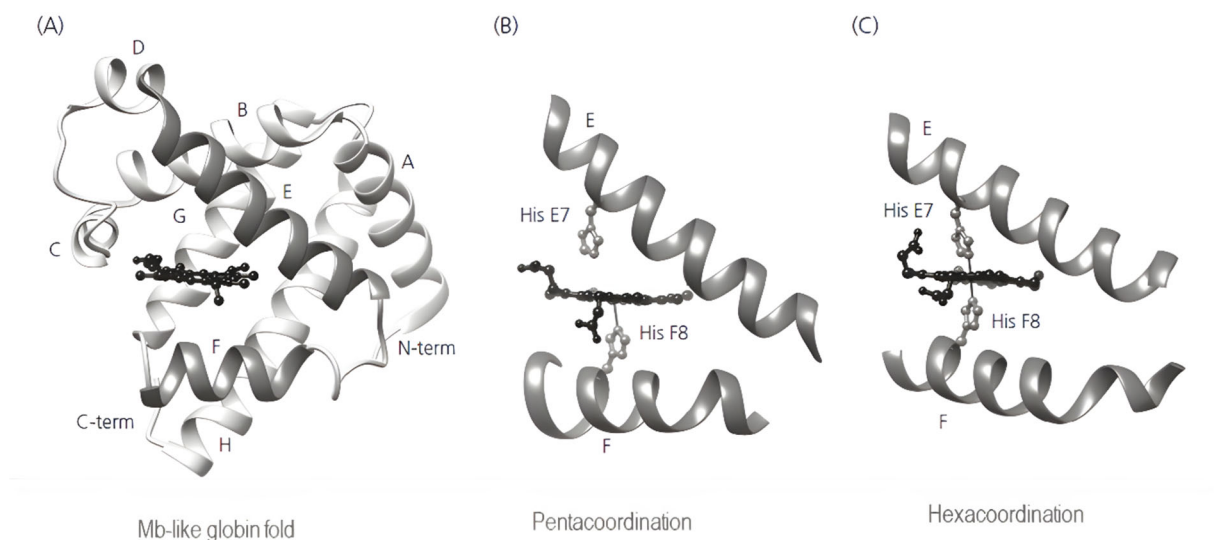


Figure 1-1: Structural models of the globin fold with its heme group (black). (A) 3-over3- helix structure of Mb. (B) Penta-coordinated sperm whale Mb. The distal binding site is accessible for ligand binding. (C) Hexa-coordinated Neuroglobin (Ngb). The distal binding site is occupied by HisE7, which competes with the external ligand binding. Helices of the globin fold are named A-H starting from the N-terminus. Adapted from Storz, 2018.

Cytoglobin is mainly expressed in fibroblasts and related cell types and therefore found in many organs (Kawada et al., 2001; Burmester et al., 2002; Nakatani et al., 2004). In addition, it is expressed in the smooth muscle cells of the vascular system, where it regulates blood pressure via the consumption of NO (Liu et al., 2017). Melanocytes and some cell lines derived from melanomas show elevated levels of Cygb (Fujita et al., 2014), as well as adipocytes and adipose tissues (Doğan et al., 2017; Thuy et al., 2020; Elena Porto, unpublished data).

Globin X (GbX) was first identified via sequence analysis in frogs and fish (Rösner et al., 2005) and is the only membrane bound member of the vertebrate globin family (Blank, Wollberg, et al., 2011). GbX expression could be observed in a variety of tissues, such as brain, eye, heart, intestine, liver and spleen and differs between taxa (Blank, Wollberg, et al., 2011; Schwarze et al., 2015; Gallagher & Macqueen, 2017). In vitro studies suggest that it could protect cells from ROS (Koch & Burmester, 2016) and it was unexpectedly reported in red blood cells of fish, where it could act as a fast nitrite reductase (Corti et al., 2016).

Globin E (GbE) demonstrates an even higher oxygen affinity than the specialized Mb isoforms in diving mammals, making it an ideal candidate for oxygen supply (Blank, Kiger, et al., 2011; Wright & Davis, 2015). Its expression pattern is very distinct: in birds and turtles, GbE mRNA was detected only in the eye (Kugelstadt et al., 2004; Schwarze et al., 2015). In lungfish, where

the GbE gene underwent duplications, all gene copies are exclusively expressed in the ovary (Lüdemann et al., 2019). The exact function of this globin, however, is still not clear.

Of all vertebrate globins, **Globin Y** (GbY) has been studied the least. It was initially discovered in the genome of *Xenopus tropicalis* (Fuchs et al., 2006) and later also identified in several fish species and turtles. Like Cygb, it is widely expressed and the cell- and organ-specific expression sites vary depending on the species analysed. In fish, GbY is expressed in gills and the spleen whereas in turtles it is mostly found in kidney, lung and brain (Schwarze et al., 2015; Gallagher & Macqueen, 2017).

Androglobin (Adgb), discovered in 2012, is the focus of this thesis, and will therefore be discussed in more detail below (Chapter 1.4).

1.2 Globin phylogeny: ancestral and vertebrate-specific lineages

Globins have been described in bacteria and archaea (Vinogradov et al., 2013), indicating that they were already present in the last universal common ancestor. Horizontal gene transfer during endosymbiotic events may have led to the incorporation of the first globin sequence into a eucaryotic genome (Vinogradov et al., 2007). Their rather short sequences of about 150 amino acids with high amounts of primary sequence divergence pose difficulties for phylogenetic reconstructions, however, the highly conserved folding pattern allows for robust identification of family members (Herman et al., 2014). In addition, globin genes carry characteristic introns at fixed positions, namely one in the 12th amino acid of the B-Helix (B12.2) as well as one in the G-Helix (G7.0). These introns may have already been present in the last common ancestor of plants and animals (as reviewed in: Hardison, 1996) and also do facilitate the identification of potential globin gene family members. In Figure 1-2, an overview over the globin lineages is given, as well the current hypothesis about their phylogenetic relationship:

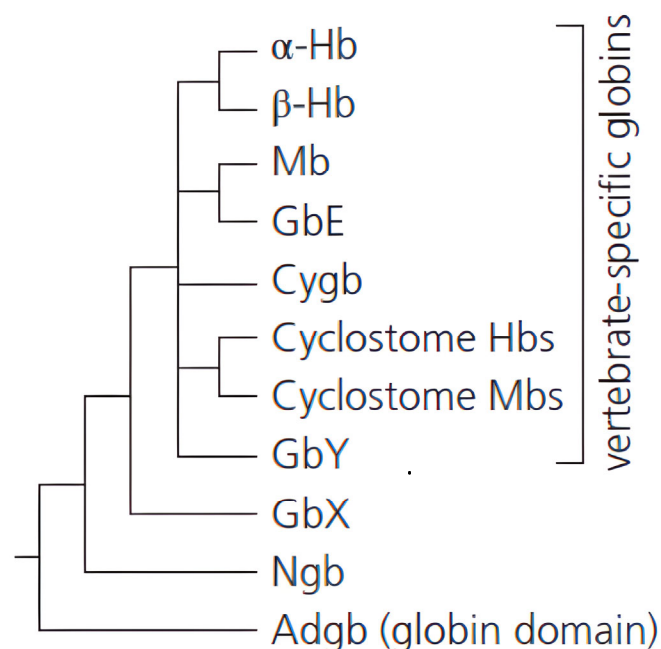


Figure 1-2: Phylogeny of vertebrate globins. Adgb, Ngb and GbX are the ancestral metazoan globin lineages, whereas the vertebrate specific globins developed during the whole genome duplication events which preceded the emergence of the vertebrate subphylum. Adapted from Storz, 2018.

The emergence of Ngb, one of the “oldest” globins, dates back prior to the split of protostomes and deuterostomes (Burmester et al., 2000). In phylogenetic analyses, vertebrate Ngb clusters with globin sequences from sea urchins and cephalochordates as well as annelid nerve globins. A possible Ngb ortholog was found in the placozoan *Trichoplax*, thus predating the emergence of a real nervous system (Dröge et al., 2012; Hankeln and Ebner, unpublished). It has been proposed that Ngbs represent the most ancient clade of the classically arranged, contemporary globins (Hoffmann et al., 2012), but was independently lost several times during evolution, as it is the case in hemichordates and tunicates, some cartilaginous fishes (Schwarze et al., 2014; Opazo et al., 2015) and also in arthropods (Prothmann et al., 2020).

The evolution of GbX is special: it is, on the one hand, an ancient globin such as Ngb with a very wide distribution over protostomes as well as deuterostomes (Hoffmann et al., 2012; Prothmann et al., 2020). On the other hand, the GbX repertoire has also been expanded on several occasions during globin evolution. Phylogenetic analysis of the globin repertoires of the acorn worm and branchiostoma suggest that there are two main GbX lineages, GbX and GbX-like (or Clade 2 and 3), which reflect a duplication early in evolution (Hoffmann et al., 2012; Blank & Burmester, 2012). In line with this, members of both lineages were identified in arthropod genomes and phylogenetic reconstruction clusters both at the base of the phylogenetic tree (Prothmann et al., 2020). Many vertebrate taxa only harbour one variant of GbX, whereas the GbX-L lineage was lost. However, during comparative phylogenetic analyses it became clear that this perceived orthology was in fact hidden paralogy: The GbX locus was duplicated during one of the two WGD events which precede the emergence of vertebrates (Figure 1-3). One of these proto-GbX variants was then respectively lost in cyclostomes and gnathostomes and the second duplicated, probably due to additional WGD or polyploidization events (Opazo et al., 2015; Gallagher & Macqueen, 2017; Hoffmann et al., 2021). In gnathostome evolution, both copies were differentially retained and duplicated several times, and sometimes copies of closely related taxa such as *Xenopus* and the salamanders are only paralogous (Figure 1-3). In contrast, Androglobin, which is the third ancient metazoan globin lineage is almost exclusively present as single copy. In fact, Adgb originated already at the very base of the metazoan tree and can be traced back to the last common ancestor of choanoflagellates and animals (Hoogewijs et al., 2012).

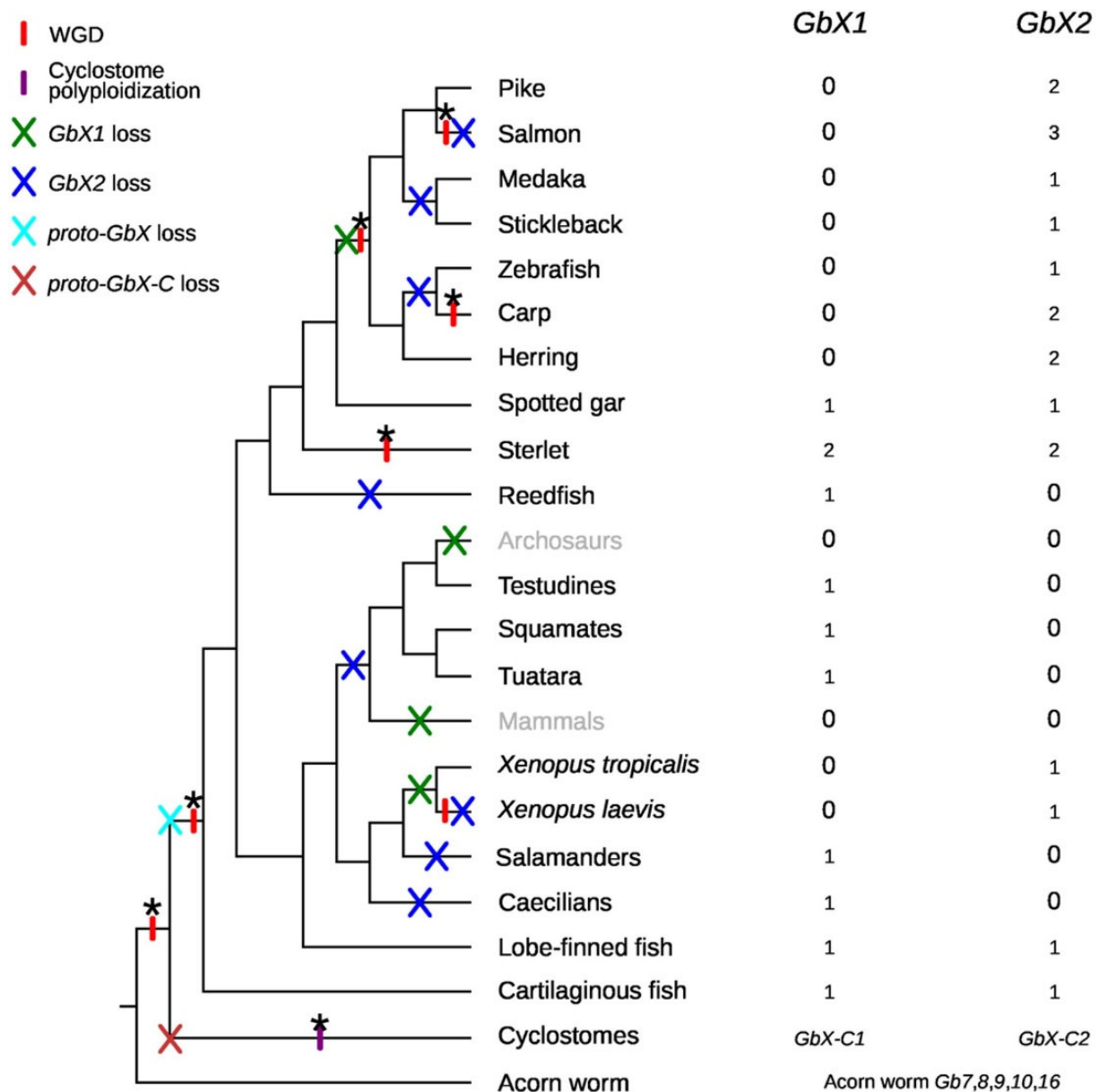


Figure 1-3: Model of the evolution of GbX in vertebrate species. The GbX locus was duplicated during one of the two WGD events which precede the emergence of vertebrates. One of these proto-GbX variants was then independently lost in cyclostomes and gnathostomes and the second duplicated (probably due to additional WGD events). Note that this model reflects the hypothesis that gnathostomes and cyclostomes only shared one of the two rounds of WGD of vertebrates. During gnathostome evolution, additional duplications and loss events occurred and although many of the analysed genomes only contain one GbX copy, they are not necessarily orthologous to each other. Adapted from Hoffmann et al. 2021.

The lineage of the globin family, which consist of vertebrate-specific members, can be considered a textbook example of how whole genome and gene duplication events may drive diversification of gene families (Storz et al., 2013; Hoffmann et al., 2021). All contemporary vertebrate-specific globin genes are probably derived from a single common ancestor. Through subsequent whole genome duplication events during vertebrate evolution, 4 paralogous genomic loci arose: Cygb, Mb, Hb and a fourth globin locus which later on lost its globin gene copy. Consecutively, a tandem gene duplication in the Mb locus in the stem lineage of jawed vertebrates created the GbE gene. GbE was secondarily lost again multiple times and now only remains in birds, turtles (Schwarze et al., 2015), and the coelacanth (Schwarze & Burmester, 2013). GbX, GbE and GbY were all secondarily lost in the mammalian lineage. Currently, the coelacanth and turtles are the only taxa known to harbour the full set of globin diversity (Schwarze et al., 2015). A summary of the distribution of all globins among the vertebrate lineages can be found in Figure 1-4. The retention of both the proto-Hb and Mb in the ancestor of jawed vertebrates allowed for the specialisation towards O₂ storage (Mb) or transport function (Hb). Additional duplication of the proto-Hb locus resulted in the two subunits α and β , which permitted the formation of a multimeric and heterogenous Hb and laid the foundation of development of sophisticated regulatory transport mechanism such as allostery. Intriguingly, the specialized O₂ transport in blood via Hb seems to have evolved convergently in cyclostomes and jawed vertebrates (Hoffmann et al., 2010; Schwarze et al., 2014). Also in cyclostomes, transport of oxygen is realised by a tetrameric and storage by a monomeric globin variant, but the underlying molecular cooperativity mechanisms as well as the oxygen affinity show significant differences (Fago & Weber, 1995; Qiu et al., 2000; Fago et al., 2001).

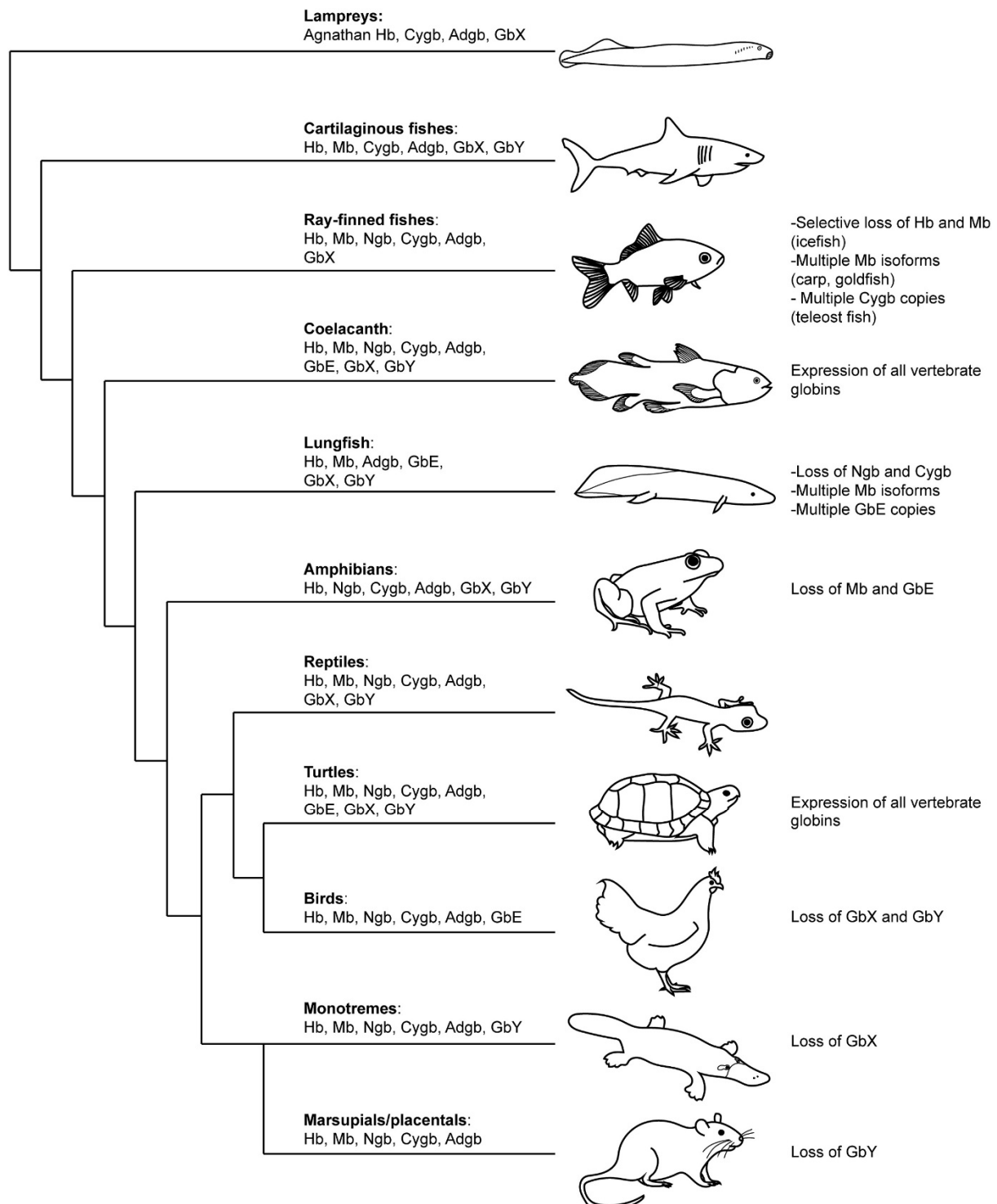


Figure 1-4: Occurrence of globins among vertebrates. On the branch of the tree, the globin variants are listed which could be identified in the corresponding clade. Events of losses or species specific duplications are described on the right. Note that of the extant lineages, only turtles and the coelacanth harbour the full set of vertebrate globins. A more detailed depiction of the evolutionary events is found in the text. Adapted from Keppner et al., 2020.

1.3 Functions of conventional globins in health and disease

Monomeric globins are famously known for their role in oxygen storage and transport. However, the precursor globin, which all our contemporary representatives were derived from, probably served a completely different function, since oxygen was rather a cytotoxin than an enabler of aerobic energy production in early evolution more than 1 billion years ago (Hardison, 1998; Vinogradov et al., 2007). Conclusively, research on the properties of globins has shown that they do not only bind and provide oxygen but fulfil a variety of additional functions (Burmester & Hankeln, 2014; Keppner et al., 2020). Cygb, for example, regulates NO bioactivity (as reviewed in: Mathai et al., 2020) and thus influences blood pressure and vascular tone in mammals *in vivo* (Liu et al., 2017). Mouse work has shown that Mb can act as an NO scavenger to protect cardiac function (Gödecke 2003). Depending on its oxygenation state, it may also produce NO, a signalling molecule, under low oxygen environments (Cossins 2008, Hendgen-Cotta 2008). Cygb and Mb also serve as protectors against oxidative stress: Mb exhibits catalase and peroxidase activity which neutralizes toxic H₂O₂ in mammals (Mannino et al., 2019) as well as fish (Helbo et al., 2012). In vitro studies have shown that Cygb expression is beneficial under oxidative stress (e.g.: Fordel et al., 2006; Nishi et al., 2011). Furthermore, work in additional mouse models confirms its antioxidant role, as aged Cygb knockout mice showed signs of an imbalanced antioxidant defence system with severely damaged organs (Thuy et al., 2016). In line with this, Cygb overexpressing mice were more resistant to induced oxidative stress and less prone to hepatic fibrosis (Thi Thanh Hai et al., 2018). Also for Ngb, a cytoprotective function towards ROS and RNS has been proposed (Burmester & Hankeln, 2009; Ascenzi et al., 2016). Ectopic overexpression of Ngb enhanced H₂O₂ tolerance in cell culture models (Fordel et al., 2006; Li et al., 2007) and ameliorated recovery after ischemic injury in mouse and rat models (as reviewed in: Van Acker et al., 2018). In vitro data suggests that Ngb may act as a NO scavenger in its oxygenated form (Brunori et al., 2005), but up to now the reduction system that would regenerate the heme has not been found (Smagghe et al., 2008). In its deoxygenated form, Ngb might reduce Nitrate to NO, which could then inhibit mitochondrial respiration and thus have cytoprotective effects under hypoxia (Petersen et al., 2008; Tiso et al., 2011).

The hypoxic environment and prevalence of ROS/RNS are widely known hallmarks of cancer and may serve as basis for therapeutical approaches (Trachootham et al., 2009; Sosa et al., 2013). Whereas high amounts of ROS lead to extensive cell damage, low concentrations could promote cell growth and proliferation and ultimately enhance malignancy (Sosa et al., 2013). Since globins are able to produce or scavenge ROS depending on their oxygenation state, they can both promote and inhibit tumour growth and are therefore promising candidates for cancer research. For some family members, a deregulation in cancer entities has already been described. Mb is upregulated in a variety of epithelial tumours, such as breast, lung, ovary, and colon carcinomas (Flonta et al., 2009; Kristiansen et al., 2010), and can even have prognostic value (as reviewed in: Elkholi et al., 2022). In brain tumours, for example, ectopic Mb expression correlates with a more aggressive tumour phenotype (Elsherbiny et al., 2021), whereas in breast and neck cancer, expression of Mb is considered advantageous and leads to a better outcome for the patient (Kristiansen et al., 2010, 2011; Meller et al., 2016). Cygb, on the other hand, is mostly downregulated in cancer entities (as reviewed in: Oleksiewicz et al., 2011). In gliomas and pancreatic ductal adenocarcinomas, reduced endogenous expression of Cygb was correlated with a more fatal tumour phenotype and higher rates of recurrence (Xu et al., 2013; Kono et al., 2021). These findings could be confirmed in in vivo Cygb KO models. Two independent mouse lines (Thi Thanh Thuy et al., 2011; Yassin et al., 2018) showed higher tumour incidence upon Cygb ablation in liver, lung and colon tissue. Ngb, in contrast, seems to have a dualistic function in cancer: it has been described both as an oncogene as well as a tumour suppressor in different entities. In breast cancer as well as glioma cell culture models, Ngb might promote proliferation and suppress apoptosis (Zhang, B. et al., 2018; Fiocchetti et al., 2018). In line with this potential oncogene function, antibody stainings showed a potential upregulation in several tumour samples (Emara et al., 2009, 2010). However, it has been contested whether the cell culture models that were used to study Ngb's oncogenic attributes were suitable: a comprehensive reanalysis of Ngb expression sites found no evidence of endogenous Ngb mRNA expression in a variety of cell lines and most of these cancer studies relied on antibodies which were not thoroughly validated (Fabrizius et al., 2016). In addition, antibody-free in situ hybridisation analysis of Ngb in a cancer tissue array showed no significant expression of Ngb in a variety of malignancies (Gorr et al., 2011). This is also reflected in the transcriptomic analysis performed by the Cancer

Genome Atlas project¹ (e.g.: Blum et al., 2018; Hutter & Zenklusen, 2018): Ngb is downregulated in all cancer entities derived from endogenous expression sites (Carina Osterhof and Thomas Hankeln, unpublished), which strongly argues against an oncogenic function. It has also been proposed that Ngb could act as a tumour suppressor in hepatocellular carcinoma (Zhang, J. et al., 2013), but also here, the expression analysis shows no evidence of Ngb mRNA expression in healthy tissues. It is therefore unknown whether Ngb plays any role in malignancies *in vivo*.

¹ <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>

1.4 Androglobin – an ancient globin with a peculiar structure

Androglobin (Adgb) is the most recent addition to the diverse family of globins. It was discovered in 2012 in parallel in the echinoderm *Strongylocentrotus purpuratus* and the cephalopod *Branchiostoma floridae* (Hoogewijs et al., 2012). Through an extensive search of public datasets, the Adgb gene could additionally be identified in most taxa of the metazoans, including mammals, as well as in some unicellular species of choanoflagellates (Figure 1-7). Unfortunately, it was lost in several genetical model species, such as *Caenorhabditis elegans* and *Drosophila melanogaster*, which has complicated functional studies so far.

Adgb was discovered rather late given its widespread existence, which is probably due to its unique globin domain structure (cf. Figure 1-5): instead of the typical A-H sequence of α -helical segments, it starts with helices C-H followed by a putative IQ calmodulin binding site. Helices A and B form the C-terminal end of the domain, including the characteristic globin intron at B12.2. This rearranged domain presumably resulted from multiple mutation events, starting with a tandem duplication of the globin domain followed by truncation of the now redundant additional helices. The IQ domain could have been acquired secondarily: the motif is enclosed by two introns of the same phase, meaning that both introns are inserted at the same codon position. This indicates that Adgb gained this motif via an exon-shuffling event. Regardless of these profound structural differences, alignments of mammalian globins and subsequent 3D modelling of helices C-H suggest that Adgb forms a functional globin fold (Hoogewijs et al., 2012), as it was already described for a comparably truncated form of Mb (Ribeiro & Ramos, 2005). Intriguingly, the recombinantly expressed truncated human Adgb globin domain exhibited absorption spectra characteristic of a typical hexacoordinated globin – despite the substitution of HisE7 by glutamine.

In contrast to all other globins, which are usually comprised of about 150 amino acids with only small N- or C-terminal extensions, Adgb is more than ten times larger (~1600 amino acids) and contains additional protein domains (as shown exemplarily for the human variant in Figure 1-5). At the N-terminus, the protein sequence exhibits high similarity to the catalytic domain of human calpain-7. However, only one of the three important active site residues is conserved, which indicates that Adgb is possibly not proteolytically active. In addition, it carries two large domains with currently unknown function: one between the calpain and globin domain (~300 aa) and a second, larger one at the C-terminus (~750 aa).

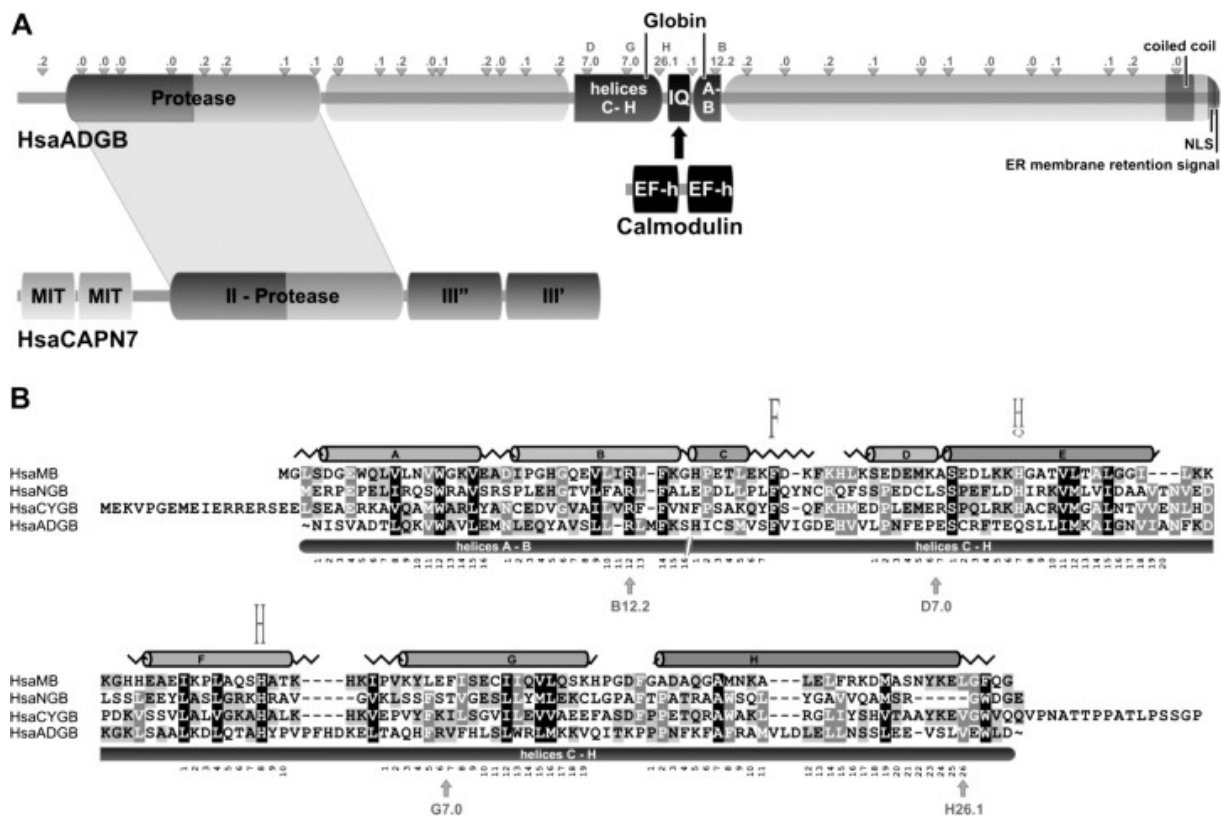


Figure 1-5: Structure of Androglobin and comparison of its globin domain with other vertebrate globins. A) Schematic overview of Adgb's protein domains and intron positions. The protein starts with a putative calpain protease domain. The globin domain is situated in the centre, but helices A and B are shifted towards the end. The domain is also interrupted by a calmodulin binding IQ-motif. At the C-terminus, a coiled coil domain and a nuclear localisation signal were predicted bioinformatically. B) multiple sequence alignment of human MB, CYGB, NGB and the manually rearranged globin domain of ADGB. Adapted from (Hoogewijs et al., 2012)

qRT-PCR analysis on different tissues of mice showed that Adgb is mainly expressed in testis and, at a tenfold lower level, in lungs. Minor levels of Adgb mRNA could also be detected in brain and heart (Figure 1-6, A). A more detailed analysis of testis development revealed that Adgb mRNA expression surges at postnatal day 25, which concurs with the postmeiotic stage of sperm differentiation (Figure 1-6, B). In situ hybridization experiments on cryo-sections of mouse testis tissue displayed stronger signal intensities towards the lumen of the seminiferous tubules, confirming this association with late stages of spermatogenesis (Figure 1-6, C (Hoogewijs et al., 2012)). Reanalysis of public microarray data comparing expression levels of fertile and infertile males even suggested a 4-fold downregulation of ADGB in infertile human individuals, hinting at a crucial function of ADGB in reproduction.

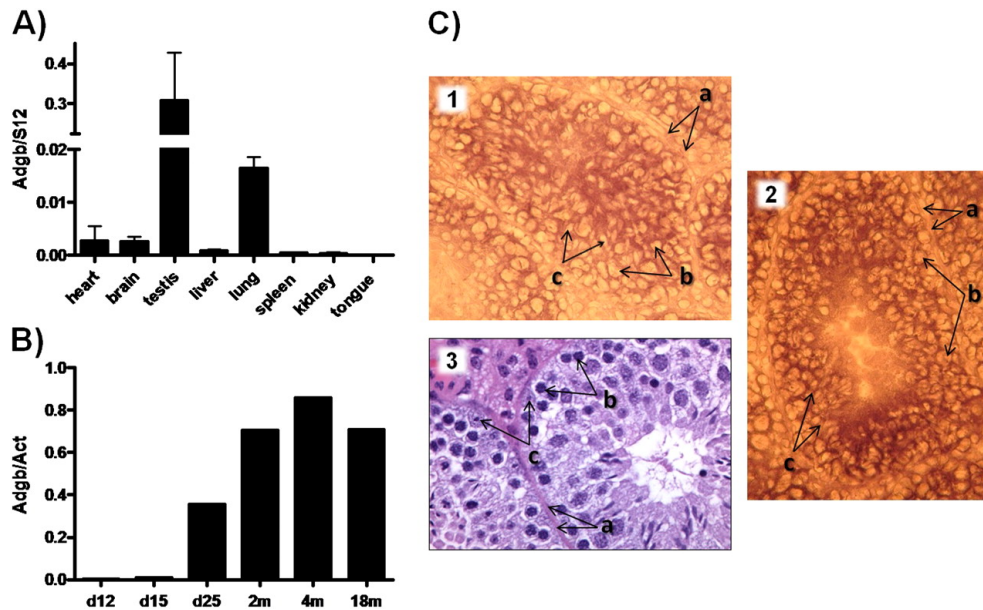


Figure 1-6: mRNA expression analysis of Adgb in mice. Adgb mRNA expression was measured by qRT-PCR in tissues (A) and different stages of testis development (B). C) mRNA ISH experiments performed on sections of testis from mice. The signal intensifies towards the lumen of the seminiferous tubules. a) smooth muscle cells; b) spermatogonia; c) Sertoli cells. Adapted from (Hoogewijs et al., 2012)

Analogous to its other family members (Chapter 1.3), the structurally complicated and functionally elusive Adgb has been studied in the context of cancer as well. Adgb expression was downregulated experimentally in a glioblastoma cell culture model via lentiviral mediated shRNA. Colony formation and MTT assays showed that Adgb-knockdown cells proliferated more slowly and displayed apoptotic attributes (Huang, B. et al., 2014). These results suggest that Adgb has oncogenic function in glioma cell lines. A second study (Chen et al., 2020) analysed Adgb in pancreatic cancer. The authors focused initially on STXBP5-AS1, a long non-coding RNA that was already described as a tumour suppressor before (Huang, J. et al., 2018; Cen et al., 2019). They found that downregulation of STXBP5-AS1 was correlated with a poor prognosis in pancreatic cancer and that upregulation compromised stemness and metastasis in several pancreas-derived cell lines. Mechanistically, they proposed that the lnc-RNA recruits EZH2, which then epigenetically silences the neighbouring ADGB gene. Immunoprecipitation analysis of EZH2 confirmed binding of the transcription factor to the ADGB promotor region. In line with this, ectopic co-expression of ADGB and STXBP5-AS1 reversed the positive effects of the latter, supporting ADGBs role as an oncogene in this pancreatic cancer model (Chen et al., 2020). It remained unclear, however, if ADGB is expressed in real tumour entities and whether it has any prognostic value there.

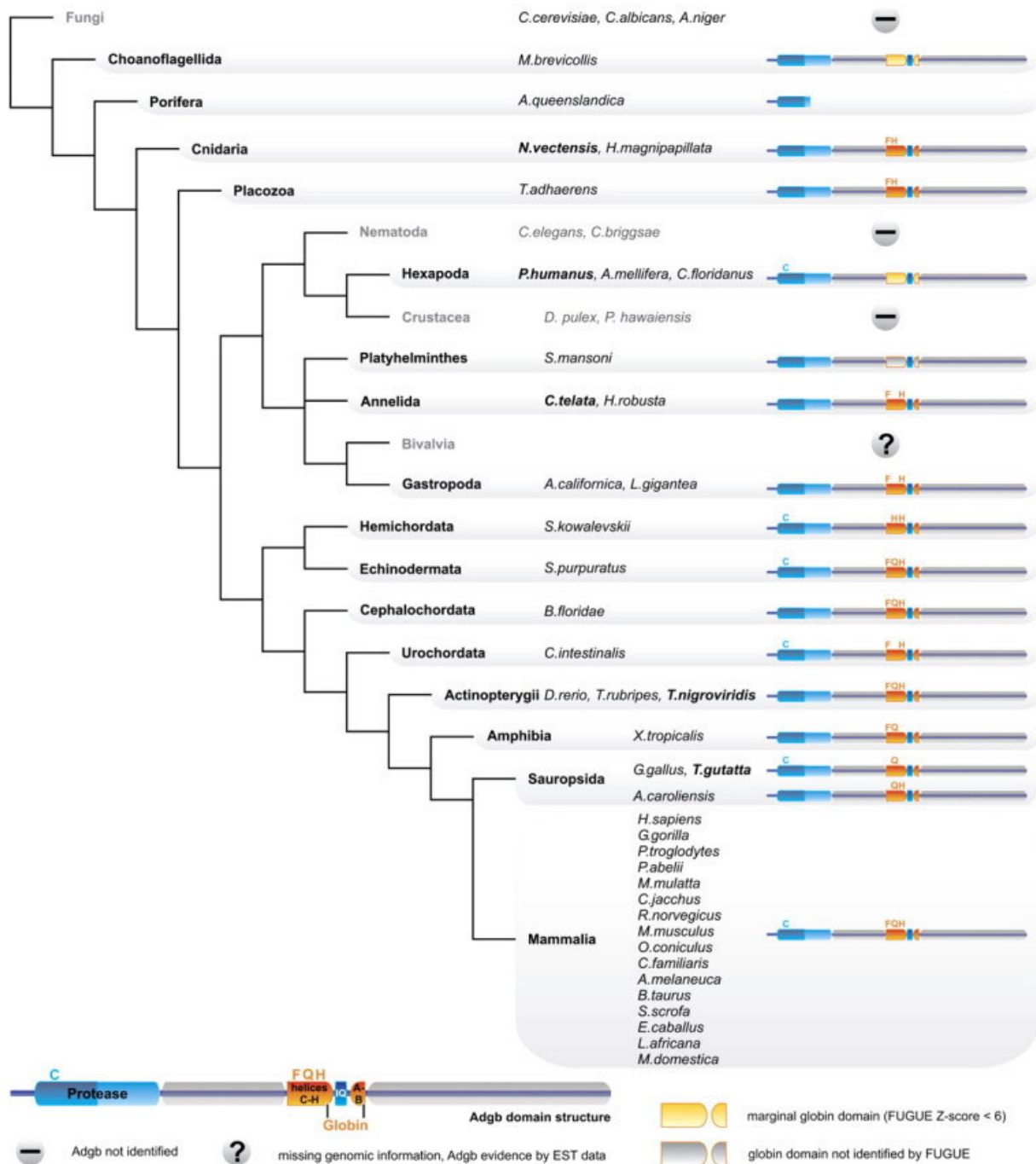


Figure 1-7: Phylogenetic tree of Androglobin homologues identified through bioinformatic analysis of genome databases. On the right: schematic depiction of conserved elements of Adgb domain structure and globin domain confidence scores. Note the absence of Adgb orthologs in Nematodes and Crustaceans and possibly truncated version in sponges. Adapted from Hoogewijs et al. 2012.

1.5 Aim of this study

Although 10 years have passed since its initial description, research on Adgb is still in its early stages. We will therefore take a basic approach to characterize Adgb's role by focusing on its expression patterns across a variety of taxa and tissues.

The following questions will be considered:

1) *When and where is Androglobin expressed?*

Early studies showed high amounts of Adgb mRNA in testis tissue and lower amounts in both lungs and brain. However, it is not clear whether this expression pattern is driven by a common cell type (which would suggest a similar function in all tissues) or whether these tissues should be addressed separately in the functional analysis of Adgb. Since its initial description in 2012, a vast amount of additional expression data was generated by omics techniques which allows a more detailed view on Adgb mRNA expression across tissues. We will therefore integrate bulk as well as single cell mRNA sequencing data from a variety of sequencing efforts to elucidate the underlying expressing cell types. Bioinformatical findings will then be validated in the wet lab via qPCR and immunohistological stainings (Chapter 2.1).

2) *Is the expression pattern of Adgb conserved?*

Some globins show very restricted expression patterns (e.g. Ngb, which is predominantly expressed in neuronal tissues) while others display a diverse set of expression sites in different cell types and organs, which may even change in different taxa (e.g. GbX, as described above). Due to its evolutionary antiquity, Androglobin probably has a conserved function across the metazoan tree. Unfortunately, genetic model organisms such as *Drosophila melanogaster* or *Caenorhabditis elegans* have lost the Adgb gene and are consequently not useful for functional studies here. We will therefore analyse bulk and single cell mRNA sequencing data from a diverse set of metazoan taxa to elucidate the expression pattern of Adgb across the phylogenetic tree and to identify suitable model organisms for further functional analysis of Adgb (Chapter 2.4).

3) *Which impact does the modulation of endogenous Adgb levels have?*

In humans and mice, Adgb is expressed at the highest level in testis tissue and comparative expression studies on fertile and infertile men suggest an association of ADGB with fertility. It is thus not unlikely that Adgb exerts a function in a basal process such as spermatogenesis. We will therefore analyse the transcriptomic changes upon ablation of Adgb in testis tissue of mice (Chapter 2.2). Together with the phenotypic analysis of these Adgb KO mice performed by our cooperation partners, we will elucidate Adgb's potential function in sperm formation.

4) *Is Androglobin expressed in cancer entities, and does this have prognostic value?*

In vitro studies have suggested that Adgb acts as oncogene, but no studies have been performed on real patient samples. We will therefore analyse the expression pattern of Adgb in published RNA-Seq datasets of tumours (e.g. from the TCGA¹ consortium) to detect changes in expression during tumour progression. In addition, we will stably overexpress Adgb in a lung cancer cell model. Subsequent RNA-Seq analysis may add further evidence of Adgb's oncogenicity and provide insight into the mode of action of Adgb in cancer entities (Chapter 2.3).

2 Results

The results presented in this thesis were already published in scientific journals (Chapters 2.1 & 2.2) or composed in a way which would allow publishing (Chapters 2.3 & 2.4).

2.1 Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis

Teng Wei Koay*, Carina Osterhof*, Ilaria M.C. Orlando, Anna Keppner, Daniel Andre, Schayan Yousefian, María Suárez Alonso, Miguel Correia, Robert Markworth, Johannes Schödel, Thomas Hankeln[#], David Hoogewijs[#] (2021) Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis. *Journal of Biological Chemistry* 296:100291

*/[#] Authors contributed equally

Supplementary files: <https://doi.org/10.1016/j.jbc.2021.100291>

Own contributions to this publication:

- Mapping and quantification of published bulk RNA-Seq data on tissues from human and cattle, revealing a new expression site for Androglobin in the female reproductive tract
- Differential gene expression and correlation analysis on bulk RNA-Seq data to identify genes strongly associated with the presence of Adgb mRNA; subsequent GO-term analysis and interpretation of gene lists
- Experimental validation of the new expression site via qRT-PCR and correlation analysis with known ciliary marker genes
- Immunohistochemical verification of Adgb protein expression in testis and fallopian tube tissue from cattle
- Clustering analysis of publicly available single cell sequencing data on murine lung and testis

Planning of experiments, analysis, and interpretation of data as well as drafting of the corresponding part of the manuscript were realized together with Prof. Dr. T. Hankeln. The project was managed by Prof. Dr. David Hoogewijs and Prof. Dr. T. Hankeln.

Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis

Received for publication, August 26, 2020, and in revised form, December 17, 2020 Published, Papers in Press, January 13, 2021,
<https://doi.org/10.1016/j.jbc.2021.100291>

Teng Wei Koay^{1,†}, Carina Osterhof^{2,†,✉}, Ilaria M.C. Orlando¹, Anna Keppner¹, Daniel Andre², Schayan Yousefian², María Suárez Alonso¹, Miguel Correia¹, Robert Markworth³, Johannes Schödel⁴, Thomas Hankeln^{2,✉}, and David Hoogewijs^{1,*,#}

From the ¹Section of Medicine, Department of Endocrinology, Metabolism and Cardiovascular System, University of Fribourg, Fribourg, Switzerland; ²Faculty of Biology, Institute of Organismic and Molecular Evolution, Molecular Genetics & Genome Analysis, Johannes Gutenberg University Mainz, Mainz, Germany; ³Institute of Physiology, University of Duisburg-Essen, Duisburg, Germany; ⁴Department of Nephrology and Hypertension, Universitätsklinikum Erlangen and Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Edited by John Denu

Androglobin (ADGB) represents the latest addition to the globin superfamily in metazoans. The chimeric protein comprises a calpain domain and a unique circularly permuted globin domain. ADGB expression levels are most abundant in mammalian testis, but its cell-type-specific expression, regulation, and function have remained unexplored. Analyzing bulk and single-cell mRNA-Seq data from mammalian tissues, we found that—in addition to the testes—ADGB is prominently expressed in the female reproductive tract, lungs, and brain, specifically being associated with cell types forming motile cilia. Correlation analysis suggested coregulation of ADGB with FOXJ1, a crucial transcription factor of ciliogenesis. Investigating the transcriptional regulation of the *ADGB* gene, we characterized its promoter using epigenomic datasets, exogenous promoter-dependent luciferase assays, and CRISPR/dCas9-VPR-mediated activation approaches. Reporter gene assays revealed that FOXJ1 indeed substantially enhanced luciferase activity driven by the *ADGB* promoter. ChIP assays confirmed binding of FOXJ1 to the endogenous *ADGB* promoter region. We dissected the minimal sequence required for FOXJ1-dependent regulation and fine mapped the FOXJ1 binding site to two evolutionarily conserved regions within the *ADGB* promoter. FOXJ1 overexpression significantly increased endogenous ADGB mRNA levels in HEK293 and MCF-7 cells. Similar results were observed upon RFX2 overexpression, another key transcription factor in ciliogenesis. The complex transcriptional regulation of the *ADGB* locus was illustrated by identifying a distal enhancer, responsible for synergistic regulation by RFX2 and FOXJ1. Finally, cell culture studies indicated an ADGB-dependent increase in the number of ciliated cells upon overexpression of the full-length protein, confirming a ciliogenesis-associated role of ADGB in mammals.

Globins are small globular metallo-proteins consisting of about 150 amino acids, which comprise eight α -helical segments in a characteristic 3-over-3 α -helical sandwich structure. This conserved “globin fold” identifies them as members of a large protein superfamily. Globins contain a heme prosthetic group, by which they can reversibly bind gaseous ligands such as O₂, CO, and NO. Historically, the familiar vertebrate O₂-binding hemoglobin (HB), a tetramer of α - and β -globins, and the monomeric myoglobin (MB) were among the first proteins whose sequences and structures were determined already over 50 years ago. Genomic analyses have considerably altered and extended our view of the globin family in mammals, leading to the discovery of novel globin types such as neuroglobin (NGB) and cytoglobin (CYGB), which are expressed in nerve and fibroblast-like cells, respectively (1, 2). Both globin types perform yet-to-be-illuminated functions, which possibly reside in antioxidant defense, reactive oxygen species signaling, or even lipid metabolism (3, 4).

Recently, a novel family of large, chimeric proteins containing a globin-like domain was discovered and termed androglobin (ADGB) based on its preferential expression in mammalian testis tissue (5). ADGB is a chimeric protein of about 1500 amino acids, which contains an embedded globin domain. This globin domain is permuted with respect to its characteristic α helices and interrupted by a calmodulin-binding motif. Nevertheless, the globin domain appears to be able to bind oxygen *in vitro* (5). The N-terminal domain of ADGB shows high sequence similarity to the human protease calpain 7, although functionally important amino acid residues are mutated. ADGB was shown to be highly conserved throughout the metazoan tree of life, and orthologous copies of the *ADGB* gene could be found from humans and other vertebrates down to very basal taxa such as the cnidarian *Nematostella vectensis*, the placozoan *Trichoplax adherens*, and even the choanoflagellate *Monosiga brevicollis* (5), which suggests an elementary and possibly conserved function in metazoans. ADGB is predominantly expressed in later stages of spermatogenesis in mammalian testes and, to a much lower

This article contains [supporting information](#).

[†] Contributed equally.

[#] Joint senior authors.

* For correspondence: David Hoogewijs, david.hoogewijs@unifr.ch.

ADGB is involved in ciliogenesis and a target of FOXJ1

extent, in the lung and brain tissue (5). An important role of ADGB in spermatogenesis was supported by analysis of published microarray data revealing that endogenous levels of human ADGB mRNA were lower in the testes of infertile men *versus* their healthy counterparts. An *in vitro* cell culture study suggested that ADGB could act as an oncogene in glioma formation and an ADGB knockdown could inhibit the growth of glioma cell lines (6). Overall, studies on ADGB expression patterns and gene regulation were scarce, and the functional role of ADGB has remained elusive.

Since the expressional profile of a gene, specifically addressing the organs and their cell types, can provide a valuable hint at its possible function (as illustrated *e.g.*, by the specific presence of HB in erythrocytes), we revisited the expression patterns of ADGB using an integrative approach of bioinformatical data mining. In particular, novel RNA-Seq datasets from bulk and single-cell experiments were analyzed with the aim to recognize common patterns with functional implications. The data yielded valuable insight into the properties of ADGB-expressing cell types, which led us to characterize in detail the gene-regulatory landscape determining ADGB expression. We comprehensively mined epigenomic databases for accessible chromatin and promoter/enhancer-associated histone marks, identified transcription factors binding to the *ADGB* locus using reporter gene assays and chromatin immunoprecipitation (ChIP) experiments, and further characterized several functional distal enhancers in the *ADGB* locus. Finally, we performed ADGB overexpression *in vitro* to elicit a cellular phenotype. These different lines of experimental evidence converged and convincingly pointed out that the cellular function of ADGB is associated with the presence of motile cilia.

Results

ADGB expression in female reproductive tract, lung, and brain suggests functional association with ciliary structures

The wealth of gene expression data, which have been produced since the initial description of ADGB in 2012, enabled us to define a much more detailed expression profile of the gene in mammalian tissues and cell types. As such, the bulk RNA-Seq data of the Human Protein Atlas (7) revealed the fallopian tube of the female reproductive tract as a novel expression site of ADGB mRNA (Fig. 1A). Transcript levels were even higher than in the lung, which was initially described as the second highest ADGB-expressing human organ (5). To study this further, and noticing a shortage of data from healthy human samples, we evaluated bulk RNA-Seq data from the female reproductive tract of cattle. The bovine data sets confirmed *Adgb* expression in the oviducts, showing the highest amount of *Adgb* expression of all cattle organs analyzed, and in endometrial tissue (Fig. 1B). Human endometrial data appeared largely devoid of ADGB RNA, but sequencing data of separate stromal and epithelial fractions (8) revealed restriction of ADGB expression to the epithelial fraction only (Fig. S1).

Both human and cattle RNA-Seq data revealed a high interindividual variability in expression intensity, which suggested a temporal and/or spatial restriction of *Adgb* expression in the female reproductive tract. To study the gene expression footprint of the hypothetical *Adgb*-expressing cell type involved, we subdivided the bovine endometrial samples into 2 groups, “*Adgb*-high” (TPM >20, *n* = 4) and “*Adgb*-low” (TPM <5, *n* = 4), and performed differential gene expression analysis to infer genes associated with either high or low levels of *Adgb*. Subsequent overrepresentation analysis (Table 1, Supplemental File 1) revealed that genes associated with high amounts of *Adgb* were connected to GO-terms such as “cilium and axoneme assembly”, “dynein-dependent microtubular transport”, “microtubular movement” and, interestingly, the “sperm flagellum.” An independent clustering approach to identify genes with an *Adgb*-type expression pattern using an additional data set of human fallopian tube samples (9) generated a smaller subset of genes, which were even more strongly associated with cilia-related processes such as “cilium movement”, “determination of left-right-symmetry” and the “differentiation of lung epithelial cells” (Table 2, Supplemental File 2). Among these approximately 100 *Adgb*-associated genes, we found *Foxj1*, the master transcription factor of ciliogenesis (10), and *Dnah5*, a protein known for its specific localization to motile cilia of the respiratory tract (11). The tissue with the highest amount of *Adgb* expression, however, was inconsistent between the samples of the two species (Fig. 1, A and B). A possible explanation could be that the samples were at different stages of the menstrual cycle, given that ciliogenesis is estrogen-dependent (12, 13). Additionally, depending on the part of the oviduct that was dissected, the ratio of epithelial cells to connective tissue and thus the overall number of ciliated cells may vary between samples (14). Reproductive aging (*i.e.*, menopause), which decreases the number of ciliated cells (15), may also have contributed to the observed *Adgb* expression differences.

Experimentally, we confirmed the fallopian tube and the endometrium as novel expression sites *via* RT-qPCR analysis in cattle. In addition, we also determined the amount of *Dnah5* and *Foxj1* mRNA in these samples (Fig. 1C). *Foxj1* transcript levels showed a positive correlation with *Adgb* expression in the endometria and oviducts ($R^2 = 0.73$ and $R^2 = 0.72$, respectively). The association between *Dnah5* and *Adgb* in the endometrium was even stronger ($R^2 = 0.93$). Though not as prominent, there was also a significant positive correlation between expression of *Adgb* and *Dnah5* in the oviducts ($R^2 = 0.74$). Immunohistochemistry analysis further confirmed the localization of *Adgb* protein in the epithelia in the bovine endometrium and specifically in multiciliated cells in the oviduct (Fig. S2).

We previously reported the lung to show the second highest *Adgb* mRNA expression, after the testes (5). Bulk RNA-Seq data from this tissue, however, led to inconclusive results with high interindividual variability and overall low levels of expression, or, as in some human samples, no expression at all (Fig. 1). Therefore, we considered analyzing available single-cell RNA-Seq data obtained from the murine lung (16). To

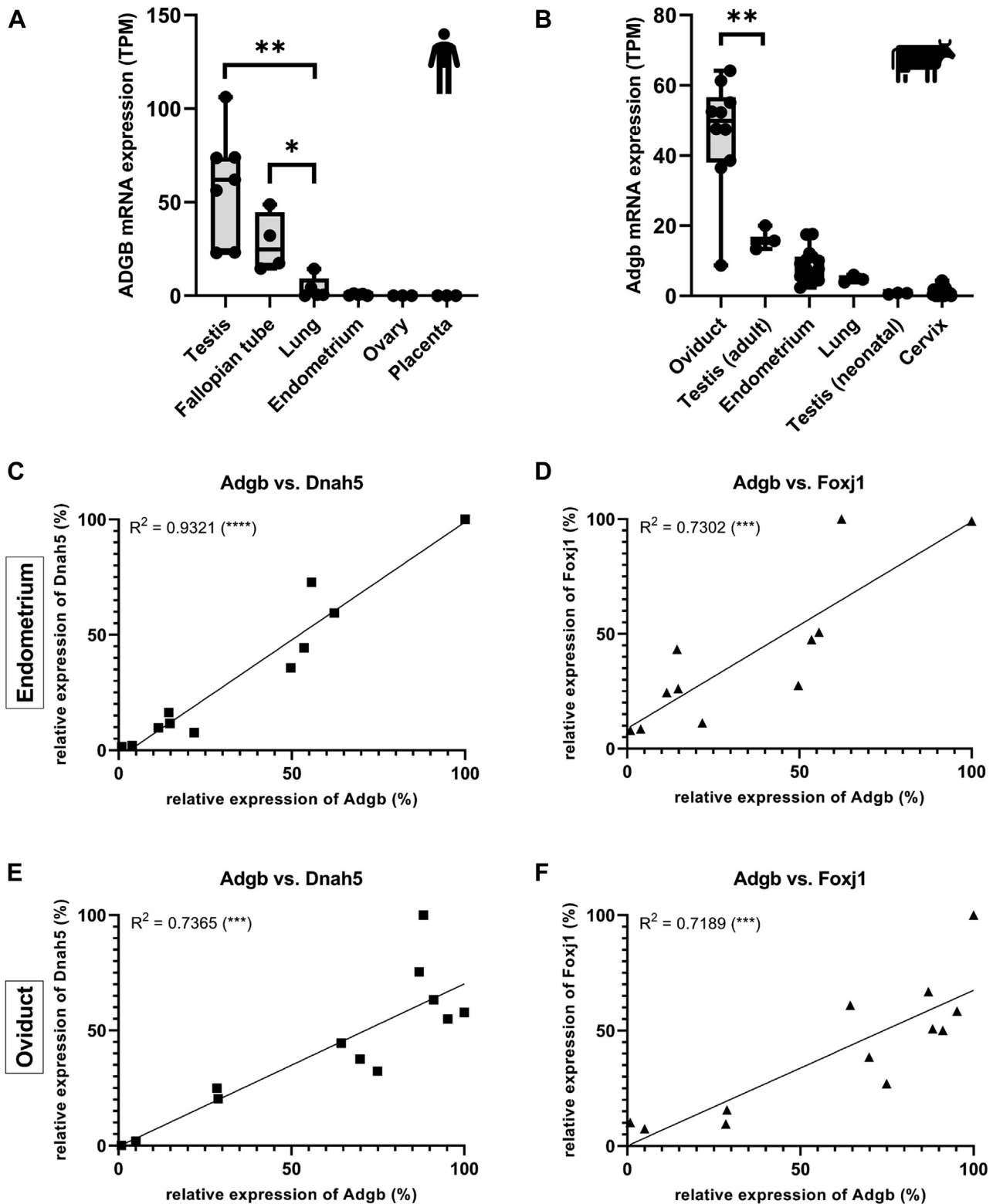


Figure 1. Novel Adgb expression sites correlate with cilia-associated genes. A and B, expression levels of Adgb mRNA in human and bovine tissues as determined by bulk RNA-Seq. A, high levels of expression are found in human testis, but also in fallopian tubes of females. B, transcript levels of Adgb in the oviducts of cattle exceed expression in bovine testis. C–F, correlation analysis of Adgb mRNA expression and mRNA levels of cilia-associated genes Dnah5 (left) and Foxj1 (right). Expression was measured by RT-qPCR analysis in the endometrium (n = 11, upper panels) and oviduct (n = 12, lower panels) of cattle. Adgb shows very strong correlation with Dnah5 ($R^2 = 0.93$) in the endometria (A), strong correlation ($R^2 = 0.73$) with Dnah5 in the oviducts (C), and strong correlations with Foxj1 in the endometria ($R^2 = 0.73$, B) as well as oviducts ($R^2 = 0.72$, D). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

ADGB is involved in ciliogenesis and a target of FOXJ1

Table 1

GO term enrichment analysis of genes coregulated with Adgb in differential gene expression analysis of bovine endometrial samples

Gene set	Description	Enrichment ratio	FDR
Biological process			
GO:0000226	Microtubule cytoskeleton organization	8.37	0
GO:0060271	Cilium assembly	17.66	0
GO:0007018	Microtubule-based movement	16.96	0
GO:0035082	Axoneme assembly	52.00	0
GO:0003341	Cilium movement	45.50	0
GO:0070286	Axonemal dynein complex assembly	58.50	0
GO:0007368	Determination of left/right symmetry	19.31	2.56E-09
GO:0030317	Flagellated sperm motility	15.36	6.96E-05
GO:0001539	Cilium or flagellum-dependent cell motility	29.25	1.11E-04
GO:0042073	Intraciliary transport	24.38	2.97E-04
GO:0044458	Motile cilium assembly	27.00	0.0022
Cellular component			
GO:0015630	Microtubule cytoskeleton	6.16	0
GO:0005929	Cilium	14.60	0
GO:0031514	Motile cilium	27.16	0
GO:0005930	Axoneme	43.15	0
GO:0036126	Sperm flagellum	26.81	0
GO:0097729	9+2 Motile cilium	25.74	6.18E-15
GO:0097223	Sperm part	17.10	3.49E-14
GO:0005815	Microtubule organizing center	6.19	5.86E-12
GO:0044447	Axoneme part	56.57	8.32E-12
GO:0005858	Axonemal dynein complex	68.95	1.69E-09
GO:0005875	Microtubule-associated complex	14.45	7.82E-09
Molecular function			
GO:0003774	Motor activity	15.80	5.35E-06
GO:0003777	Microtubule motor activity	21.23	1.81E-05
GO:0015631	Tubulin binding	7.11	4.41E-05
GO:1990939	ATP-dependent microtubule motor activity	30.79	0.0021
GO:0008017	Microtubule binding	6.73	0.0061
GO:0008092	Cytoskeletal protein binding	3.23	0.0087
GO:0045504	Dynein heavy chain binding	37.53	0.0087
GO:0045503	Dynein light chain binding	37.53	0.0087

All GO terms show a strong connection to the motile cilium. The associated gene list and the full list of enriched terms are provided in [Supplemental File 1](#).

prove that this method was sensitive enough to detect Adgb mRNA expression, we first reanalyzed single-cell RNA-Seq data from murine testis (17). We could show that, in accordance with Hoogewijs *et al.* (5), Adgb mRNA expression was

restricted to later stages of spermatogenesis, where round spermatids differentiate into elongating spermatids and form the flagellum, a motile microtubular structure very similar to a motile cilium. Fully differentiated condensed spermatids,

Table 2

Enriched GO terms in genes showing correlation with ADGB expression in human fallopian tube samples

Gene set	Description	Enrichment ratio	FDR
Biological process			
GO:0007017	Microtubule-based process	6.02	1.28E-04
GO:0035082	Axoneme assembly	28.76	1.96E-04
GO:0003341	Cilium movement	28.30	1.96E-04
GO:0007018	Microtubule-based movement	9.71	5.76E-04
GO:0001578	Microtubule bundle formation	19.71	8.64E-04
GO:0060271	Cilium assembly	6.70	0.0088
GO:0000226	Microtubule cytoskeleton organization	5.41	0.0146
GO:0007368	Determination of left/right symmetry	12.71	0.0439
GO:0060487	Lung epithelial cell differentiation	41.77	0.0439
GO:0009855	Determination of bilateral symmetry	11.89	0.0452
Cellular component			
GO:0005929	Cilium	12.87	0
GO:0005930	Axoneme	30.26	3.82E-10
GO:0097014	Ciliary plasm	30	3.82E-10
GO:0031514	Motile cilium	19.38	2.28E-08
GO:0005858	Axonemal dynein complex	76.67	2.73E-05
GO:0005875	Microtubule-associated complex	16.32	2.86E-05
GO:0005874	Microtubule	6.01	0.0109
GO:0097729	9+2 Motile cilium	14.68	0.0109
GO:0036157	Outer dynein arm	76.67	0.0191
GO:0070160	Tight junction	10.95	0.0284
Molecular function			
GO:1990939	ATP-dependent microtubule motor activity	45.47	6.49E-06
GO:0045503	Dynein light chain binding	66.68	9.42E-06
GO:0003777	Microtubule motor activity	24.10	1.06E-04
GO:0003774	Motor activity	14.82	0.0014
GO:0051959	Dynein light intermediate chain binding	34.49	0.0228
GO:0045505	Dynein intermediate chain binding	33.34	0.0228

All GO terms reveal a strong connection to the motile cilium. The associated gene list and the full list of enriched terms are provided in [Supplemental File 2](#).

however, did not express *Adgb* mRNA anymore (Fig. S3). Following this proof of principle, we performed clustering analysis on single-cell RNA-Seq data sets from epithelial fractions of murine lungs (dataset from Montoro *et al.* (16)). This revealed a distinct entity of lung cells expressing *Adgb*. Using known cell-type markers from literature, and in accordance with our original report, we identified these cells as being multiciliated (Fig. 2). As we had observed in the correlation analysis on fallopian tube samples, *Adgb* expression correlated well with *Dnah5* and *Foxj1*, although the overall number of *Adgb*-positive cells was lower. An additional round of clustering of these ciliated cells revealed no subtypes with noticeable differences in *Adgb* expression, so that we assume that *Adgb*-negative ciliated cells are due to dropout artifacts because of rather low endogenous levels of *Adgb* mRNA (Fig. S4). No *Adgb* expression was observed in progenitors of multiciliated cells, such as basal cells (Fig. 2). Cell subcluster 4 (Fig. S4) showed slightly lower levels of both, *Adgb* and *Foxj1*, but a higher amount of expression of the basal cell marker *Aqp3* (16). This could indicate that *Adgb* expression rises during differentiation and is rather associated with later stages of ciliogenesis or with a maintenance function in cells with already established cilia.

Cells with multiple motile cilia are not only found in the airways and the reproductive tract, but also in the ventricles of the brain, where they maintain proper circulation of cerebrospinal fluid (reviewed in (18)). To obtain further evidence for a functional association of *Adgb* and motile cilia, and looking to explain the previously reported low expression in brain tissue (5), we reanalyzed single-cell RNA-Seq data from mouse brains enriched for ependymal cells and their neuronal progenitors (19). As expected, we could specifically detect *Adgb* mRNA expression in fully mature ependymal cells, although only in a small proportion of cells (Fig. 3). In addition, a subpopulation of tanycytes (designated as “2”) showed a moderate amount of *Adgb* positive cells. GO term analysis of genes overrepresented in ependymocytes and tanycytes “2” again showed a high amount of cilia-associated genes (Fig. 3D). Further analysis revealed that *Adgb*-positive tanycytes belong to the α -subtype, whereas β -tanycytes were *Adgb*-negative. Although not multiciliated such as lung epithelial cells or ependymocytes, α -tanycytes can be biciliated with the motile 9 + 2 microtubule conformation, whereas β -tanycytes only form 9 + 0 immotile cilia, if any (20). Altogether, these data point at an association of *Adgb* with cilia formation and/or function and a possible regulation by *Foxj1*.

The upstream sequence of the ADGB gene displays promoter activity and is inducible by CRISPRa

Gene expression is determined to a great extent by epigenetics and regulatory elements at promoters. As information on this for *ADGB* is scarce, we first inspected data derived from the ENCODE consortium. ENCODE data illustrate that the upstream region surrounding the *ADGB* first exon displays strong DNase hypersensitivity, enrichment of the promoter histone mark H3K4me3, and substantial transcription factor

occupancy, all indicating chromatin accessibility and suggesting putative promoter activity (Fig. S5). Furthermore, chromatin segmentation states coupled to HMM motifs suggest promoter activity of this region in six different cell lines. Additional analysis of epigenetic modifications typical of active chromatin regions showed that H3K4me3 was also enriched at this region in multiple additional cell lines. (Fig. S5). This epigenetic profile reflecting open chromatin is in striking contrast to the rather limited, cell-type-specific expression of *ADGB* (see Discussion).

To experimentally explore the basal activity of the putative human *ADGB* promoter, several potential promoter fragments (431 bp, 1031 bp, and 1981 bp long and starting at –33 bp upstream of the transcriptional start site—TSS) were cloned in a pGL3-luciferase basic vector (Fig. 4A). Reporter gene assays were performed in three cell lines able to form cilia (21–23) and displaying reasonable mRNA expression levels of *FOXJ1* and *RFX2*. Following transfection in HeLa and MCF-7 cells, moderate but consistent basal promoter activity could be observed (Fig. 4A). No substantial changes were seen in HEK293 cells. Based on screening of ENCODE-integrated ChIP-sequencing data for candidate promoter regulating factors, these vectors were cotransfected in HeLa cells and consistently increased *ADGB* promoter-dependent luciferase activity (Fig. S6A). Additional cotransfection experiments with increasing amounts of GATA-3 encoding plasmids indicated GATA-3-dependent regulation of the *ADGB* promoter in a dose-dependent way (Fig. S6B). Next, we employed CRISPR activation (CRISPRa) technology to activate transcription at the *ADGB* promoter. CRISPRa is based on a fusion of catalytically inactive Cas9 (dCas9) with the activation domains of three potent transcription factors, VP64, p65, and Rta (dCas9-VP64), which is targeted to a specific genomic region with single guide RNA (sgRNA) to trigger locus-specific transcriptional activation (24). Several gRNAs, designed to bind upstream of the *ADGB* TSS region, were tested for their capacity to induce *ADGB* promoter-driven luciferase activity and endogenous *ADGB* expression. Using two gRNA sequences (termed gRNA AP-1 and gRNA AP-2), the CRISPR-based system was able to substantially induce *ADGB* promoter-driven luciferase activity in HEK293 and MCF-7 cells, validating functionality of these gRNAs (Fig. 4B). Similarly, the CRISPR-based system also robustly activated endogenous *ADGB* gene expression on mRNA level in both cell lines (Fig. 4C). Interestingly, combined transfection of gRNA AP-1 and gRNA AP-2 additively facilitated expression of the *ADGB* gene. On the protein level, a band could be observed of slightly lower molecular weight compared with predicted endogenous full-length *ADGB* in HEK293 (Fig. 4D). Immunoblotting experiments displayed similar results in MCF-7 (Fig. 4D). Taken together these results confirm that the upstream *ADGB* gene region possesses promoter activity.

The ADGB locus contains functional enhancers

The cell-type-specific regulation of *ADGB* expression is likely to be under the control of multiple *cis*-regulator

ADGB is involved in ciliogenesis and a target of FOXJ1

elements apart from the promoter alone. To further delineate the *ADGB* regulatory landscape, we mined ENCODE and ReMap-based data (25, 26) within the large *ADGB* locus. Multiple regions with strong transcription factor occupancy and DNase hypersensitivity are detectable within the *ADGB* locus (Fig. S7). Furthermore, GeneHancer-derived data suggest that the *ADGB* promoter is regulated by distal enhancer elements that come in close proximity with the promoter by long-range chromatin looping. More precisely seven different potential enhancers (GH06J146620, GH06J146700, GH06J146770, GH06J146808, GH06J146812, GH06J146815, GH06J146819) display looping to the *ADGB* promoter, based on correlations between epigenetic marks and the gene-enhancer distance algorithm implemented by the GeneHancer database (27). Five of these potential *ADGB* enhancer elements are situated within different introns of the *ADGB* gene and two are located immediately downstream of the last *ADGB* exon (exon 36) (Fig. 5A). All of them coincide with strong DNase hypersensitivity and substantial transcription factor occupancy, as well as frequent or occasional (depending on the enhancer) enrichment of enhancer histone marks (H3K4me1, H3K4me2, H3K27ac, and H3K9ac) in multiple mammalian cell lines (Fig. S8). Moreover, chromatin segmentation state tools suggest activity of all enhancers in several cell lines (Fig. S8). In order to experimentally investigate their functionality, we first analyzed their ability to drive SV40 promoter-dependent luciferase activity and cloned all seven potential enhancer elements (for convenience renamed as *ADGB* enhancers (AE) based on intronic or 3' position in: Int1-AE, Int12-AE, Int29-AE, Int35-AE1, Int35-AE2, 3'-AE1, and 3'-AE2) in a pGL3Prom system (Fig. 5B). Reporter gene assays in MCF-7 cells displayed enhancing effects on the SV40 promoter in the presence of Int35-AE1, 3'-AE1, and Int12-AE (Fig. 5B), indicating that these DNA segments possess promoter-enhancing capability. Subsequently, all potential enhancer elements were cloned in the presence of the endogenous *ADGB* promoter (–1 to –464 bp upstream of the *ADGB* TSS). Corresponding with the SV40 promoter-driven luciferase assays, Int35-AE1 and 3'-AE1, but not Int12-AE, increased *ADGB* promoter-driven luciferase activity (Fig. 5C). 3'-AE1 displayed a more profound enhancing effect than Int35-AE1, and the effect of Int35-AE1 diminished to basal levels when the experiment was carried out in HeLa and HEK293 cells (Fig. 5C). Intriguingly, Int35-AE1 and 3'-AE1 display substantial sequence similarity (~63% identity), with the entire 3'-AE1 sequence found within Int35-AE1, with some differences indicative of insertional or substitutional mutations (data not shown). Finally, we employed CRISPRa technology to activate transcription at the 3'-AE1 enhancer and validated enhancer capacities of 3'-AE1 in an endogenous context. Whereas gRNA-3'-AE1 could modestly induce 3'-AE1-dependent *ADGB* promoter-driven luciferase activity (Fig. 5D), gRNA-3'-AE1 also robustly enhanced endogenous *ADGB* mRNA levels (Fig. 5E), albeit to a considerably lower extent as compared with those targeting the *ADGB* promoter. Importantly, dose-dependent overexpression of the 3'-AE1 enhancer targeting gRNA increased endogenous *ADGB*

mRNA levels accordingly. Collectively, these data indicate a complex transcriptional regulation of the *ADGB* locus.

FOXJ1 activates the ADGB promoter via direct binding

The *ADGB* expression data described above suggested a regulation of the gene by FOXJ1, an essential transcriptional regulator of motile cilia formation. To investigate the potential activation of the *ADGB* promoter by FOXJ1, we employed reporter gene assays on cloned *ADGB* promoter of varying lengths. Overexpression of FOXJ1 significantly increased *ADGB* promoter-driven luciferase activity in MCF-7, HeLa, and HEK293 cells, substantiating that FOXJ1 represents an *ADGB* promoter-targeting transcription factor (Fig. 6A). As FOXJ1-mediated activation of the *ADGB* promoter was observed in promoter segments of different but overlapping lengths, the binding site of FOXJ1 might be situated in the smallest –33 to –464 bp region, present in all three of the cloned *ADGB* promoter constructs, while the presence of multiple interaction sites along the longest –33 to –2014 bp fragment cannot be excluded. To analyze FOXJ1-DNA binding to the endogenous *ADGB* promoter region, ChIP assays were performed using anti-FLAG and anti-FOXJ1 antibodies in HEK293 cells transiently transfected with FLAG-tagged FOXJ1 constructs. To control for FOXJ1 overexpression, we analyzed endogenous mRNA levels of four established FOXJ1 target genes (Fig. 6B). Whereas FOXJ1 expression levels were strongly upregulated, also mRNA levels of its target genes DNAAF1, TEK1, CCDC151, and DNAL1 were robustly induced following transient transfection of FOXJ1. Similarly, immunoblotting confirmed expression of the chimeric protein using anti-FLAG and anti-FOXJ1 antibodies (Fig. 6B). Quantitative ChIP analysis revealed more than tenfold FOXJ1 enrichment at the endogenous promoter region compared with the IgG control using two primer pairs spanning the upstream proximal *ADGB* region (Fig. 6C), but not at two more distal upstream and downstream regions, neither at an independent region on chromosome 7 (Fig. 6C). Consistently, no binding was observed in nontransfected cells (Fig. S9). These results confirm that FOXJ1 also binds to the endogenous *ADGB* promoter.

Evolutionary conserved nucleotides within –71 ± 30 bp upstream of the ADGB gene are required for FOXJ1 binding.

To narrow down the search for the FOXJ1-binding site within the *ADGB* promoter, we further dissected the longest *ADGB* promoter segment (–33 to –2014 bp) into three nonoverlapping segments (Fig. 7A) indicating the absence of FOXJ1-mediated activation in more distal *ADGB* promoter segments. In contrast, FOXJ1 overexpression significantly increased the promoter activity of segment –1 to –464 bp, suggesting that the FOXJ1-binding site is limited to this segment closest to the *ADGB* TSS (Fig. 7A). Further refinement of the FOXJ1 responsive region by dividing the –1 to –464 bp segment into three nonoverlapping segments indicated that only the –1 to 140 bp segment closest to the *ADGB* TSS was highly activated by FOXJ1 (Fig. 7A). Next, we

further trimmed down the length of the -1 to -140 bp segment from both the 5'- and 3'- ends by 10 bp, 20 bp, 30 bp, and 40 bp (Fig. 7B). The incremental reduction of -1 to -140 bp segment from both ends at 10 bp intervals did not abolish the FOXJ1-mediated increase in *ADGB* promoter-driven luciferase activity, although a drop in the luciferase signals could be observed in the smaller *ADGB* promoter fragments of -21 to -120 bp/-31 to -110 bp/-41 to -100 bp (Fig. 7B). These results suggest that the FOXJ1 interaction site remains in all of these segments. When the -1 to -140 bp fragment was divided into two equally long, nonoverlapping parts, FOXJ1-mediated increase in promoter activity was abolished (Fig. 7C). This indicated that the mid-region of -70 bp might be important for FOXJ1 interaction, or the -1 to -70 bp and -71 to -140 bp divided segments each contain part of the FOXJ1 interaction site. Multiple sequence alignments of several vertebrate species based on the MULTIZ algorithm within -71 ± 30 bp upstream of the *ADGB* TSS indicated the presence of evolutionary conserved nucleotides within this segment of the promoter (Fig. 8A). To further fine-map the FOXJ1 binding site, we separately mutated three regions within the -71 ± 30 bp *ADGB* promoter region, one containing a single conserved nucleotide (termed *Cons1*), one region displaying evidence of evolutionary constraint as reflected by phyloP and phastCons scores (termed *Cons2*), and one within the mid-point at -71 bp, as the separation of this region resulted in abolished FOXJ1-mediated activation (Fig. 7C). It was suggested that Fox TF-binding sites are approximately 8 to 10 bp in length (28). Therefore, we mutated these regions by substitution of 5 to 6 residues with tandem A and/or T, which are likely to be sufficient to disrupt potential FOXJ1 binding (Fig. 8A). For *Cons2*, two separate mutants of proximal and distal parts were constructed due to the 12 bp size of the conserved region. Interestingly, mutations on *Cons1* and *Cons2* abolished the FOXJ1-mediated increase in *ADGB* promoter-driven luciferase activity, whereas the mid-point mutation did not. (Fig. 8B). This result suggests that both conserved regions (*Cons1* and *Cons2*) within the promoter might be important for FOXJ1 interaction. Moreover, these conserved regions are probably mutually dependent on each other to mediate FOXJ1 interaction as the absence of either part disrupted FOXJ1-mediated activation on the *ADGB* promoter. This might also explain the abolished FOXJ1 activation in pGL3B-AP70-1 and pGL3B-AP140-71 (Fig. 7C), as both of these interdependent FOXJ1 interaction sites were separated in these constructs. Similar findings were obtained with an *ADGB* promoter fragment of reduced size (Fig. S10). Finally, to independently validate the FOXJ1-binding site endogenously, we employed the CRISPR/dCas9 approach with *ADGB* promoter gRNAs to block the genomic-binding site in the presence of exogenous FOXJ1. The docking of a dCas9 variant uncoupled from VPR onto the putative FOXJ1-binding site hinders the interaction of FOXJ1 with the *ADGB* promoter. Similar to the negative control gRNA cotransfection of two more remotely located gRNAs (-590 and -119 bp upstream of the *ADGB* TSS) had no effect on the FOXJ1-dependent activation of *ADGB* expression. In contrast, transfection of three

different *Cons2*-overlapping gRNAs as well as a gRNA 8 bp upstream of *Cons2* all significantly reduced FOXJ1-mediated increase of *ADGB* transcription in both HEK293 and MCF-7 cells (Fig. 8C). Taken together, these data strongly indicate that the FOXJ1 interaction site is located within -71 ± 30 bp upstream of the *ADGB* TSS and involves two evolutionary conserved regions.

Overexpression of FOXJ1 and RFX2 induces endogenous *ADGB* mRNA levels

To investigate the effect of FOXJ1 in the regulation of *ADGB* expression, we overexpressed this transcription factor in HEK293 cells, expressing no endogenous *ADGB* as well as very little endogenous FOXJ1, and subsequently measured endogenous *ADGB* mRNA expression levels. Overexpression of FOXJ1 in HEK293 cells profoundly increased expression levels of endogenous *ADGB* (Fig. 9A), further confirming a FOXJ1-dependent regulation of *ADGB* expression. Given the cooperative functional association between FOXJ1 and RFX2, another essential transcriptional regulator of ciliogenesis (28), we also assessed a potential RFX2-dependent regulation of *ADGB*. Consistent with the RNA-Seq results of *Rfx2*-deficient mice suggesting a RFX2-dependent regulation of *ADGB* transcription (29), also overexpression of RFX2 in HEK293 cells increased endogenous *ADGB* expression levels (Fig. 9A), albeit to a lower extent than FOXJ1. FLAG-tag-based immunoblotting experiments excluded that this discrepancy arose from differences in plasmid expression (Fig. 9B). As the 36-exon containing *ADGB* pre-mRNA might be alternatively spliced to produce different variants of the protein, we confirmed these results by employing multiple exon-exon primer pairs for RT-qPCR across the whole *ADGB* transcript (Fig. S11). Similar results were obtained in MCF-7 cells (Fig. 9A). Collectively, these findings indicate that *ADGB* is a downstream effector of the two master regulators of ciliogenesis FOXJ1 and RFX2, further suggesting a potential role of *ADGB* in the formation and/or function of cilia.

FOXJ1 and RFX2 synergistically activate the *ADGB* promoter in the presence of the 3'-AE1 enhancer in reporter assays

In order to understand the role of RFX2 in regulating the *cis*-regulatory elements of the *ADGB* gene, we employed reporter gene assays to elucidate the interaction of RFX2 on these regulatory elements. Subsequently, we postulated that RFX2 might be binding to the enhancer elements that are in close proximity with the promoter. In order to exclude the possible influence of the endogenous *ADGB* promoter on the readout, we co-overexpressed RFX2 with heterologous SV40 promoter-driven luciferase reporter constructs coupled with *ADGB* enhancers. Our results displayed no RFX2-mediated activation of promoter activity despite the presence of enhancer elements (Fig. S12), suggesting that RFX2 might not interact with any of these enhancer elements in a direct manner. Next, we examined the role of RFX2-mediated activation of *ADGB* regulatory elements in the presence of FOXJ1, *ADGB* promoter and enhancer elements. Reporter assays

ADGB is involved in ciliogenesis and a target of FOXJ1

illustrated no difference in *ADGB* promoter activity across all coupled enhancers with RFX2 overexpression alone (Fig. S12). However, under FOXJ1 overexpression conditions, FOXJ1-mediated activation in each of the *ADGB* promoter-driven reporter constructs could be observed. Interestingly, co-overexpression of FOXJ1 and RFX2 displayed additive activation of *ADGB* promoter activity only in the presence of 3'-AE1 enhancer element (Fig. 9C), but none of the other *ADGB* enhancers (Fig. S12), indicating that the presence of FOXJ1 and the 3'-AE1 enhancer is a prerequisite for RFX2-mediated activation of the *ADGB* promoter. Similar experiments in two independent cell lines, HEK293 and HeLa, validated the 3'-AE1 enhancer-dependent regulation. Whereas this regulation was additive in HEK293, a synergistic regulation could be observed for HeLa cells (Fig. 9C). As this FOXJ1-RFX2 synergy was not observed in the sole *ADGB* promoter-driven reporter constructs, the 3'-AE1 enhancer is crucial in mediating the synergistic effect. Hence, these experiments strongly suggest that RFX2 supports the FOXJ1-mediated regulation of *ADGB* expression with the presence of the essential 3'-AE1 enhancer.

Ectopic *ADGB* overexpression promotes ciliogenesis

As FOXJ1 and RFX2 represent both critical regulators of ciliogenesis, a potential role of *ADGB* in the formation and/or function of cilia is plausible. To explore a putative role for *ADGB* in ciliogenesis, we examined *ADGB* requirements in cilia formation in cellular models. Due to the lack of cellular models with robust endogenous *ADGB* expression levels, we performed *ADGB* overexpression. Immunostaining with anti-acetylated tubulin revealed that the number of cilia was substantially increased following *ADGB* overexpression in human HeLa cells (Fig. 10A). Cilia formation was similarly increased in ciliated mouse cortical collecting duct cells following overexpression of *ADGB*, almost comparable with serum-starved induction of ciliogenesis (Fig. 10B). These findings are in perfect agreement with the expression analyses presented above and collectively suggest that *ADGB* is associated with ciliogenesis and could play an evolutionarily conserved role in the formation and/or maintenance of cilia.

Discussion

ADGB, the fifth member of the mammalian globin family (5), is a chimeric protein with an unusual, embedded globin domain that is circularly permuted and exhibits hallmarks of a hexacoordinated heme-binding scheme (30). Intriguingly, abundant expression of *ADGB* in various species seemed to be restricted to the testis tissue (hence its name) and, more specifically, to postmeiotic stages of spermatogenesis. The function of *ADGB*, however, has remained unclear. Since the gene's initial description (5), a wealth of transcriptome data has been produced by the scientific community, facilitating a re-evaluation of *ADGB*'s expression profile. Our extended expression analysis of *ADGB* in mammalian tissues based on bulk and single-cell RNA-Seq data, including confirmation by RT-qPCR and immunohistochemistry, revealed that *ADGB* is consistently detected in cells carrying motile cilia or flagella.

Specifically, substantial *ADGB* expression was also observed in the *female* reproductive tract, which obviously adds a completely new perspective on its original designation suggesting a predominant role in males. Prominent *ADGB* mRNA expression was also observed in the lung epithelial cells and in the brain ependymocytes and α -tanycytes, altogether possessing motile cilia. A particularly interesting result was the bioinformatically inferred correlation of *ADGB* expression with the master transcription factor of ciliogenesis, Foxj1 (see below). Additional independent support for a ciliogenesis-associated role of *ADGB* came from the CiliaCarta database, a multiomics-based comprehensive ciliary compendium suggesting that *ADGB* is a human ciliome component with a high rank and probability score (31). A recent evolutionary proteomics approach (32) indicated that the association of *ADGB* with ciliary structures may be phylogenetically ancient and can even be detected down to the flagellum-containing choanoflagellates, which is in agreement with our own phylogenetic reconstruction of *ADGB* ancestry (5). In fact, such a phylogenetic perspective lends additional weight to the proposed functional association of *ADGB* and cilia: despite the presence of numerous globin genes in their genomes (33–35), *Drosophila melanogaster* and *Caenorhabditis elegans* both are missing orthologues of *ADGB* (5), and both organisms are devoid of motile cilia on their somatic cells (36). Movement of spermatozoa in *C. elegans* is achieved *via* pseudopods, which carry no resemblance to classical flagella (37). *Drosophila*, on the other hand, still develops motile flagella during spermatogenesis (38). However, the fly harbors two testis-specific globins, which are not phylogenetically related to *ADGB* (39). In expression analyses, these *Drosophila* testis globins were correlated with genes characterized by GO-terms such as sperm axoneme assembly and motility (40). It is therefore tempting to speculate that these globins at least partially compensate for the loss of *ADGB* in the fruit fly. In addition, *ADGB* orthologues appear to be missing in the phylum of crustaceans (5), which form specialized, mostly immobile spermatozoa (41).

The observed strictly cell-specific expression pattern of *ADGB* prompted us to comprehensively investigate its transcriptional regulation. In accordance with a suggested role in ciliogenesis, our experimental data provided direct evidence that the *ADGB* gene is indeed regulated by FOXJ1. For further confirmation, we inspected transcriptome screens of FOXJ1 knockout and overexpression models in mouse, zebrafish, and frogs (42–45) and detected consistent evidence for FOXJ1-dependent expression levels of *ADGB* in those data sets. Furthermore, a recent *in silico* study of FOXJ1-mediated regulatory and signaling networks predicted *ADGB* as one of the direct FOXJ1-regulated genes (46). During spermatogenesis, FOXJ1 expression coincides with the timely stages of flagella formation where it probably orchestrates the expression of genes essential for flagella biogenesis (47, 48). The broader role of FOXJ1 as the master regulator of motile ciliogenesis has been reported as well (10). In addition to FOXJ1-dependent expression, our reporter gene assays and overexpression experiments revealed that the *ADGB* gene is

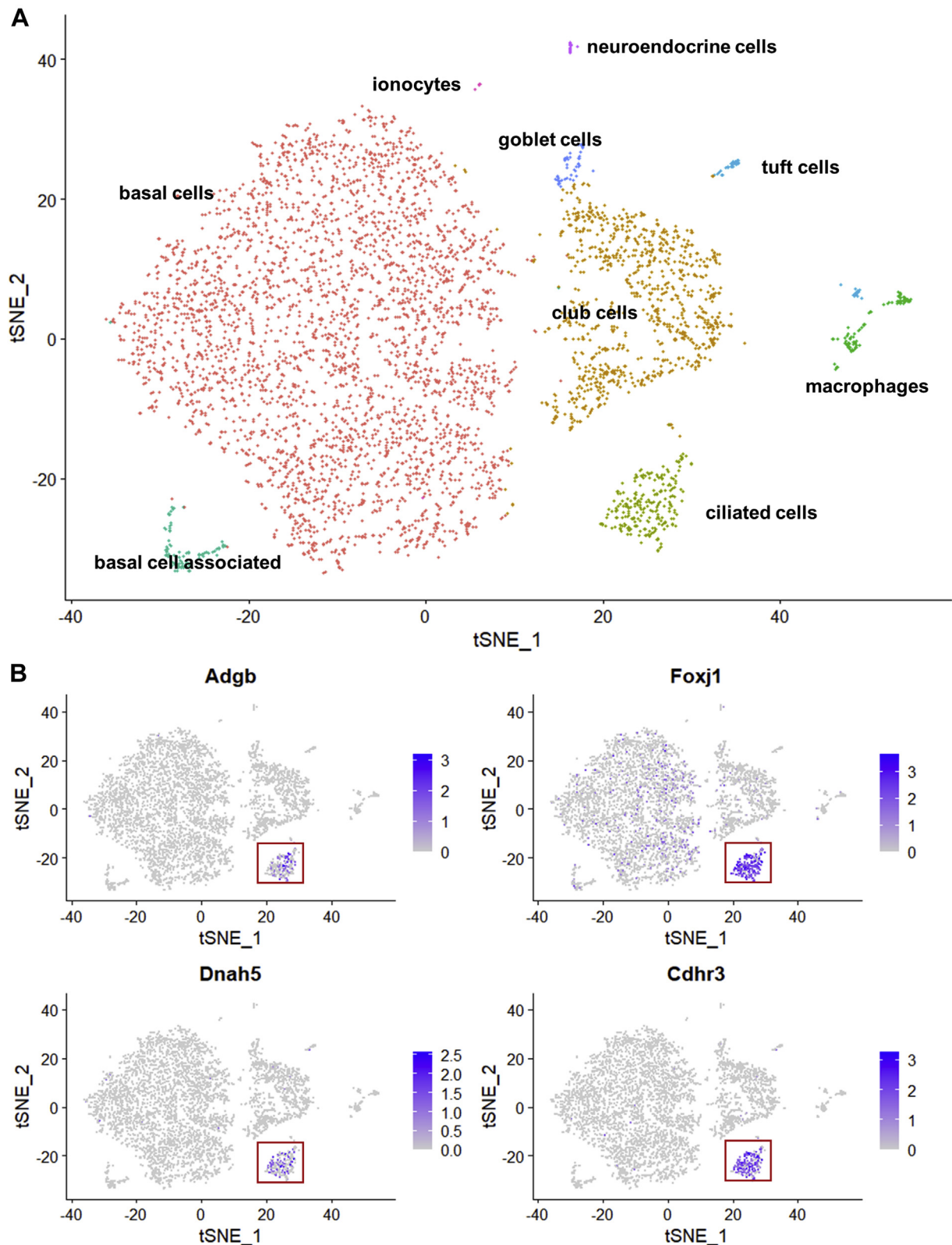


Figure 2. Clustering analysis of single cell RNA-Seq data from murine lungs. A, tSNE representation of cell clusters (named in accordance with (16)). B, visualization of all clusters expressing mRNA of *Adgb* and ciliary marker genes *Dnah5*, *Foxj1* and *Cdhr3*. *Adgb* expression is restricted to ciliated cells in murine lung epithelia.

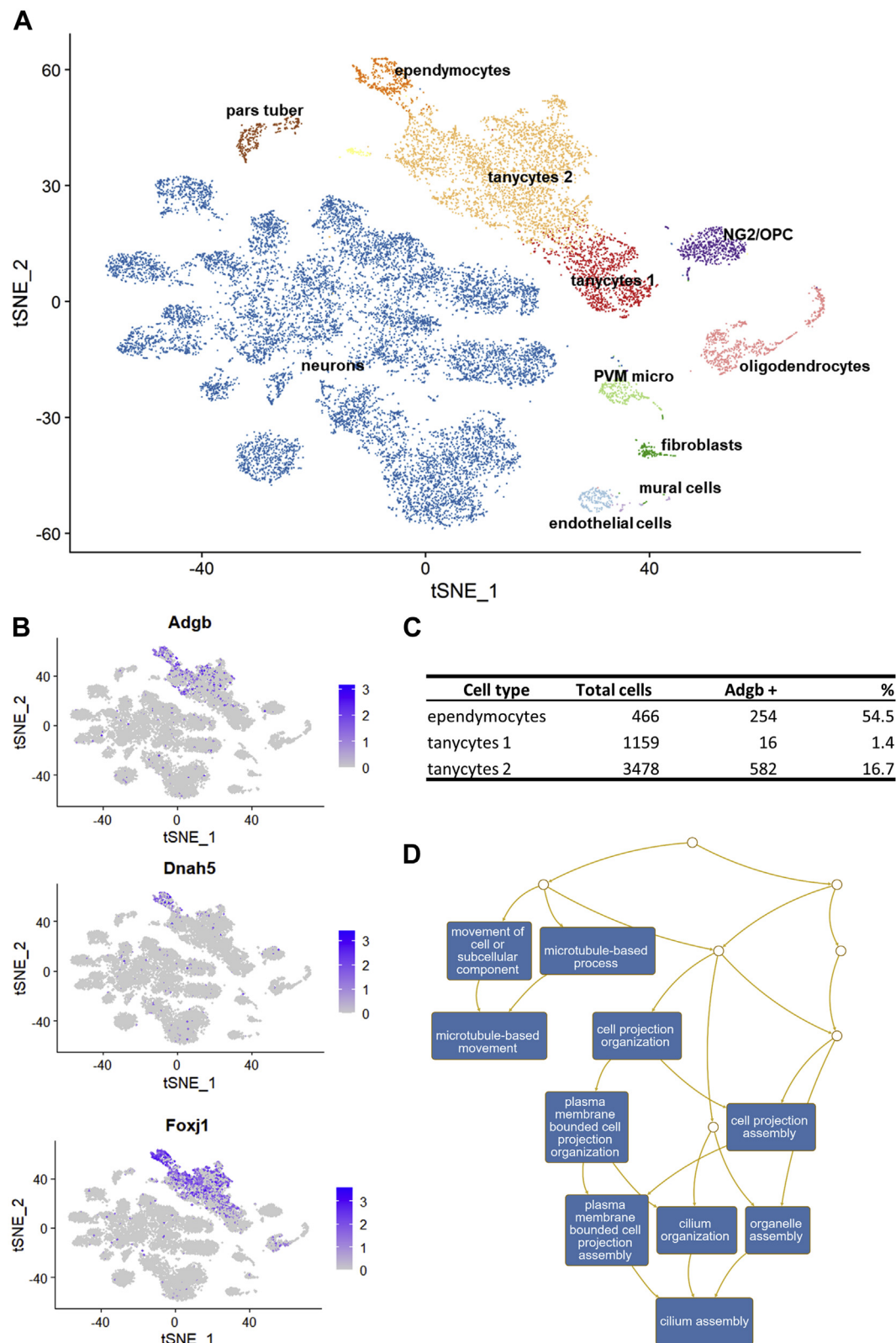


Figure 3. Analysis of single cell RNA-Seq data from mouse hypothalamus. A, tSNE representation of brain cells clustered by levels of expression similarity. Cell types were named in accordance with the initial publication (19). B, mRNA expression levels of *Adgb*, *Dnah5* and *Foxj1*. *Adgb* expression is most prominent in ependymocytes, but also in subpopulation “2” of tanycytes. *Foxj1* expression is also found in these two clusters and absent in tanycytes “1”. *Dnah5* expression is restricted to ependymocytes. C, percentage of *Adgb*-positive cells in ependymocytes and tanycytes subtypes. D, gene ontology analysis of genes overrepresented in ependymocytes and tanycytes “2”. *Adgb* positivity correlates with terms connected to cilia.

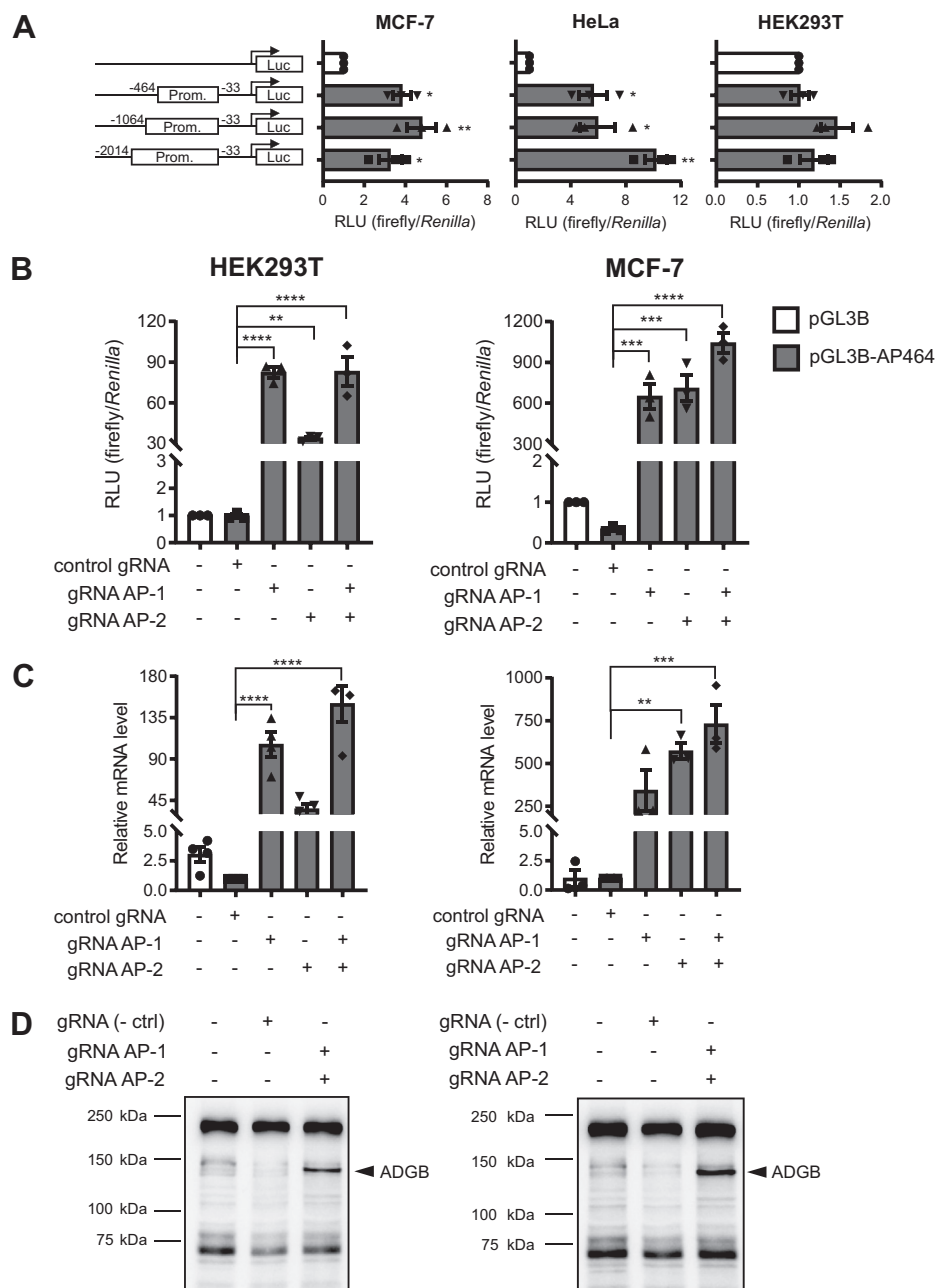


Figure 4. The upstream region of the *ADGB* transcriptional start site contains promoter activity and is inducible by CRISPRa. A, luciferase reporter assays of *ADGB* promoter (AP) elements of three different lengths from –33 bp to –2014 bp, –1064 bp, or –464 bp, respectively, upstream of the *ADGB* TSS in MCF-7 cells, HeLa cells, and HEK293T cells, showing consistent increase in *ADGB* promoter-driven luciferase activity in MCF-7 and HeLa ($n = 3$ independent experiments). Results are displayed as ratios of firefly to *Renilla* luciferase activities in relative light units (RLU) and normalized to results from pGL3-Basic control transfected cells. Schematic representation of cloned fragments upstream of the *ADGB* gene in a pGL3-Basic vector is shown with numbers representing positions corresponding to the first nucleotide of the TSS. B, HEK293T and MCF-7 cells were transfected with dCas9-VPR along with *ADGB* promoter (AP)-targeting gRNAs (gRNA AP-1 and/or gRNA AP-2) and *ADGB* promoter (pGL3B-AP464)-driven luciferase constructs. The gRNA used as negative control contains a nonspecific sequence as present in the pSPgRNA plasmid. Cas9-VPR-based activation of *ADGB* promoter (pGL3B-AP464)-driven luciferase constructs results in activation of the *ADGB* promoter construct ($n = 3$ independent experiments). Results are displayed as ratios of firefly to *Renilla* luciferase activities in RLU. C, HEK293T and MCF-7 cells were transfected with dCas9-VPR along with *ADGB* promoter-targeting gRNA AP-1 and/or gRNA AP-2, and relative *ADGB* transcript levels were quantified by RT-qPCR using a negative control gRNA as reference. Single-guide activation of the *ADGB* promoter with gRNA AP-1 and gRNA AP-2 results in substantial increment in *ADGB* transcript levels ($n = 4$ independent experiments). Simultaneous expression of gRNA AP-1 and gRNA AP-2 leads to synergistic activation of endogenous *ADGB* expression ($n = 4$ independent experiments). D, immunoblotting of immunoprecipitated *ADGB* from HEK293T and MCF-7 cells after gRNAs-dCas9-VPR-activation for 72 h detects endogenous *ADGB* expression. Data represent mean $\pm$ S.E.M (error bars); * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

also robustly regulated by RFX2, a transcriptional activator of spermiogenesis (29). Again, this is in strong agreement with transcriptome data from *Rfx2*-deficient mice (29) showing

Rfx2-dependent expression of *Adgb*. The latter study also listed *Rfx2* binding to the mouse *Adgb* locus in ChIP-sequencing experiments.

ADGB is involved in ciliogenesis and a target of FOXJ1

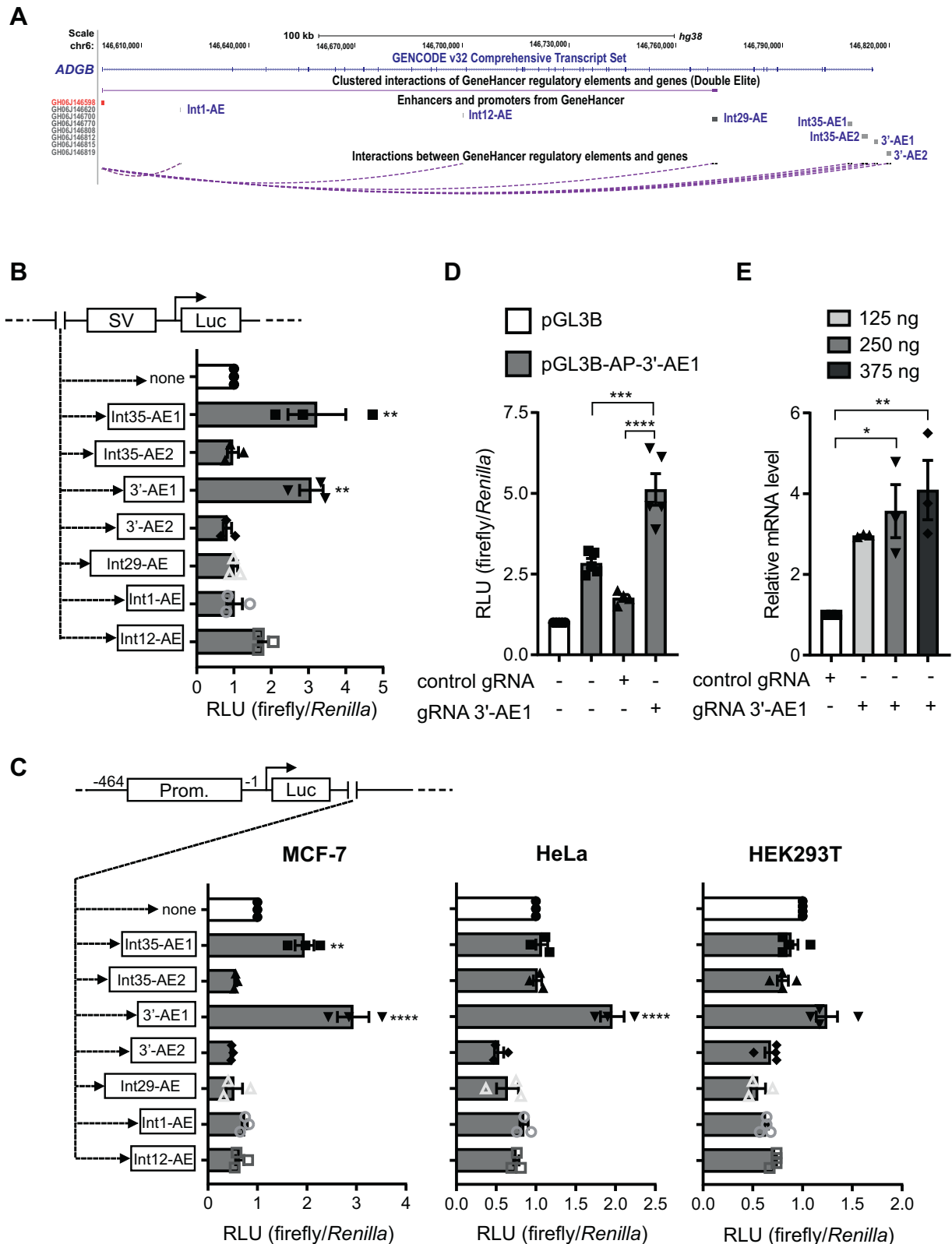


Figure 5. Distal enhancers regulate *ADGB* promoter-driven gene transcription. A, potential *ADGB* regulatory enhancer elements interacting via long-range looping with the *ADGB* promoter region, derived from the GeneHancer database, are displayed in the UCSC Genome Browser. B, seven potential enhancer elements predicted to be in close proximity with the *ADGB* promoter region were cloned at -27 bp upstream of the SV40 promoter in a luciferase vector. These constructs were cotransfected into MCF-7 cells together with a *Renilla* control plasmid, and the effect of these enhancer elements was assessed. Results are displayed as ratios of firefly to *Renilla* luciferase activities in relative light units (RLU). The presence of Int35-AE1, 3'-AE1 and Int12-AE enhancer candidates increased SV40 promoter-driven luciferase activity ($n = 3$ independent experiments). C, the seven potential enhancer elements were cloned at 268 bp downstream of the 464 bp *ADGB* promoter-driven luciferase reporter gene. Consistent with (B), Int35-AE1 and 3'-AE1 sequences, but not the Int12-AE sequence, increased *ADGB* promoter-driven luciferase activity in MCF-7 cells. When tested in HeLa and HEK293T cells, 3'-AE1 sequence also

A recent study reported that the cooperation of FOXJ1 and RFX2 has a prominent role at promoters of ciliary genes compared with other established cilia transcription factors. Both transcription factors were found to be positioned at the anchor end point of chromatin loops, where RFX2 was suggested to act as a scaffolding factor to stabilize the distal enhancer element with the proximal promoter, thus bringing the enhancer-binding FOXJ1 closer to the promoter (28). Consistent with these findings, our study describes that the remote enhancer 3'-AE1, located downstream of the *ADGB* gene, is important for RFX2 to cooperate in synergy with FOXJ1 in activating *ADGB* promoter-driven luciferase activity. It is thus likely that RFX2 acts as the mediator that enables the connection between enhancer 3'-AE1 and the *ADGB* promoter, which could explain the lack of RFX2-mediated transactivation of 3'-AE1-dependent SV40-driven luciferase activity. Correspondingly, RFX2 also shows little to no activation on enhancer 3'-AE1 coupled with the *ADGB* promoter, suggesting a possible scaffolding promoter–enhancer mediator role with no transcriptional activity. It remains to be determined if, in an endogenous genomic context, RFX2 is crucial to establish the connection between the *ADGB* promoter and enhancer 3'-AE1, explaining the modest upregulation of endogenous *ADGB* upon RFX2 overexpression. In slight contrast to the study of Quigley and Kintner (28), our findings from reporter assays indicate a functional interaction of FOXJ1 at the *ADGB* promoter rather than at its distal enhancers. In our reporter gene studies, RFX2 synergistically activates the *ADGB* promoter with FOXJ1 only in the presence of the distal enhancer 3'-AE1, which is in line with Quigley and Kintner (28) and indicates that FOXJ1 is stabilized at promoters of cilia genes through cooperative interactions with RFX2.

Our study provided extensive efforts in refining the interaction site(s) of FOXJ1 on the *ADGB* promoter, which enabled the identification of evolutionarily conserved nucleotides that are crucial for FOXJ1-mediated activation of the *ADGB* promoter. The FOXJ1-binding motif has not been fully annotated so far. A single computational study deduced the preferential binding of FOXJ1 to the consensus sequence NNN[G/A]TAAACAAANN, with N representing any nucleotide (46). However, only a sole motif with this consensus can be found within the –2014 to –1065 bp upstream *ADGB* promoter sequence, whereas motifs with less stringent sequence similarity can be found within –465 to –2014 bp upstream of the *ADGB* TSS. From our experimental data, this part of the *ADGB* promoter shows no FOXJ1-mediated activation. In another study employing *Xenopus laevis*, an RFX-based analysis for binding motifs in the promoters of multi-cilia-related genes has proposed a consensus binding motif in human orthologs (TTCCTGGAAAC). Although this motif was suggested to be the binding site for RFX TFs, also enrichment of

FOXJ1 in this RFX-motif was reported, probably due to cobinding of FOXJ1 and RFX factors (28). Strikingly, this binding motif displays very strong sequence similarity to the *Cons2* region in the *ADGB* promoter whose mutation abolished FOXJ1-mediated activation. Therefore, our conservation-based analysis of FOXJ1 binding on the *ADGB* promoter is independently validated by the *in silico* analysis of FOXJ1-enriched motifs.

The transcription factor p73 plays a major role in ciliogenesis and acts upstream of FOXJ1 and RFX2 (49). Nemajerova *et al.* (50) reported that TP73 deficiency broadly attenuates ciliary gene expression by transcriptome analysis of mouse tracheal epithelial cells (mTEC) derived from WT and TAp73-deficient mice. In line with a role of *ADGB* in ciliogenesis, mTEC air–liquid interfaces (ALI) cultured for 0, 4, 7, and 14 days of differentiation displayed increasingly abundant *ADGB* expression. Moreover, RNA-Seq-based transcriptome analysis of ALI cultured mTECs derived from TAp73-deficient mice showed significantly reduced *ADGB* levels, further substantiating a Tp73-dependent regulation, either directly or more likely *via* its downstream targets *Rfx2*/*Foxj1*, which both displayed downregulated mRNA levels in TAp73 knockout mice. Simultaneously, ChIP-Seq experiments (50) linked p73 directly to FOXJ1/RFX2 and, most interestingly, revealed p73 binding to the distal *ADGB* enhancer 3'-AE1 of the *ADGB* locus. This exquisite dependency of *ADGB* expression on ciliogenesis-associated transcription factors is accompanied by open chromatin marks at the *ADGB* promoter. Surprisingly, this epigenetic feature was also observed in a variety of transcriptionally silent cell types, which—for unknown reasons—may thus contain poised promoters.

In conclusion, our study provides first-time evidence that *ADGB* is specifically expressed in cell types with motile cilia, that its cellular role is most probably associated with cilia biogenesis and function, and that it is a direct regulatory target of FOXJ1 in a complex regulatory landscape. The exact role of *ADGB* in ciliogenesis remains to be established. Future investigations involving the generation of new animal models with conditional knockout of *Adgb* in ciliated tissues will hopefully reveal the intriguing physiological role of *ADGB* in cilia formation and the contribution of FOXJ1- and RFX2-dependent gene regulation.

Experimental procedures

Analysis of bulk RNA sequencing, single-cell RNA sequencing, and microarray data

Publicly available transcriptome raw data (Table S1) were downloaded from either NCBI or ENA web servers (<https://www.ncbi.nlm.nih.gov/sra>; <https://www.ebi.ac.uk/ena>). We

displays an enhancing effect on *ADGB* promoter-driven luciferase activity ($n = 3$ independent experiments). *D*, dCas9-VPR-based activation of *ADGB* enhancer-dependent *ADGB* promoter (464 bp)-driven luciferase constructs using a 3'-AE1 *ADGB* enhancer-targeting gRNA (gRNA-3'-AE1) results in increased luciferase activity. *E*, HEK293T cells were transfected with dCas9-VPR along with different amounts of the 3'-AE1 *ADGB* enhancer-targeting gRNA (gRNA-3'-AE1) and relative *ADGB* transcript levels were quantified by RT-qPCR using the negative control as reference ($n = 3$ independent experiments). Data represent mean $\pm$ S.E.M (error bars); * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

ADGB is involved in ciliogenesis and a target of FOXJ1

only included data from Illumina machines with a minimal read length of 50 nt. For organism-wide gene expression (PRJEB6971 (Hsa; <https://www.proteinatlas.org/>) and PRJNA263600 (Bta)), we focused on data from large sequencing consortia, to ensure comparability. Trimming parameters were assessed for each data set *via* inspection with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter and quality trimming were performed with

BBDuk (<https://sourceforge.net/projects/bbmap/>). We did not apply the same trimming parameters to all the data sets to account for differences in quality and sequencing length; however, differential expression analysis was only performed on data sets from the same study with the same trimming mode. After processing, the reads were mapped against the corresponding reference genomes of either *Homo sapiens* (GRCh 38) or *Bos taurus* (Bta UMD3.1) with HISAT2 (51).

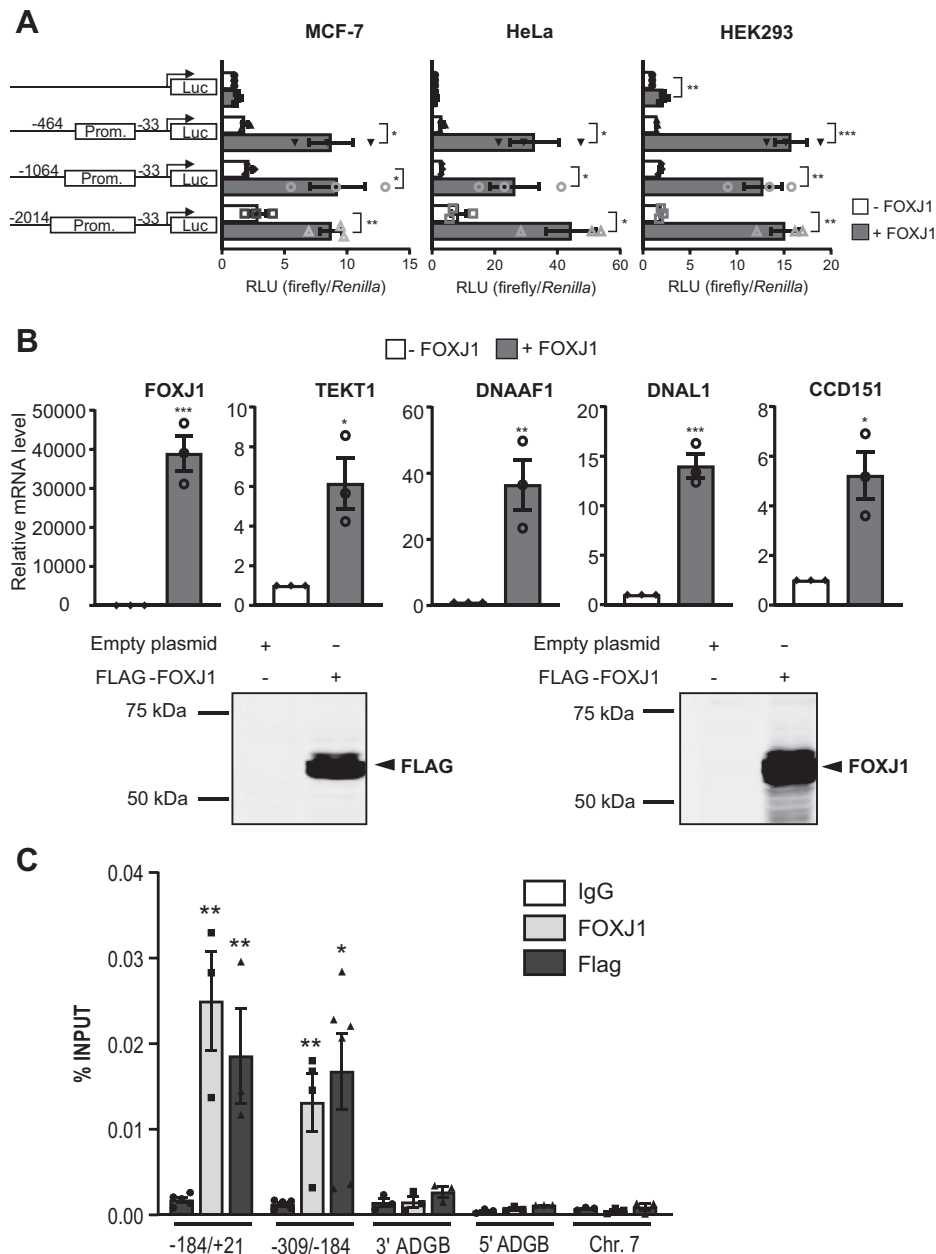


Figure 6. FOXJ1 activates the ADGB promoter via direct binding. A, reporter gene assays of ADGB promoter elements in MCF-7, HeLa, and HEK293T cells with or without co-overexpression of FOXJ1 display a FOXJ1-induced activation of ADGB promoter-driven luciferase activity. This increase in promoter activity is seen in all promoter elements of different lengths (–33 bp to –464, –1064, and –2014 bp upstream of the ADGB TSS). Results are displayed as ratios of firefly to Renilla luciferase activities in relative light units (RLU) and normalized to the pGL3-basic vector (n = 3 independent experiments). B, mRNA and protein experiments of HEK293 cells transiently transfected with a FLAG-FOXJ1 vector. FOXJ1 target gene mRNA levels were measured by RT-qPCR and normalized to β -actin mRNA levels. Immunoblotting using anti-FLAG and anti-FOXJ1 antibodies also confirmed FOXJ1 overexpression. C, the amount of coprecipitated chromatin derived from the proximal ADGB promoter region using two primer pairs (covering +21 to –184 and –184 to –309 upstream of the ADGB TSS, selected based on the reporter gene assays), its upstream (5') and downstream (3') regions as well as an independent region on chromosome 7 in the EPO locus (64), was determined by qPCR. Data represent mean $\pm$ S.E.M (error bars); * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.

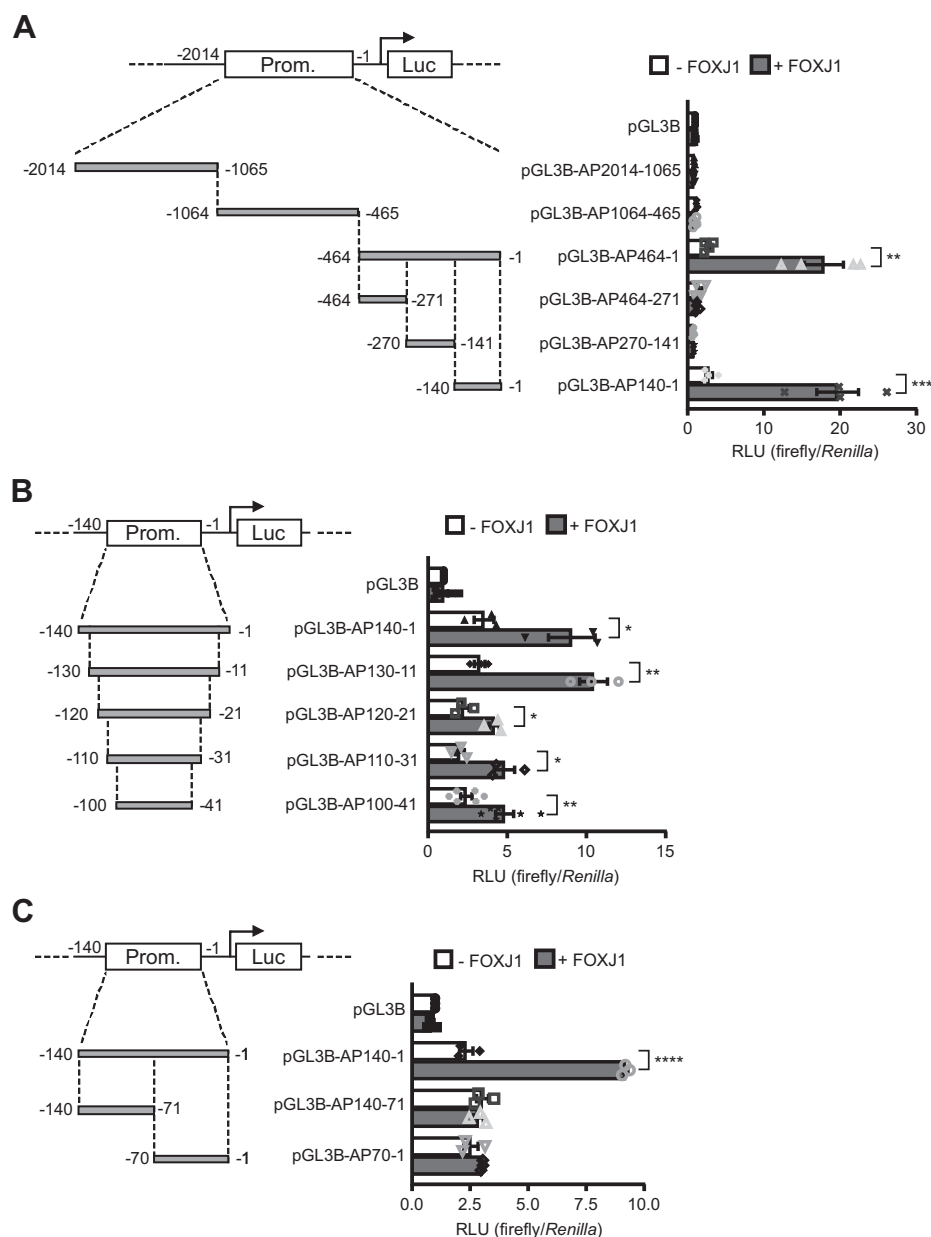


Figure 7. The proximal ADGB promoter confers FOXJ1-mediated increase in luciferase activity. A, the longest ADGB promoter (AP) element, AP2014 (–33 to –2014 bp upstream of TSS), displaying FOXJ1-induced promoter activity was first divided into three nonoverlapping segments (–1065 to –2014 bp, –465 to –1064 bp, and –1 to –464 bp upstream of the ADGB TSS) and cloned individually upstream of the firefly luciferase gene. Only the segment –1 to –464 bp upstream of the ADGB TSS displays FOXJ1-induced promoter activity, and no induction is observed with inserts covering the –465 to –2014 bp upstream of the ADGB TSS (n = 4 independent experiments). Schematic representation of different ADGB promoter segments is shown with numbers representing positions corresponding to the first nucleotide of the TSS. The –1 to –464 bp ADGB promoter construct was further subdivided into three nonoverlapping segments (–271 to –464 bp, –141 to –270 bp, and –1 to –140 bp upstream of the ADGB TSS) and cloned individually upstream of the firefly luciferase gene. From those only insert –1 to –140 bp displays FOXJ1-induced promoter activity (n = 4 independent experiments). B, the –1 to –140 bp ADGB promoter construct was sequentially reduced at both the 5'- and 3'- ends by 10 bp, 20 bp, 30 bp, and 40 bp, to generate smaller segments of –11 to –130 bp/–21 to –120 bp/–31 to –110 bp/–41 to –100 bp upstream of the ADGB TSS. All segments display FOXJ1-induced promoter activity including the smallest 60 bp segment of –41 to –100 bp (n = 3 independent experiments). C, –1 to –140 bp ADGB promoter element was subdivided into two segments of 70 bp, –1 to –70 bp and –71 to –140 bp. In both constructs, FOXJ1-induced promoter activity was abolished (n = 3 independent experiments) suggesting a disruption to the FOXJ1 binding site or missing binding sites for proper FOXJ1 interaction.

Calculation of *transcripts-per-million* values (TPM) was performed with StringTie (52) in guided mode. Differential gene expression analysis (Bta) as well as hierarchical clustering of genes (Hsa) was done in R (<https://www.r-project.org/>) with the help of Bioconductor's DESeq2 package (53). Gene ontology analysis was performed with WebGestalt 2019 (54).

Mapped single-cell data (Table S1) were downloaded from NCBI as UMI count tables. In the lung data set, we only included mice #2 and #3, which share the same genetic background, retaining 5010 out of 7193 cells from the original analysis (16). Graph-based clustering of cells was performed using Seurat 3 (55). Our clusters were mostly in accordance

ADGB is involved in ciliogenesis and a target of FOXJ1

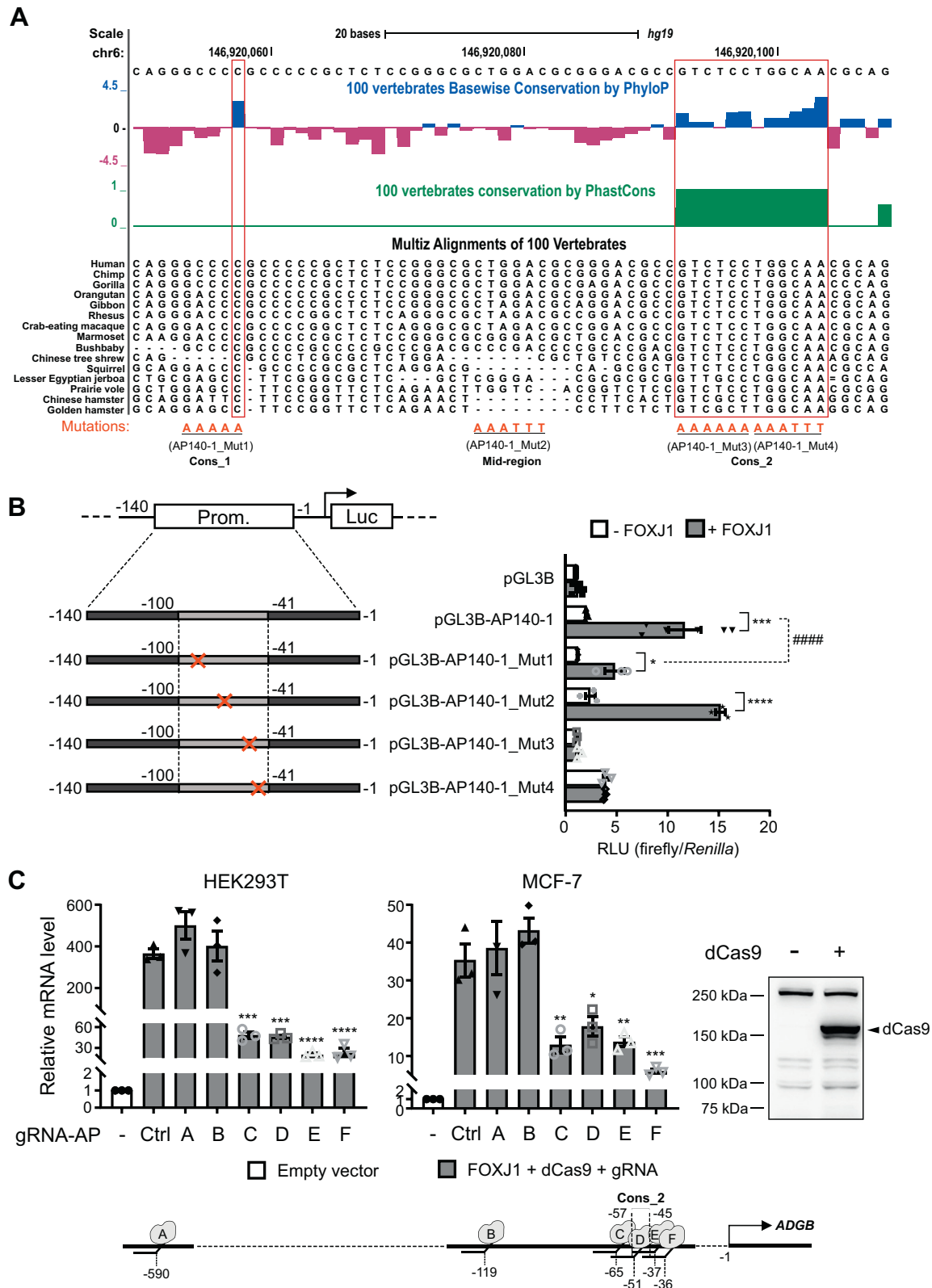


Figure 8. Conserved nucleotides within -71 ± 30 bp upstream of *ADGB* TSS are crucial for FOXJ1 binding. A, UCSC Genome Browser output (*hg19*) of evolutionarily conserved nucleotides within -41 bp to -100 bp upstream of the *ADGB* TSS based on a subset of vertebrate sequences extracted from the 100-MULTIZ whole-genome multiple sequence alignment algorithm. Basewise conservation scored by PhyloP indicates conserved and variable nucleotides in blue and red bars, respectively. Highly conserved nucleotides, also supported by the PhastCons track within this region of the *ADGB* promoter are boxed in red. Mutation strategy of conserved nucleotides is illustrated at the bottom. B, substitution-based mutation at -96 to -92 bp (Mut1), -57 to -52 bp (Mut3), and -51 to -46 bp (Mut4) results in loss of FOXJ1-dependent increase in *ADGB* promoter activity. Whereas the mutation at -73 to -68 bp (Mut2) did not abolish the FOXJ1-mediated activation ($n = 3$ independent experiments). C, sgRNA-mediated docking of dCas9 onto the Cons2 region in the *ADGB* promoter results in reduced FOXJ1-dependent activation of *ADGB* expression while

with the ones published before, although we did not prefilter for contaminating cell types such as macrophages to ensure impartial analysis.

Animals

Cattle tissue was obtained from young females (heifers) immediately after slaughter in a regional commercial slaughterhouse. One uterine horn was opened lengthwise and the endometrial portion of the uterus was dissected. Tissue was either flash-frozen on dry ice (RNA extraction) or fixed in 4% para-formaldehyde (immunofluorescence).

RNA extraction and reverse-transcription quantitative PCR (RT-qPCR)

RNA extraction of cattle tissues was performed from snap-frozen samples with the RNeasy Plus Universal Mini kit (Qiagen) according to the instructions of the manufacturer. Approximately 50 mg of tissues was grinded and homogenized with a MiniLys (Precellys) system using mixed ceramic beads (Precellys Lysing CKMix). Difficult tissues such as endometria were pregrinded manually on dry ice with a cool scalpel. RNA was eluted in nuclease-free water. RNA quality was assessed with a Bioanalyzer (Agilent), and only samples with RIN >7 were used for further analysis. RNA was quantified *via* Qubit measurement using the Broad Range RNA Assay Kit (Thermo Fisher) and was stored at -80°C until further use. To confirm the bioinformatical findings, we performed reverse-transcription quantitative PCR (RT-qPCR) on tissues from the female reproductive tract of cattle. 1000 ng of total cattle RNA per sample was used for reverse transcription with the SuperScript III enzyme (10,000 units per assay; Invitrogen) using an Oligo-dT primer. In the absence of validated reference genes, the amount of mRNA expression was normalized on the adjusted total amount of carefully quantified total RNA. To additionally control for differences in cDNA synthesis, 100 ng of *Drosophila* total RNA was added to the reaction as a spike-in control. RT-qPCR was carried out using GoTaq qPCR Master Mix (Promega) on the ABI Prism 7500 Fast Detection System (SDS, Applied Biosystems) and interpreted using 7500 Software Version 2.3. Quantification of ADGB-cDNA molecules was done in absolute numbers applying a calibration standard curve with known amounts of target PCR product, previously cloned into the pGEM T-easy vector system (Promega). *Foxj1* and *Dnah5* expressions were measured as relative values only and the sample with the highest expression was set to 100%. Copies of the *Drosophila* Globin 1 (Glob1) cDNA of the internal control were measured in parallel to identify samples with substandard reverse transcription. All primers used are listed in Table S2. For HEK293 cells, total RNA was extracted as previously described (56). Total RNA (2 μg) was reverse transcribed (RT) using the Prime Script RT reagent kit (Takara Bio USA) and cDNA levels were estimated by qPCR using the primers listed in Table S2 and a KAPA

SYBR FAST qPCR reagent kit (Sigma-Aldrich) in a CFX96 C1000 Thermal Cycler (BioRad). Transcript levels were calculated as described before (57) and displayed as relative expression levels.

Expression plasmid constructs

pENTR233-FOXJ1 entry clone obtained from the DNASU plasmid repository (58) and pcDNA3.1/nV5-DEST mammalian expression vector were Gateway-recombined according to the manufacturer's instruction (Invitrogen) to generate a pcDNA3.1-nV5-FOXJ1 expression vector. The pcDNA3.1-HA-RFX2 plasmid was a generous gift from Prof. Zijie Sun (Stanford). N-terminally FLAG-tagged FOXJ1 and RFX2 were cloned by amplifying FOXJ1 and RFX2 genes using primer pairs with *SalI/KpnI* and *BglIII/SalI* overhangs, respectively as listed in Table S2. Amplicons were subsequently digested with their respective restriction enzymes as designed on the primers and ligated into linearized pFLAG-CMV-6a vector to generate pFLAG-FOXJ1 and pFLAG-RFX2.

Luciferase constructs

ADGB promoter elements spanning from -33 bp to -464 bp, -1064 bp, and -2014 bp upstream of the ADGB transcriptional start site (TSS) were cloned into a pGL3-Basic (Promega) vector at -67 bp upstream of the firefly luciferase reporter gene. Promoter elements were amplified from a pool of genomic DNA extracted from three human cell lines, MCF-7, HEK293T, and Hep3B cells, by PCR using the primer pairs described in Table S2. PCR amplicons were digested with their respective restriction enzymes and ligated into linearized pGL3-Basic (Promega) vector digested with *KpnI* and *NheI*. Evolutionarily conserved nucleotides within -71 ± 30 bp upstream of the ADGB TSS were mutated into a tandem of 5X or 6X A or T (or both) using oligonucleotide-based cloning of mutant promoter fragments into pGL3-Basic (Promega) vector at -67 bp upstream of a firefly luciferase gene. The wild-type sequences were mutated as indicated in Table S2. Prior to cloning, phosphorylated oligo duplexes with designed 5'-*KpnI* overhang and 3'-*NheI* overhang were generated by incubating 0.5 μM of synthesized complementary oligo strands in T4 DNA ligase buffer (ThermoScientific) (40 mM Tris-HCl, 10 mM MgCl_2 , 10 mM DTT, 500 μM ATP) with T4 polynucleotide kinase (ThermoScientific) at 37°C for 1 h, followed by heating to 95°C for 5 min and slow cooling at the rate of $-5^{\circ}\text{C min}^{-1}$ to 10°C . Oligo duplexes were subsequently ligated into *KpnI* and *NheI* digested pGL3-Basic vector backbone. Potential ADGB intronic and 3' enhancer elements from the GeneHancer database (27) were cloned in a pGL3-SV40 vector (Promega), at -27 bp upstream of the SV40 promoter. These putative enhancer elements were amplified from genomic DNA, by PCR using the primer pairs as described in Table S2. PCR amplicons were digested with their

more remote control sgRNAs display no effect in HEK293 and MCF-7 cells ($n = 3$ independent experiments). gRNA positions are schematically represented. Immunoblotting analysis using a Cas9 antibody controlled for dCas9 overexpression. Data represent mean $\pm$ S.E.M (error bars); * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

ADGB is involved in ciliogenesis and a target of FOXJ1

respective restriction enzymes and ligated into linearized pGL3-SV40 (Promega) vector digested with *KpnI* and *NheI* or into the linearized *ADGB* promoter containing pGL3B-AP464 vector, at 269 bp downstream of the firefly luciferase gene, digested with *BamHI*.

Mammalian cell culture and DNA transfection

MCF-7, HEK293T, and HeLa cells were cultured and maintained in DMEM (Thermo Fisher) medium supplemented with 10% fetal bovine serum (Chemie Brunschwig) and 100 µg/ml Pen/Strep Glutamine (Thermo Fisher), DMEM/FBS/PS for simplification, and incubated at 37 °C in a humidified incubator with 5% CO₂. mCCD_{cl1} cells (59) were maintained at 37 °C and 5% CO₂ in DMEM/F12 (Thermo Fisher) supplemented with 5 µg/ml insulin (Sigma), 50 nM dexamethasone (Sigma), 60 nM selenium (Sigma), 5 µg/ml transferrin (Sigma), 1 nM triiodothyronine (Sigma), 5 ng/ml mouse EGF (Sigma), 100 µg/ml Pen/Strep Glutamine (Thermo Fisher), and 2% decomplexed fetal bovine serum (Thermo Fisher). MCF-7, HEK293T, and HeLa cells were transfected using Roti-Fect (ROTH),

according to manufacturer's instruction. Prior to transfection, cells were seeded on 24-well plates at a density of 5.2×10^4 cells (for MCF-7 and HEK293T) or 2.6×10^4 cells (for HeLa) per cm² of the dish surface area. For calcium phosphate precipitation method of DNA delivery, overnight medium was aspirated and replaced with DMEM/FBS/PS containing 25 µM chloroquine. In each well, 10% (v/v) of transfection mixture was introduced to the cells, consisting of 50% (v/v) 1.2 to 1.5 µg plasmid DNA/250 mM CaCl₂ and 50% (v/v) HBS buffer (50 mM HEPES, 280 mM NaCl, 10 mM KCl, 1.5 mM Na₂HPO₄, 12 mM glucose). Transfected cells were incubated at 37 °C/5% CO₂ for 24 h before the medium was replaced with fresh DMEM/FBS/PS and continued to incubate at 37 °C/5% CO₂ for another 24 h. Cells were harvested 48 h posttransfection. mCCD_{cl1} cells were transfected using Polyplus JetPrime according to manufacturer's instructions. For immunocytochemistry experiments, HeLa and mCCD cells were seeded in 6-well plates on glass coverslips at 0.2×10^6 cells per well. After 24 h, the cells were transfected and either left untreated or serum-starved for another 24 h.

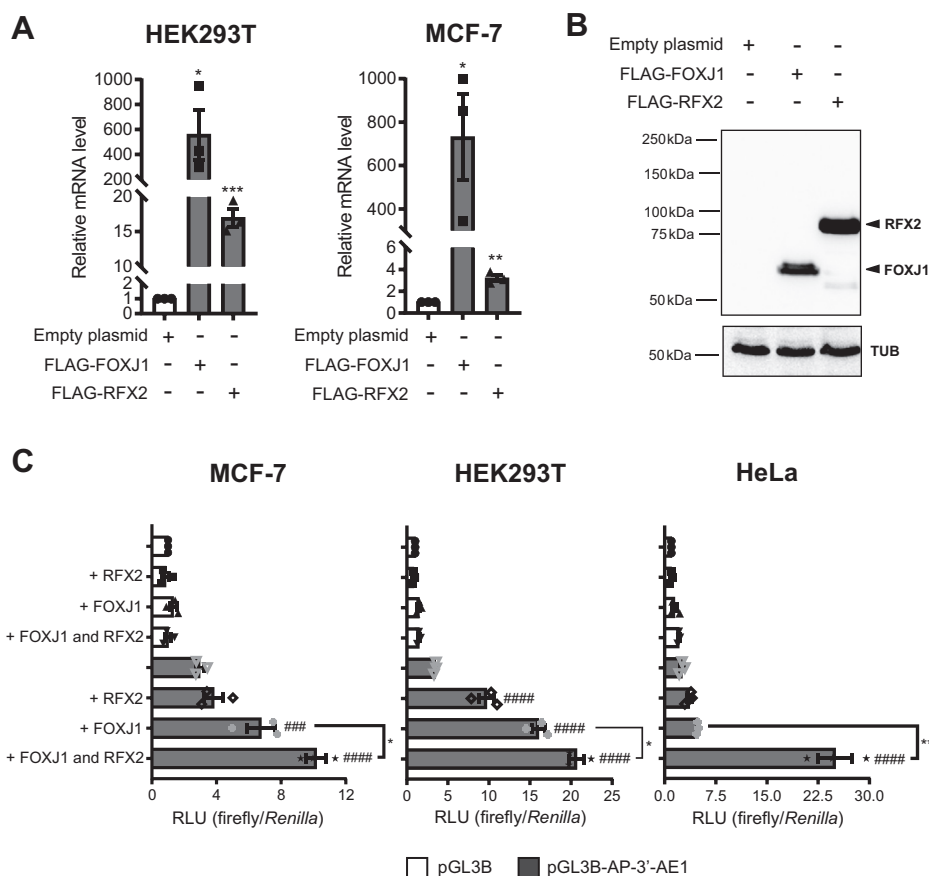


Figure 9. FOXJ1 and RFX2 transcription factors induce endogenous *ADGB* transcription and cooperate to regulate *ADGB* expression in the presence of the 3'-AE1 enhancer element. A, HEK293T and MCF-7 cells were transiently transfected with a FLAG-tagged FOXJ1 or RFX2 expressing plasmid and *ADGB* mRNA levels were measured with RT-qPCR. *ADGB* expression levels were normalized to β -actin and displayed as relative values to cDNA of HEK293T or MCF-7 cells transfected with equal amount of empty vector. B, FLAG-tagged FOXJ1 and RFX2 plasmid-derived expression levels were controlled by immunoblotting using an anti-FLAG antibody. Tubulin (TUB) was used as loading control. C, 3'-AE1 *ADGB* enhancer-dependent *ADGB* promoter-driven reporter gene assays. Following overexpression of FOXJ1, FOXJ1-mediated activation of the *ADGB* promoter activity is observed as expected from previous experiments. Co-overexpression of FOXJ1 and RFX2 shows synergistic activation of the *ADGB* promoter construct in the presence of 3'-AE1 enhancer element (n = 3 independent experiments). Similar experiments were performed in HEK293T and HeLa (n = 3 independent experiments). Data represent mean $\pm$ S.E.M (error bars); **p* < 0.05; ***p* < 0.01; ****p* < 0.001; ####*p* < 0.0001.

Luciferase reporter gene assays

Fifty nanograms of promoter- and/or enhancer-containing firefly luciferase plasmid was cotransfected along with 1 ng of pRL-SV40 *Renilla* luciferase to control for differences in transfection efficiency and extract preparation. For the study of FOXJ1 and RFX2 activity on *ADGB* promoters, 300 ng of pcDNA3.1-nV5-FOXJ1 or pcDNA3.1-HA-RFX2 was cotransfected with promoter plasmid, the total amounts of plasmid DNA used were normalized with pcDNA3.1-nV5-HisA empty vector. Luciferase activities were determined using the Dual Luciferase Reporter Assay System (Promega) as described before (60). Reporter activities were expressed as relative firefly/*Renilla* luciferase activities. All reporter gene assays were performed at least three times independently.

dCas9-VPR-mediated activation of endogenous *ADGB* promoter

Nuclease-null-Cas9 with tandem fusion of VP64-p65-Rta tripartite activator (dCas9-VPR, Addgene #63798) (24) was delivered along with gRNAs as described before (61) to activate *ADGB* promoter activity. gRNAs candidates targeting between -1 and -1700 bp upstream of *ADGB* TSS were cloned into, and expressed from, pSPgRNA plasmid (Addgene #47108), which was a generous gift from Prof. Charles Gersbach (62) (Table S2). Prior to cloning, phosphorylated oligo duplexes were generated by incubating 0.5 μ M of synthesized complementary oligo strands (5'-3') in T4 DNA ligase buffer (ThermoScientific) (40 mM Tris-HCl, 10 mM MgCl₂, 10 mM DTT, 500 μ M ATP) with T4 polynucleotide kinase (ThermoScientific) at 37 °C for 1 h, followed by heating to 95 °C for 5 min and slow cooling at the rate of -5 °C min⁻¹ to 10 °C. Oligo duplexes with sticky ends complementary to *Bbs*I-digested pSPgRNA vector were then ligated into the vector. dCas9-VPR and gRNAs were delivered to HEK293T cells with Roti-Fect (ROTH) according to the manufacturer's instruction. Cells were transfected in 24-well plates, seeded with 5.2 $\times$ 10⁴ cells per cm² surface area, 24 h prior to transfection. In each well, 375 ng of dCas9-VPR was delivered together with 125 ng of gRNA(s) in antibiotic-free DMEM medium. For the transfection of gRNAs in combinations, the total amount of gRNAs was equally distributed to a total of 125 ng. Cells were incubated for 24 h before the transfection medium was replaced with fresh DMEM/FBS/PS and allowed to grow for another 24 h before RNA extraction for analysis. For dCas9-VPR activation of endogenous *ADGB* and subsequent immunoblot analysis, HEK293T cells were seeded on 6-well plates at a density of 5.2 $\times$ 10⁴ cells per cm² of the dish surface area and were allowed to grow at 37 °C/5% CO₂ for 24 h. For transfection, 2250 ng of dCas9-VPR and 750 ng of gRNAs mix (equal amount of gRNA AP-1 and gRNA AP-2) were delivered to HEK293T cells with Roti-Fect (ROTH), and cells were incubated for 24 h before the transfection medium was replaced with fresh DMEM/FBS/PS and allowed to grow for another 48 h.

dCas9-mediated interference of FOXJ1 binding

Nuclease-null-Cas9 was cloned by amplifying the dCas9 gene using primer pairs with *Kpn*I and *Not*I overhangs listed in Table S2. The amplicon was subsequently digested with restriction enzymes as designed on the primers and ligated into a linearized pcDNA3 vector to generate pcDNA3-dCas9. Candidate gRNAs targeting the *Cons2* region within the *ADGB* promoter and control gRNAs targeting more distal regions (Table S2) were cloned into and expressed from a pSPgRNA plasmid (Addgene #47108) as described above. For dCas9-mediated interference of FOXJ1 binding, 800 ng of pFLAG-FOXJ1 was cotransfected with 1000 ng of pcDNA3-dCas9 and 600 ng of gRNA into HEK293T or MCF-7 cells with calcium phosphate precipitation and Roti-Fect (ROTH), respectively. Cells were incubated for 24 h before the transfection medium was replaced with fresh DMEM/FBS/PS, and allowed to grow for another 24 h.

Chromatin immunoprecipitation

ChIP was carried out as described before with some modifications (63, 64). Briefly, cells were cross-linked by adding 1% (w/v) formaldehyde and incubated for 20 min at RT with gentle shaking. Cell fixation was interrupted by adding 110 mM glycine. Cells were scraped off and resuspended in lysis buffer following the iDeal ChIP-qPCR kit protocol (Diagenode, Liège, Belgium). To obtain genomic DNA fragments between 500 and 100 bp, cell lysates were sonicated for four rounds of ten cycles (30 s ON/30 s OFF) using the Bioruptor Pico (Diagenode) at high power setting. For immunoprecipitations, the following antibodies were used: 1 μ g of rabbit polyclonal anti-IgG (C15410206, Diagenode) as negative control IP; 1 μ g of mouse monoclonal anti-FOXJ1 (14-9965-82, Thermo Fisher Scientific) and 3.8 μ g of mouse monoclonal anti-FLAG (F1804, Sigma Aldrich). Chromatin-antibody complexes were immunoprecipitated by DiaMag Protein A-coated magnetic beads (Diagenode). DNA isolation and de-cross-linking was carried out as described by the iDeal ChIP-qPCR kit protocol (Diagenode). Coprecipitated DNA was quantified by real-time qPCR using the primers listed in Table S2.

Immunocytochemistry and cilia counting

The cells were washed with phosphate-buffered saline (PBS) and fixed for 10 min in 4% para-formaldehyde (PFA), followed by 3 $\times$ 5 min PBS washes, permeabilization in PBS/Triton X-100 0.2% for 10 min, 3 $\times$ 5 min PBS washes, and blocking in PBS/BSA 1% for 1 h. The cells were then incubated overnight with mouse anti-acetylated tubulin (Santa Cruz) diluted 1/500 in PBS/BSA 0.1%. On the following day, cells were washed 3 $\times$ 5 min with PBS and incubated for 1 h with secondary Alexa Fluor goat anti-mouse IgG (Invitrogen) diluted 1/300 in PBS. The slides were mounted with Fluoromount mounting medium containing DAPI (Southern Biotech) and visualized on a Nikon Eclipse fluorescent microscope (Nikon Corporation). Cilia counting was performed using ImageJ software.

ADGB is involved in ciliogenesis and a target of FOXJ1

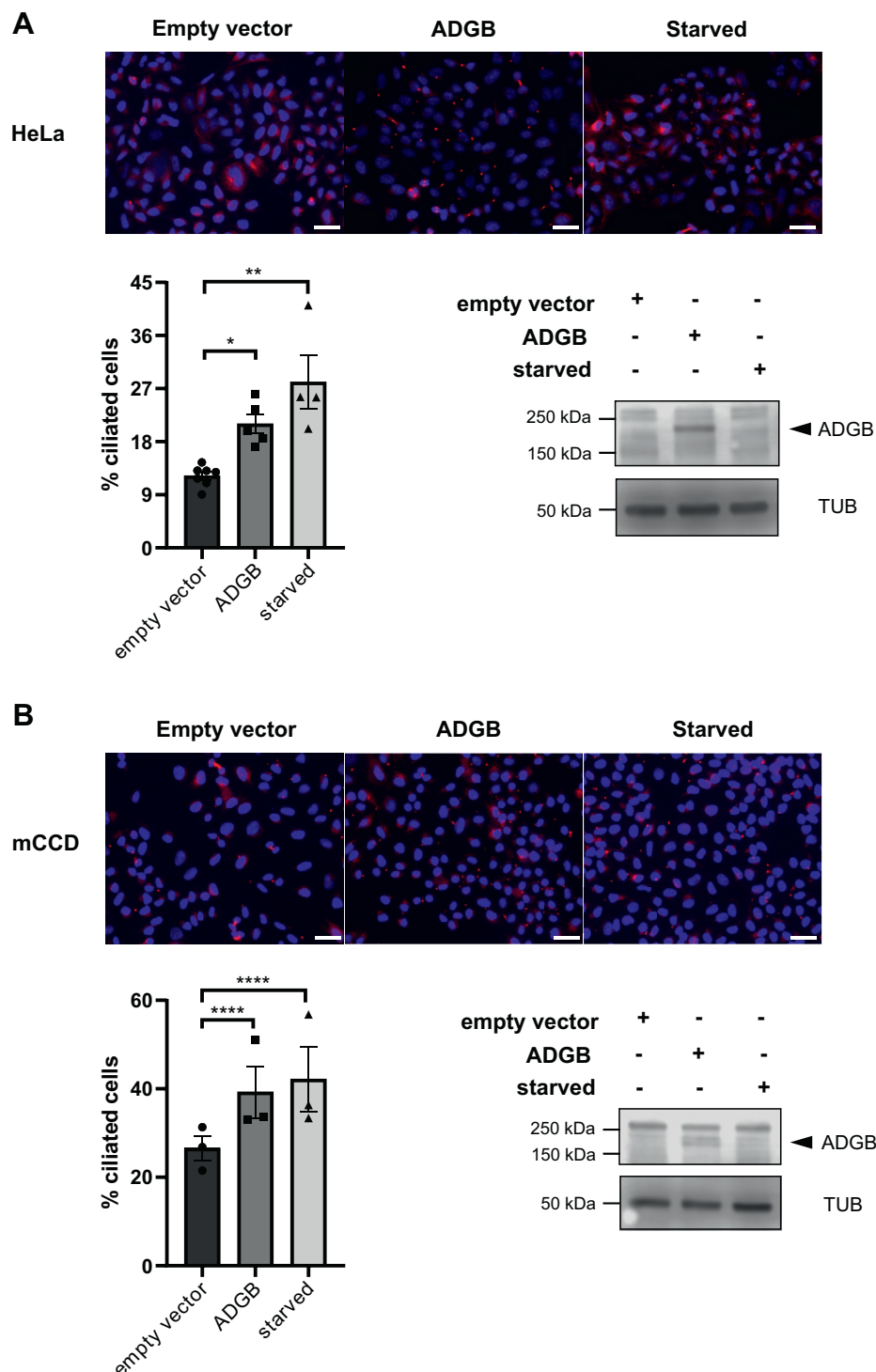


Figure 10. ADGB overexpression promotes ciliogenesis in two independent cell lines. *A*, representative pictures of acetylated α -tubulin (red fluorescence) and DAPI (blue fluorescence) in HeLa cells under basal conditions transfected with empty vector, following transfection with 2 μ g ADGB or following 24 h serum starvation (starved), and corresponding cilia quantifications (expressed as percentage of ciliated cells) ($n = 3-4$ independent experiments, 3-5 pictures were counted for each condition and per experiment). Scale bar represents 100 μ m. Overexpression of ADGB was verified by immunoblotting. Tubulin (TUB) was used as loading control. *B*, Similar representative immunofluorescence pictures of acetylated α -tubulin, corresponding cilia quantifications and immunoblotting following transfection with 2 μ g ADGB in mCCD_{cl1} cells. Scale bar represents 100 μ m. Data represent mean $\pm$ S.E.M (error bars); * $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$.

Immunohistochemistry of bovine tissue

After 20 h of fixation, the tissue was washed twice in PBS and cryoprotected in 20% saccharose solution. To avoid

desiccation, the tissue was enveloped in parafilm and stored at -80°C until further use. The samples were embedded in Neg-50 Frozen Section Medium (Thermo Scientific) and

sectioned in a cryostat at -20°C . Before immunostaining, heat-induced epitope retrieval was performed for 30 min at pH 6. Slides were permeabilized (PBS with 0.1% Triton X), blocked in blocking buffer (10% Horse serum in PBS with 0.1% Triton X), and probed with primary rabbit anti-ADGB antibody (1/100) (Sigma-Aldrich, HPA036340) in blocking buffer overnight. Incubation with secondary goat anti-rabbit CF 488 antibody (1/250) (Sigma-Aldrich, SAB4600036) or goat anti-rabbit Alkaline Phosphatase (1/500) (Sigma-Aldrich, A3687) was performed for 1 h. Fluorescent samples were counterstained with DAPI (Roche) and embedded in RotiMount Fluorcare antifading solution (Roth). For colorimetric staining, slides were incubated in NBT-BCIP (Roche) under exclusion of oxygen. Image acquisition was done either with a Leica SP5 or a fluorescence microscope BX61 (Olympus).

Protein extraction and immunoblotting

Protein extraction and immunoblotting were performed as before with some modifications (65). Cells were lysed in NP-40 lysis buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 400 mM NaCl, 0.1% NP-40, $2\text{ }\mu\text{g ml}^{-1}$ leupeptin, $2\text{ }\mu\text{g ml}^{-1}$ pepstatin, $2\text{ }\mu\text{g ml}^{-1}$ aprotinin, 1 mM PMSF) or in Triton X-100 buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton X-100, $2\text{ }\mu\text{g/ml}$ aprotinin, $4\text{ }\mu\text{g/ml}$ leupeptin, $2\text{ }\mu\text{g/ml}$ pepstatin and 1 mM PMSF), and centrifuged for 10 min at 14,000 rpm at 4°C . For the detection of CRISPRa-induced ADGB, total cell lysate was first immunoprecipitated with 0.2 μg of ADGB antibody (Sigma-Aldrich) and 25 μl bed volume of Protein-G-Sepharose beads (Sigma-Aldrich). Protein quantification was performed by Bradford assay. In total, 50 to 100 μg proteins was separated on 10% SDS acrylamide gels and electrically transferred onto nitrocellulose membranes (Amersham Protran Western blotting membranes). Following transfer membranes were blocked for 1 h with 5% dried milk/TBS-tween at room temperature and incubated overnight in 1% dried milk/TBS-tween with primary rabbit anti-ADGB (1/500) (Sigma-Aldrich, HPA036340), mouse anti-FOXJ1 (1/300) (Invitrogen, 14-9965-82), mouse anti-HA (1/100) (Santa Cruz, sc-57592), mouse anti FLAG (1/1000) (Sigma-Aldrich, F3165), or rabbit anti-CAS9 (1/2500) (ABclonal, A14997) antibodies. Membranes were then incubated for 1 h with HRP-conjugated goat anti-rabbit secondary antibody (GE Healthcare), and the signal was revealed using ECL Prime (Amersham) on a C-DiGit western blot scanner (LI-COR Biosciences) or fluorescent signals were visualized *via* a Li-COR Odyssey infrared imaging system.

Data analysis

Results are shown as mean values $\pm$ SEM of at least three independent experiments. Statistical analysis was performed applying two-tailed or paired Student's *t*-test or one-way analysis of variance (ANOVA) for multiple comparisons, using GraphPad Prism Version 7.0 (GraphPad Software).

Data availability

All data are contained within the article.

Acknowledgments—We thank Darko Maric (University of Fribourg) for helpful advice and discussions. CO and TH would like to thank Klaus-Dieter Fuchs and Dr Heinrich Dahmen (Fuchs GmbH EU Zerlege-und Schlachtbetrieb, Prüm, Germany) for generous donation of tissue and assistance in sample acquisition and Timo Fredi Kopp and Julian Thomas Mohr (JGU Mainz) for assistance in laboratory work.

Author contributions—T. W. K., C. O., I. M. C. O., A. K., D. A., S. Y., M. S. A., M. C., R. M., and D. H. data curation, formal analysis, and investigation; T. W. K., C. O., I. M. C. O., A. K., J. S., T. H., and D. H. methodology; T. W. K., C. O., and D. H. visualization; T. W. K., C. O., T. H., and D. H. writing-original draft; T. H. and D. H. funding acquisition, conceptualization, and supervision.

Funding and additional information—This work was supported by the Swiss National Science Foundation to D. H. (grant 31003A_173000) and the German Research Foundation to D. H. (HO 5837/1-1), T. H. (HA 2103/9-1) and J. S. (Projektnummer 387509280; SFB 1350).

Conflict of interest—The authors declare no conflicts of interest in regard to this article.

Abbreviations—The abbreviations used are: ADGB, androglobin; ALI, air-liquid interface; ANOVA, analysis of variance; ChIP, chromatin immunoprecipitation; CYGB, cytoglobin; HB, hemoglobin; MB, myoglobin; mTEC, mouse tracheal epithelial cells; NGB, neuroglobin; PFA, para-formaldehyde; RT, reverse transcription; sgRNA, single guide RNA; TSS, transcriptional start site.

References

- Burmester, T., Weich, B., Reinhardt, S., and Hankeln, T. (2000) A vertebrate globin expressed in the brain. *Nature* **407**, 520–523
- Burmester, T., Ebner, B., Weich, B., and Hankeln, T. (2002) Cytoglobin: A novel globin type ubiquitously expressed in vertebrate tissues. *Mol. Biol. Evol.* **19**, 416–421
- Burmester, T., and Hankeln, T. (2014) Function and evolution of vertebrate globins. *Acta Physiol. (Oxf.)* **211**, 501–514
- Keppner, A., Maric, D., Correia, M., Koay, T. W., Orlando, I. M. C., Vinogradov, S. N., and Hoogewijs, D. (2020) Lessons from the post-genomic era: Globin diversity beyond oxygen binding and transport. *Redox Biol.* **37**, 101687
- Hoogewijs, D., Ebner, B., Germani, F., Hoffmann, F. G., Fabrizio, A., Moens, L., Burmester, T., Dewilde, S., Storz, J. F., Vinogradov, S. N., and Hankeln, T. (2012) Androglobin: A chimeric globin in metazoans that is preferentially expressed in mammalian testes. *Mol. Biol. Evol.* **29**, 1105–1114
- Huang, B., Lu, Y. S., Li, X., Zhu, Z. C., Li, K., Liu, J. W., Zheng, J., and Hu, Z. L. (2014) Androglobin knockdown inhibits growth of glioma cell lines. *Int. J. Clin. Exp. Pathol.* **7**, 2179–2184
- Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, L., Edlund, K., Lundberg, E., Navani, S., Szegedy, C. A., et al. (2015) Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419
- Altmae, S., Koel, M., Vosa, U., Adler, P., Suhorutsenko, M., Laisk-Podar, T., Kukushkina, V., Saare, M., Velthut-Meikas, A., Krjatskov, K., Aghajanova, L., Lalitkumar, P. G., Gemzell-Danielsson, K., Giudice, L., Simon, C., et al. (2017) Meta-signature of human endometrial receptivity: A meta-analysis and validation study of transcriptomic biomarkers. *Sci. Rep.* **7**, 10077

9. Schuster, H., Peper, J. K., Bosmuller, H. C., Rohle, K., Backert, L., Bilich, T., Ney, B., Löffler, M. W., Kowalewski, D. J., Trautwein, N., Rabsteyn, A., Engler, T., Braun, S., Haen, S. P., Walz, J. S., *et al.* (2017) The immunopeptidomic landscape of ovarian carcinomas. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E9942–E9951
10. Yu, X., Ng, C. P., Habacher, H., and Roy, S. (2008) Foxj1 transcription factors are master regulators of the motile ciliogenic program. *Nat. Genet.* **40**, 1445–1453
11. Whitfield, M., Thomas, L., Bequignon, E., Schmitt, A., Stouvenel, L., Montantin, G., Tissier, S., Duquesnoy, P., Copin, B., Chantot, S., Dastot, F., Faucon, C., Barbotin, A. L., Loyens, A., Siffroi, J. P., *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
12. Comer, M. T., Leese, H. J., and Southgate, J. (1998) Induction of a differentiated ciliated cell phenotype in primary cultures of Fallopian tube epithelium. *Hum. Reprod.* **13**, 3114–3120
13. Haider, S., Gamperl, M., Burkard, T. R., Kunihs, V., Kaindl, U., Junttila, S., Fiala, C., Schmidt, K., Mendjan, S., Knofler, M., and Latos, P. A. (2019) Estrogen signaling drives ciliogenesis in human endometrial organoids. *Endocrinology* **160**, 2282–2297
14. Barton, B. E., Herrera, G. G., Anamthakmakula, P., Rock, J. K., Willie, A., Harris, E. A., Takemaru, K. I., and Winuthayanon, W. (2020) Roles of steroid hormones in oviductal function. *Reproduction* **159**, R125–R137
15. Ferenczy, A., Richart, R. M., Agate, F. J., Jr., Purkerson, M. L., and Dempsey, E. W. (1972) Scanning electron microscopy of the human fallopian tube. *Science* **175**, 783–784
16. Montoro, D. T., Haber, A. L., Biton, M., Vinarsky, V., Lin, B., Birket, S. E., Yuan, F., Chen, S., Leung, H. M., Villoria, J., Rogel, N., Burgin, G., Tsankov, A. M., Waghay, A., Slyper, M., *et al.* (2018) A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324
17. Lukassen, S., Bosch, E., Ekici, A. B., and Winterpacht, A. (2018) Characterization of germ cell differentiation in the male mouse through single-cell RNA sequencing. *Sci. Rep.* **8**, 6521
18. Spassky, N., and Meunier, A. (2017) The development and functions of multiciliated epithelia. *Nat. Rev. Mol. Cell Biol.* **18**, 423–436
19. Campbell, J. N., Macosko, E. Z., Fenselau, H., Pers, T. H., Lyubetskaya, A., Tenen, D., Goldman, M., Verstegen, A. M., Resch, J. M., McCarroll, S. A., Rosen, E. D., Lowell, B. B., and Tsai, L. T. (2017) A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.* **20**, 484–496
20. Mirzadeh, Z., Kusne, Y., Duran-Moreno, M., Cabrales, E., Gil-Perotin, S., Ortiz, C., Chen, B., Garcia-Verdugo, J. M., Sanai, N., and Alvarez-Buylla, A. (2017) Bi- and unciliated ependymal cells define continuous floor-plate-derived tanycytic territories. *Nat. Commun.* **8**, 13759
21. Gerdes, J. M., Liu, Y., Zaghloul, N. A., Leitch, C. C., Lawson, S. S., Kato, M., Beachy, P. A., Beales, P. L., DeMartino, G. N., Fisher, S., Badano, J. L., and Katsanis, N. (2007) Disruption of the basal body compromises proteasomal function and perturbs intracellular Wnt response. *Nat. Genet.* **39**, 1350–1360
22. Kowal, T. J., and Falk, M. M. (2015) Primary cilia found on HeLa and other cancer cells. *Cell Biol. Int.* **39**, 1341–1347
23. Tang, Z., Lin, M. G., Stowe, T. R., Chen, S., Zhu, M., Stearns, T., Franco, B., and Zhong, Q. (2013) Autophagy promotes primary ciliogenesis by removing OFD1 from centriolar satellites. *Nature* **502**, 254–257
24. Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., E, P. R. I., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., Ter-Ovanesyan, D., Braff, J. L., Davidsohn, N., Housden, B. E., Perrimon, N., *et al.* (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328
25. Consortium, E. P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74
26. Cheneby, J., Gheorghe, M., Artufel, M., Mathelier, A., and Ballester, B. (2018) ReMap 2018: An updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res.* **46**, D267–D275
27. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., and Cohen, D. (2017) GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**
28. Quigley, I. K., and Kintner, C. (2017) Rfx2 stabilizes Foxj1 binding at chromatin loops to enable multiciliated cell gene expression. *PLoS Genet.* **13**, e1006538
29. Kistler, W. S., Baas, D., Lemeille, S., Paschaki, M., Seguin-Estevez, Q., Barras, E., Ma, W., Duteyrat, J. L., Morle, L., Durand, B., and Reith, W. (2015) RFX2 is a major transcriptional regulator of spermiogenesis. *PLoS Genet.* **11**, e1005368
30. Bracke, A., Hoogewijs, D., and Dewilde, S. (2018) Exploring three different expression systems for recombinant expression of globins: *Escherichia coli*, *Pichia pastoris* and *Spodoptera frugiperda*. *Anal. Biochem.* **543**, 62–70
31. van Dam, T. J. P., Kennedy, J., van der Lee, R., de Vrieze, E., Wunderlich, K. A., Rix, S., Dougherty, G. W., Lambacher, N. J., Li, C., Jensen, V. L., Leroux, M. R., Hjeij, R., Horn, N., Texier, Y., Wissinger, Y., *et al.* (2019) CiliaCarta: An integrated and validated compendium of ciliary genes. *PLoS One* **14**, e0216705
32. Sigg, M. A., Menchen, T., Lee, C., Johnson, J., Jungnickel, M. K., Choksi, S. P., Garcia, G., 3rd, Busengdal, H., Dougherty, G. W., Pennenkamp, P., Werner, C., Rentzsch, F., Florman, H. M., Krogan, N., Wallingford, J. B., *et al.* (2017) Evolutionary proteomics uncovers ancient associations of cilia with signaling pathways. *Dev. Cell* **43**, 744–762.e711
33. Burmester, T., Storf, J., Hasenjager, A., Klawitter, S., and Hankeln, T. (2006) The hemoglobin genes of *Drosophila*. *FEBS J.* **273**, 468–480
34. Hoogewijs, D., De Henau, S., Dewilde, S., Moens, L., Couvreur, M., Borgonie, G., Vinogradov, S. N., Roy, S. W., and Vanfleteren, J. R. (2008) The *Caenorhabditis* globin gene family reveals extensive nematode-specific radiation and diversification. *BMC Evol. Biol.* **8**, 279
35. Hoogewijs, D., Geuens, E., Dewilde, S., Moens, L., Vierstraete, A., Vinogradov, S., and Vanfleteren, J. (2004) Genome-wide analysis of the globin gene family of *C. elegans*. *IUBMB Life* **56**, 697–702
36. Choksi, S. P., Lauter, G., Swoboda, P., and Roy, S. (2014) Switching on cilia: Transcriptional networks regulating ciliogenesis. *Development* **141**, 1427–1441
37. Nelson, G. A., Roberts, T. M., and Ward, S. (1982) *Caenorhabditis elegans* spermatozoan locomotion: Amoeboid movement with almost no actin. *J. Cell Biol.* **92**, 121–131
38. Acton, A. B. (1966) An unusual ciliumlike process. *J. Cell Biol.* **29**, 366–369
39. Gleixner, E., Herlyn, H., Zimmerling, S., Burmester, T., and Hankeln, T. (2012) Testes-specific hemoglobins in *Drosophila* evolved by a combination of sub- and neofunctionalization after gene duplication. *BMC Evol. Biol.* **12**, 34
40. Burmester, T., Wawrowski, A., Diepenbruck, I., Schrick, K., Seiwert, N., Ripp, F., Prothmann, A., and Hankeln, T. (2018) Divergent roles of the *Drosophila melanogaster* globins. *J. Insect Physiol.* **106**, 224–231
41. Jamieson, B. (1991) Ultrastructure and phylogeny of Crustacean spermatozoa. *Mem. Queensland Mus.* **31**, 109–142
42. Choksi, S. P., Babu, D., Lau, D., Yu, X., and Roy, S. (2014) Systematic discovery of novel ciliary genes through functional genomics in the zebrafish. *Development* **141**, 3410–3419
43. Stauber, M., Weidemann, M., Dittich-Breiholz, O., Lobschat, K., Alten, L., Mai, M., Beckers, A., Kracht, M., and Gossler, A. (2017) Identification of FOXJ1 effectors during ciliogenesis in the foetal respiratory epithelium and embryonic left-right organiser of the mouse. *Dev. Biol.* **423**, 170–188
44. Jacquet, B. V., Salinas-Mondragon, R., Liang, H., Therit, B., Buie, J. D., Dykstra, M., Campbell, K., Ostrowski, L. E., Brody, S. L., and Ghoshghaei, H. T. (2009) Foxj1-dependent gene expression is required for differentiation of radial glia into ependymal cells and a subset of astrocytes in the postnatal brain. *Development* **136**, 4021–4031
45. Campbell, E. P., Quigley, I. K., and Kintner, C. (2016) Foxn4 promotes gene expression required for the formation of multiple motile cilia. *Development* **143**, 4654–4664

46. Mukherjee, I., Roy, S., and Chakrabarti, S. (2019) Identification of important effector proteins in the FOXJ1 transcriptional network associated with ciliogenesis and ciliary function. *Front. Genet.* **10**, 23
47. Blatt, E. N., Yan, X. H., Wuerffel, M. K., Hamilos, D. L., and Brody, S. L. (1999) Forkhead transcription factor HFH-4 expression is temporally related to ciliogenesis. *Am. J. Respir. Cell Mol. Biol.* **21**, 168–176
48. Weidemann, M., Schuster-Gossler, K., Stauber, M., Wrede, C., Hegermann, J., Ott, T., Boldt, K., Beyer, T., Serth, K., Kremmer, E., Blum, M., Ueffing, M., and Gossler, A. (2016) CFAP157 is a murine downstream effector of FOXJ1 that is specifically required for flagellum morphogenesis and sperm motility. *Development* **143**, 4736–4748
49. Lewis, M., and Stracker, T. H. (2020) Transcriptional regulation of multiciliated cell differentiation. *Semin. Cell Dev. Biol.* <https://doi.org/10.1016/j.semcdb.2020.04.007>
50. Nemajerova, A., Kramer, D., Siller, S. S., Herr, C., Shomroni, O., Pena, T., Gallinas Suazo, C., Glaser, K., Wildung, M., Steffen, H., Sriraman, A., Oberle, F., Wienken, M., Hennion, M., Vidal, R., *et al.* (2016) TAp73 is a central transcriptional regulator of airway multiciliogenesis. *Genes Dev.* **30**, 1300–1312
51. Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915
52. Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295
53. Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550
54. Liao, Y., Wang, J., Jaehrig, E. J., Shi, Z., and Zhang, B. (2019) WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205
55. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019) Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e1821
56. Hoogewijs, D., Vogler, M., Zwenger, E., Krull, S., and Zieseniss, A. (2016) Oxygen-dependent regulation of aquaporin-3 expression. *Hypoxia (Auckl)* **4**, 91–97
57. Hoogewijs, D., Houthoofd, K., Matthijssens, F., Vandesompele, J., and Vanfleteren, J. R. (2008) Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in *C. elegans*. *BMC Mol. Biol.* **9**, 9
58. Seiler, C. Y., Park, J. G., Sharma, A., Hunter, P., Surapaneni, P., Sedillo, C., Field, J., Algar, R., Price, A., Steel, J., Throop, A., Fiacco, M., and LaBaer, J. (2014) DNASU plasmid and PSI:biology-materials repositories: Resources to accelerate biological research. *Nucleic Acids Res.* **42**, D1253–D1260
59. Gaeggeler, H. P., Gonzalez-Rodriguez, E., Jaeger, N. F., Loffing-Cueni, D., Norregaard, R., Loffing, J., Horisberger, J. D., and Rossier, B. C. (2005) Mineralocorticoid versus glucocorticoid receptor occupancy mediating aldosterone-stimulated sodium transport in a novel renal cell line. *J. Am. Soc. Nephrol.* **16**, 878–891
60. Storti, F., Santambrogio, S., Crowther, L. M., Otto, T., Abreu-Rodriguez, I., Kaufmann, M., Hu, C. J., Dame, C., Fandrey, J., Wenger, R. H., and Hoogewijs, D. (2014) A novel distal upstream hypoxia response element regulating oxygen-dependent erythropoietin gene expression. *Haematologica* **99**, e45–e48
61. Lauer, V., Grampp, S., Platt, J., Lafleur, V., Lombardi, O., Choudhry, H., Kranz, F., Hartmann, A., Wullich, B., Yamamoto, A., Coleman, M. L., Ratcliffe, P. J., Mole, D. R., and Schodel, J. (2020) Hypoxia drives glucose transporter 3 expression through hypoxia-inducible transcription factor (HIF)-mediated induction of the long noncoding RNA NIC1. *J. Biol. Chem.* **295**, 4065–4078
62. Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., Thakore, P. I., Glass, K. A., Ousterout, D. G., Leong, K. W., Guilak, F., Crawford, G. E., Reddy, T. E., and Gersbach, C. A. (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973–976
63. Schörg, A., Santambrogio, S., Platt, J. L., Schödel, J., Lindenmeyer, M. T., Cohen, C. D., Schrödter, K., Mole, D. R., Wenger, R. H., and Hoogewijs, D. (2015) Destruction of a distal hypoxia response element abolishes trans-activation of the *PAG1* gene mediated by HIF-independent chromatin looping. *Nucleic Acids Res.* **43**, 5810–5823
64. Orlando, I. M. C., Lafleur, V. N., Storti, F., Spielmann, P., Crowther, L., Santambrogio, S., Schodel, J., Hoogewijs, D., Mole, D. R., and Wenger, R. H. (2020) Distal and proximal hypoxia response elements cooperate to regulate organ-specific erythropoietin gene expression. *Haematologica* **105**, 2774–2784
65. Randi, E. B., Vervaeke, B., Tsachaki, M., Porto, E., Vermeylen, S., Lindenmeyer, M. T., Thuy, L. T. T., Cohen, C. D., Devuyst, O., Kistler, A. D., Szabo, C., Kawada, N., Hankeln, T., Odermatt, A., Dewilde, S., *et al.* (2020) The antioxidative role of cytoglobin in podocytes: Implications for a role in chronic kidney disease. *Antioxid. Redox Signal.* **32**, 1155–1171

2.2 Androglobin, a chimeric mammalian globin, is required for male fertility

Anna Keppner, Miguel Correia, Sara Santambrogio, Teng Wei Koay, Darko Maric, Carina Osterhof, Denise V Winter, Angèle Clerc, Michael Stumpe, Frédéric Chalmel, Sylvia Dewilde, Alex Odermatt, Dieter Kressler, Thomas Hankeln, Roland H Wenger, David Hoogewijs (2022) Androglobin, a chimeric mammalian globin, is required for male fertility. eLife 11:e72374

Supplementary files: <https://doi.org/10.7554/eLife.72374>

Own contributions to this Publication:

- mRNA and miRNA isolation of testis issue from WT and Adgb-KO mice following quality control via Nanodrop, Qubit and Bioanalyzer.
- Quality trimming and back mapping of mRNA-Seq reads of testis issue from WT and Adgb-KO mice.
- Quantification of expression levels and differential expression analysis with subsequent Gene Ontology analysis
- Pathway analysis of differentially expressed genes as well as analysis of potential transcription factors involved

Planning of experiments, analysis, and interpretation of data as well as drafting of the corresponding part of the manuscript were realized together with Prof. Dr. T. Hankeln. The project was managed by Prof Dr. David Hoogewijs.

Androglobin, a chimeric mammalian globin, is required for male fertility

Anna Keppner¹, Miguel Correia¹, Sara Santambrogio², Teng Wei Koay¹, Darko Maric¹, Carina Osterhof³, Denise V Winter⁴, Angèle Clerc¹, Michael Stumpe⁵, Frédéric Chalmel⁶, Sylvia Dewilde⁷, Alex Odermatt⁴, Dieter Kressler⁵, Thomas Hankeln³, Roland H Wenger², David Hoogewijs^{1*}

¹Department of Endocrinology, Metabolism and Cardiovascular system, University of Fribourg, Fribourg, Switzerland; ²Institute of Physiology, University of Zurich, Zurich, Switzerland; ³Institute for Organismic and Molecular Evolutionary Biology, University of Mainz, Mainz, Germany; ⁴Department of Pharmaceutical Sciences, University of Basel, Basel, Switzerland; ⁵Department of Biology, University of Fribourg, Fribourg, Switzerland; ⁶University of Rennes, Inserm, UMR_S 1085, Rennes, France; ⁷Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

Abstract Spermatogenesis is a highly specialized differentiation process driven by a dynamic gene expression program and ending with the production of mature spermatozoa. Whereas hundreds of genes are known to be essential for male germline proliferation and differentiation, the contribution of several genes remains uncharacterized. The predominant expression of the latest globin family member, androglobin (Adgb), in mammalian testis tissue prompted us to assess its physiological function in spermatogenesis. Adgb knockout mice display male infertility, reduced testis weight, impaired maturation of elongating spermatids, abnormal sperm shape, and ultrastructural defects in microtubule and mitochondrial organization. Epididymal sperm from Adgb knockout animals display multiple flagellar malformations including coiled, bifid or shortened flagella, and erratic acrosomal development. Following immunoprecipitation and mass spectrometry, we could identify septin 10 (Sept10) as interactor of Adgb. The Sept10-Adgb interaction was confirmed both *in vivo* using testis lysates and *in vitro* by reciprocal co-immunoprecipitation experiments. Furthermore, the absence of Adgb leads to mislocalization of Sept10 in sperm, indicating defective manchette and sperm annulus formation. Finally, *in vitro* data suggest that Adgb contributes to Sept10 proteolysis in a calmodulin-dependent manner. Collectively, our results provide evidence that Adgb is essential for murine spermatogenesis and further suggest that Adgb is required for sperm head shaping via the manchette and proper flagellum formation.

*For correspondence:
david.hoogewijs@unifr.ch

Competing interest: The authors declare that no competing interests exist.

Funding: See page 22

Received: 21 July 2021

Preprinted: 18 September 2021

Accepted: 13 June 2022

Published: 14 June 2022

Reviewing Editor: T Rajendra Kumar, University of Colorado, United States

© Copyright Keppner et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editor's evaluation

This manuscript demonstrates that male mice lacking androglobin, a poorly understood heme-containing protein, are infertile and have defects in late stage spermatogenesis. The revisions are thorough and the inclusion of additional data makes the manuscript solid. The lack of defects at the hypothalamus-pituitary level makes the phenotype more striking as a direct effect of the mutation at the testis level. Overall, this revised version is significantly improved.

Introduction

Spermatogenesis is a complex and dynamic differentiation process ending by the production of mature haploid spermatozoa (Hermo et al., 2010a; Hermo et al., 2010b; Hermo et al., 2010c; Hermo et al.,

2010d; Hermo et al., 2010e). While type A spermatogonia undergo mitosis, one of the daughter cells serves to replenish the stem cell population, and the other daughter cell further divides mitotically, differentiates, and eventually forms type B spermatogonia. Following another mitotic division, type B spermatogonia engender primary spermatocytes, which complete their first meiotic division and form two secondary spermatocytes. Each secondary spermatocyte completes a second meiotic division, leading to the production of two haploid round spermatids (Mecklenburg and Hermann, 2016). During the final differentiation phase of spermatogenesis, known as spermiogenesis, the spermatids undergo profound morphological changes to differentiate and elongate into spermatozoa. These changes include condensation of the nucleus and compaction of the genetic material, the formation of the acrosome (the sperm head) from the Golgi apparatus before migration of the latter to the cytoplasmic droplet (Khawar et al., 2019), the formation of the sperm flagellum from the centriole, around which mitochondria will migrate to form the midpiece, the sperm annulus, and the mobile tail, and cytoplasmic reduction, whereby all unnecessary cytoplasmic remnants are eliminated. At the end of the elongation process, the spermatozoa are released into the lumen of the seminiferous tubule to migrate into the rete testis and epididymis for final maturation. Dysfunctions in any of these tightly orchestrated steps could lead to impaired spermatogenesis, meiotic arrest, or abnormal sperm formation with direct consequences on male fertility (Neto et al., 2016). It is currently estimated that sperm defects and abnormalities remain idiopathic in about 30% of cases (Fainberg and Kashanian, 2019; Tüttelmann et al., 2018), and unraveling their molecular basis appears challenging since it is believed that over 4000 genes are possibly implicated in spermatogenesis (Jan et al., 2017).

Globins are small globular metallo-proteins, which have the capacity to reversibly bind gaseous ligands via a typical 8 alpha-helical structure in which a heme prosthetic group can be embedded. In mammals, five globin types exist: the well-established hemoglobin and myoglobin, neuroglobin in neuronal cells, cytoglobin ubiquitously expressed in fibroblasts, and the more recently identified androglobin (Adgb), predominantly expressed in mammalian testis tissue (Keppner et al., 2020). Adgb is a chimeric protein containing an N-terminal calpain-like cysteine protease domain, followed by an uncharacterized 300 amino acid long region, a central permuted functional globin domain (Bracke et al., 2018), interrupted by a potential calmodulin (CaM)-binding IQ motif, and a large 700 amino acid long C-terminal tail of unknown identity (Hoogewijs et al., 2012). Intriguingly, the chimeric nature of this globin resembles the domain structure of globin-coupled sensors found in prokaryotes (Hou et al., 2001; Thijs et al., 2007). CaM is one of the most conserved proteins in eukaryotes and serves as a secondary messenger following intracellular binding of Ca^{2+} and interaction with one of the more than 300 identified downstream target proteins (Andrews et al., 2020). CaM is the main intracellular receptor for Ca^{2+} , thereby participating in almost every biological process, including spermatogenesis and sperm maturation and function (Darszon et al., 2011). Decreased mRNA expression levels in infertile vs. fertile men (Platts et al., 2007) suggest a potential role of Adgb in spermatogenesis. Gene regulation and expression studies further suggest an association of Adgb with ciliogenesis including flagellum formation (Koay et al., 2021). However, the *in vivo* function of Adgb remains unexplored. In this study, we investigated the physiological function of Adgb during murine spermatogenesis by generating and analyzing Adgb knockout mice. We show that Adgb is mainly expressed in late steps of spermiogenesis, that it locates to the sperm flagellum, the annulus, and the midpiece, and that it is crucial for male fertility and sperm formation. Furthermore, we demonstrate that Adgb interacts with septin 10 (Sept10) and that co-localization is detectable within the sperm neck in stage 12 and stage 15 spermatids and within the annulus of stage 15 spermatids and mature sperm. Finally, *in vitro* data suggest that Adgb contributes to Sept10 proteolysis in a CaM-dependent manner.

Results

Adgb knockout mice display male infertility

A gene-trap strategy, provided by the Knockout Mouse Project (KOMP) (Skarnes et al., 2011), was applied to target exons 13 and 14 of the *Adgb* gene (Figure 1—figure supplement 1A). The correct targeting of ESCs was verified first by long-range PCR (Figure 1—figure supplement 1B) and second by Southern blotting (Figure 1—figure supplement 1C). The targeted *Adgb*^{tm1a(KOMP)Wtsi} allele (Adgb^{tm1a} mice), generated on a C57BL/6N background, displays a gene-trap DNA cassette, which was inserted into the 12th intron of the *Adgb* gene. The gene trap consists of a splice acceptor site,

an internal ribosome entry site, a β -galactosidase reporter sequence, and a neomycin resistance sequence. Breeding of Adgb tm1a mice with ubiquitously expressed CMV Cre-deleter mice allowed generation of mice deficient for exons 13 and 14 but still expressing the β -galactosidase reporter (Adgb tm1b mice) (**Figure 1—figure supplement 1A**). Furthermore, mating of Adgb tm1a mice with Flp-deleter mice enabled the generation of conditional floxed mice (Adgb tm1c) (**Figure 1—figure supplement 1A**). These mice were further crossed with CMV Cre-deleter mice to generate the full knockout animals (Adgb tm1d) (**Figure 1—figure supplement 1A**) that were used for all downstream applications if not otherwise stated. Genotyping was performed by regular PCR (**Figure 1—figure supplement 1D**) and revealed no significant differences in the Mendelian distribution of offspring for Adgb tm1d animals following interbreeding of heterozygous parents (380 pups: +/+, n=101; +/-, n=168; -/-, n=111; $\chi^2=5.62$, $p>0.05$, ns). The genetic ablation of Adgb expression was further verified by reverse transcription (RT)-quantitative (q)PCR (**Figure 1A**) and immunoblotting (**Figure 1B and C**). While female knockout mice displayed no fertility issues, male knockout mice never generated offspring, indicative for infertility. Full penetrance male infertility was also observed in homozygous tm1a and tm1b male mice, whereas homozygous tm1c animals showed normal fertility (data not shown). Accordingly, the testis weight was significantly reduced in knockouts (**Figure 1D**). Intratesticular testosterone (**Figure 1E**) as well as serum luteinizing hormone (LH) and follicle-stimulating hormone (FSH) levels (**Figure 1—figure supplement 2**) remained comparable between wild type and knockouts. Stage-specific histological examination of seminiferous tubules of control animals revealed normal architecture, normal spermatogenic maturation steps, and the presence of mature sperm with flagella extending into the lumen (**Figure 1F**). In contrast, in knockout animals, despite the presence of meiotic events (stages X–XII, **Figure 1F**), no flagella could be observed during the spermiation stage (stages VII–VIII, **Figure 1F**). The absence of mature sperm was accompanied by abnormally shaped heads, trapped stage 16 spermatids within the epithelium, and the presence of cytoplasmic material filling the lumen of the tubules (**Figure 1F**). Within the cauda epididymis, knockout animals displayed accumulations of residual bodies, cytoplasmic material, shed germ cells, and occasional abnormally shaped sperm heads but an overall absence of mature sperm as compared to wild-type animals (**Figure 1G**). No differences were detected in knockout testes at mRNA levels of nitric oxide synthases 1–3 (*Nos1–3*) or superoxide dismutases 1–3 (*Sod1–3*), the ratio of *Bax/Bcl2* was unchanged compared to wild-type testes, and terminal deoxynucleotidyl transferase dUTP nick-end labeling (TUNEL) assays on testis sections revealed no differences, suggesting no increase in oxidative stress or apoptotic events (**Figure 1—figure supplement 3**).

Absence of Adgb interferes with the maturation of elongating spermatids

To determine the temporal expression pattern of *Adgb* during spermatogenesis, wild-type embryos and pups at different post-natal ages, corresponding to the stages of the first wave of spermatogenesis during puberty, were dissected and analyzed by RT-qPCR. Whereas the expression of *Adgb* remained nearly undetectable until post-natal day 21 (corresponding to the stage of round spermatids), *Adgb* mRNA levels drastically increased to reach a peak at post-natal day 25, coinciding with the first elongating spermatids (**Figure 2A**). Bulk and single-cell RNA sequencing (scRNAseq) analysis in mouse and human datasets available at the ReproGenomics Viewer resource (*Darde et al., 2019; Darde et al., 2015*) as well as more recent murine scRNAseq data (*Kwak and Jung, 2019*) confirmed the conserved expression pattern of ADGB in spermatids (*Green et al., 2018; Jégou et al., 2017; Lukassen et al., 2018; Wang et al., 2018; Figure 2—figure supplement 1; Figure 2—figure supplement 2*). Accordingly, *Adgb* protein expression equally reached its peak at post-natal days 26–28, suggesting slightly delayed translation (**Figure 2B and C**). This finding was further confirmed by propidium iodide staining and FACS sorting on testis lysates of the different genotypes. While no variations could be detected for phases 2C (spermatogonia, secondary spermatocytes, and testicular somatic cells), S-phase (pre-meiotic spermatogonia), 4C (primary spermatocytes), and 1C (round spermatids), an abnormal accumulation of elongating and elongated spermatids (phase H) could be detected in knockout animals, suggesting a blockade in the elongation process (**Figure 2D**). Additionally, immunofluorescence (**Figure 2E**), mRNA *in situ* hybridization (**Figure 2F**), and X-gal (**Figure 2G**) stainings of testis sections from wild-type and knockout mice confirmed the presence of *Adgb* within layers containing post-meiotic cells and further intensifying toward the lumen and mature sperm

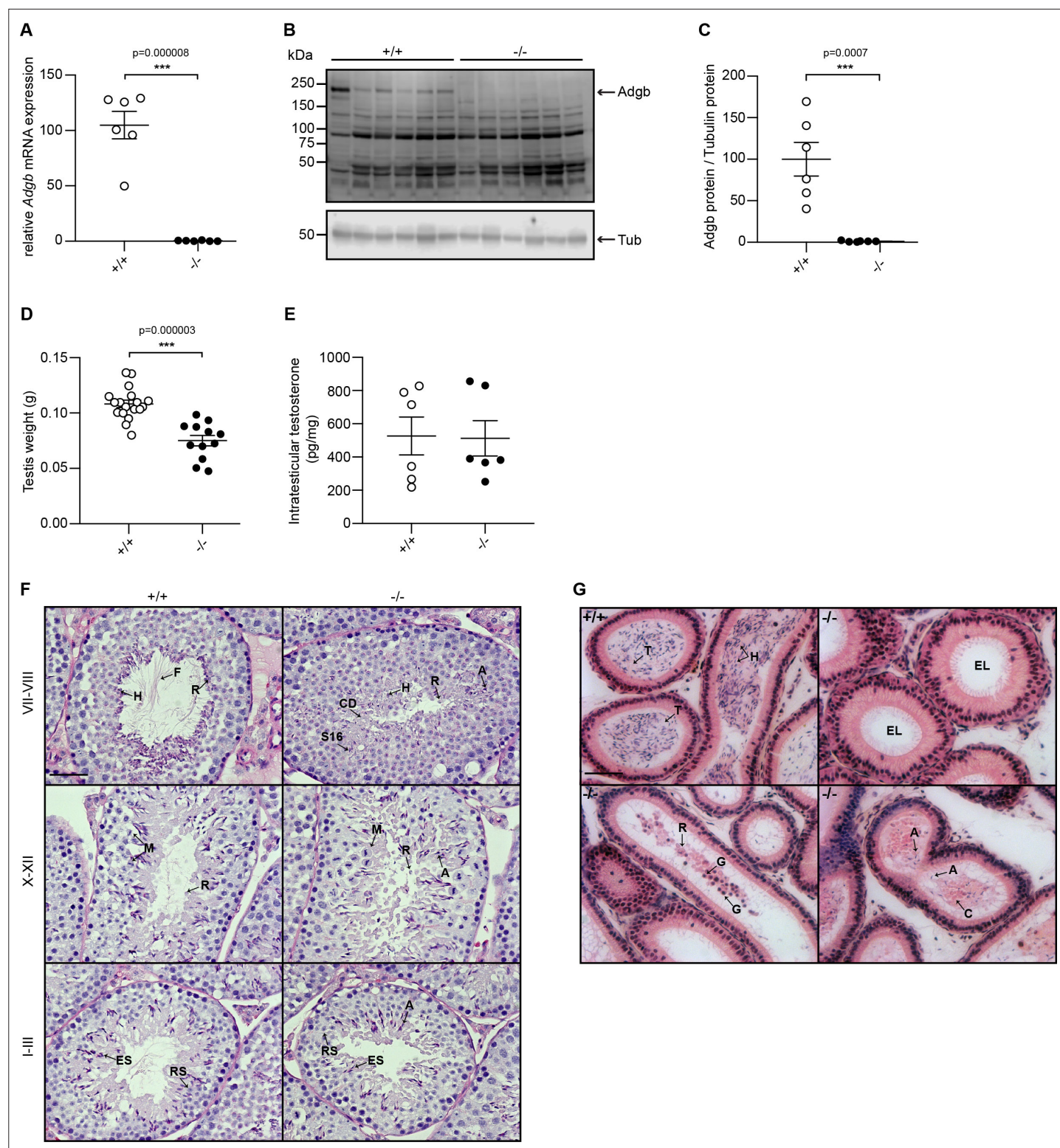


Figure 1. Validation of the knockout model and testicular phenotype. (A) Relative mRNA expression levels of *Adgb* in testes of wild-type (+/+) and knockout mice (-/-) (n=6 per genotype; p=0.000008). (B) Representative immunoblot for *Adgb* in testis lysates from wild-type (+/+) and knockout mice (-/-) (n=6 per genotype) and (C) corresponding protein quantification. Tubulin was used as loading control. p=0.0007. (D) Testis weight (g) in *Adgb* wild-type (+/+), heterozygous (+/-), and knockout (-/-) mice (n=8–13 per genotype). p=0.000003. (E) Intratesticular testosterone levels (pg/mg) in *Adgb* wild-type (+/+) and knockout (-/-) mice (n=6 per genotype). (F) Representative periodic acid Schiff (PAS)-hematoxylin-stained sections of testes from *Adgb* wild-type (+/+) and knockout mice (-/-) at the different stages of spermatogenesis. Heads (H), flagella (F), residual bodies (R), cytoplasmic debris

Figure 1 continued on next page

Figure 1 continued

(CD), meiosis (M), elongating spermatids (ES), round spermatids (RS), stage 16 spermatids (S16), and abnormal heads (A) are indicated. Note the full absence of flagella in knockout sections. Scale bar represents 50 μm . **(G)** Representative H&E stained sections of epididymides from Adgb wild-type (+/+) and knockout mice (-/-). H, tails (T), cytoplasmic bodies (C), R, germ cells (G), and A are shown. Note the empty lumen (EL) in knockout mice. Scale bar represents 50 μm . ** $p < 0.01$, *** $p < 0.001$.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1. Original uncropped immunoblots of **Figure 1B** with indication of the cropped areas.

Figure supplement 1. Generation of Adgb knockout mice.

Figure supplement 1—source data 1. Original uncropped gels of **Figure 1—figure supplement 1B,C,D** with indication of the cropped areas.

Figure supplement 2. Normal serum gonadotropin levels in Adgb knockout mice.

Figure supplement 3. Adgb knockout mice do not display changes in *Nos*, *Sod*, and apoptotic gene expression.

(Figure 2E–G). Moreover, in mature sperm, Adgb expression could be visualized within the midpiece and along the whole flagellum by both X-gal **(Figure 2G)** and immunofluorescence **(Figure 2H)** stainings.

Adgb is required for proper sperm flagellum formation

To gain additional insights into the origin of male infertility, cauda epididymal sperm was collected from both wild-type and knockout mice, and visualized under a microscope. While wild-type sperm appeared normal, very few knockout spermatozoa were found and displayed various defects of the head and/or flagellum structure, including shortened or bifid tails, loopings of the flagellum, and immature acrosomal structures **(Figure 3A)**. Knockout sperm acrosome structure appeared partially conserved, with sperm displaying either normal acrosomes, or various defects in shape or staining intensity, or total absence **(Figure 3B)**. Stage-specific transmission electron microscope (TEM) ultra-structural analysis of testis sections revealed various defects associated with sperm flagellum formation. Whereas numerous axonemes displaying the regular 9+2 structure from step 9 spermatids could be found in wild types, none were found in knockouts at the same stage. In later stages (step 12 spermatids and onward), rare and abnormal axonemes could be observed in knockouts, displaying disorganized microtubular structures and forming microtubular clusters without defined fibrous or mitochondrial sheath **(Figure 3C)**. We could further observe misshaped heads with nuclear inclusions, defective manchette elongation, and abnormal acrosomes in knockout sections **(Figure 3C–F)**.

The Adgb-dependent transcriptome reveals dysregulation of multiple spermiogenesis genes

To understand the molecular consequences of loss of Adgb in the testis, we performed RNAseq experiments on total testis RNA from wild-type and knockout mice at post-natal day 25. An elaborate set of significantly differentially expressed genes (74 genes upregulated and 204 downregulated) was identified, underscoring the crucial function of Adgb in spermatogenesis **(Figure 3—figure supplement 1A, Figure 3—source data 1)**. Functional analysis based on gene ontology term enrichments confirmed that many of these genes are related to sperm head, acrosome reaction, acrosomal membrane, sperm motility, spermatid development, and spermatid differentiation, in line with the pronounced structural changes in spermatids during spermiogenesis **(Figure 3—figure supplement 1B)**.

Adgb interacts and co-localizes with Sept10

To obtain more insights into the physiological function of Adgb, we explored the Adgb-dependent interactome. Total protein extracts from wild-type testes were immunoprecipitated (IP) with anti-Adgb or control IgG antibodies and subsequently submitted to mass spectrometry (MS) analysis to reveal potential interacting proteins of Adgb. Among the specifically enriched proteins, there were various members of the septin family, such as Sept10, Sept11, Sept2, and Sept7 **(Figure 4A and Figure 4—source data 4)**. Particular focus was put on Sept10 for further downstream experiments given its strong enrichment (4.509 log₂-fold) combined with high abundance (3.864 log₂[mean]) in the immunoprecipitation as well as substantial sequence coverage (40.5%), all reflected by its close position to Adgb among all septins in **Figure 4A**. To confirm the interaction between Adgb and

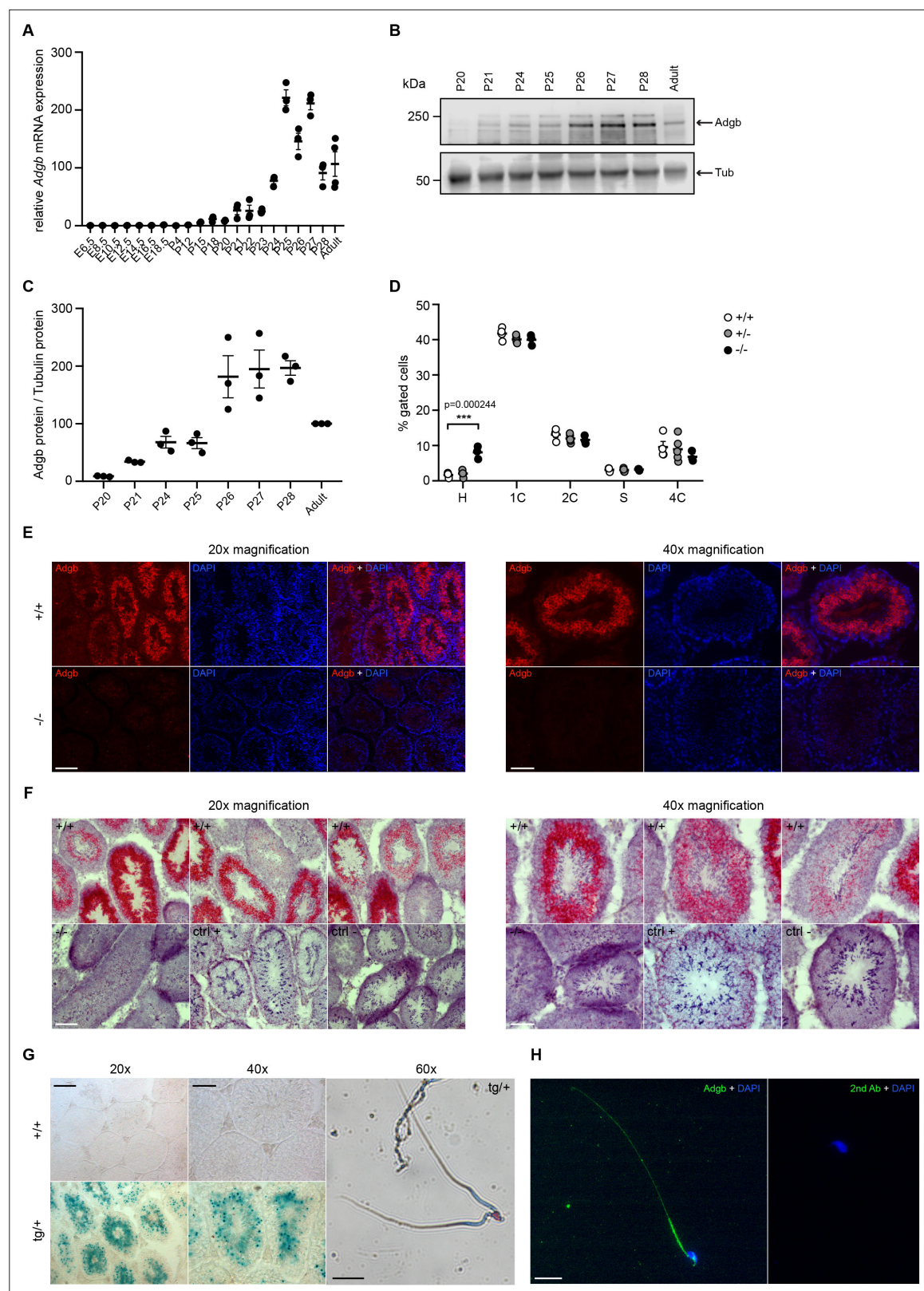


Figure 2. Testicular *Adgb* expression pattern and localization. **(A)** Relative mRNA expression levels of *Adgb* in testes of wild-type mice during embryonic development (E) and early post-natal (P) life ($n=3-4$ per condition). **(B)** Representative immunoblot for *Adgb* in testis lysates from wild-type mice at different P ages ($n=3$ per condition) and **(C)** corresponding protein quantification. Tubulin was used as loading control. **(D)** Flow cytometric analysis of spermatogenic cell populations following propidium iodide staining in *Adgb* wild-type ($+/+$, white circles, $n=4$), heterozygous ($+/-$, gray circles, $n=4$), and homozygous ($-/-$, black circles, $n=4$) mice. **(E)** Immunofluorescence images of testis sections stained for *Adgb* (red) and DAPI (blue) at 20x and 40x magnification. **(F)** Histology images of testis sections stained with H&E at 20x and 40x magnification. **(G)** Histology images of testis sections stained with H&E at 20x, 40x, and 60x magnification. **(H)** Immunofluorescence images of testis sections stained for *Adgb* (green) and DAPI (blue) at 40x magnification. Scale bars are shown in the bottom left of each panel.

Figure 2 continued

circles, n=5), and knockout (-/-, black circles, n=3) testes. H: elongating and elongated spermatids; 1 C: round spermatids; 2 C: spermatogonia, secondary spermatocytes, and testicular somatic cells; S: spermatogonia synthesizing DNA; 4 C: primary spermatocytes. p=0.00024. **(E)** Representative pictures of Adgb protein (red fluorescence) detection in testes of wild-type (+/+) and knockout (-/-) animals. Left panels 20× magnification, right panels 40× magnification, and scale bars represent 100 μm and 50 μm, respectively. Nuclei were stained with DAPI. **(F)** Representative pictures of Adgb mRNA *in situ* hybridization in testes from wild-type (+/+) and knockout (-/-) animals. Left panels 20× magnification; right panels, 40× magnification; scale bars represent 100 μm and 50 μm, respectively. Positive (ctrl +, PPIB) and negative (ctrl -, DapB) control sections are shown. **(G)** Representative pictures of β-galactosidase activity (X-gal staining) in testes from Tm1b wild-type (+/+) and Tm1b heterozygous (tg/+) mice and isolated spermatozoa from Tm1b heterozygous (tg/+) mice. Left panels, 20× magnification; middle panels, 40× magnification; right panel, 60× magnification; scale bars represent 100 μm, 50 μm, and 20 μm, respectively. Spermatozoa were counterstained with nuclear fast red. **(H)** Representative picture of Adgb protein (green fluorescence) in a single spermatozoon from wild-type (+/+) mice (left panel) and negative control (secondary antibody only, right panel). Scale bar represents 20 μm and nuclei were stained with DAPI. ** p<0.001.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Original uncropped immunoblots of **Figure 2B** with indication of the cropped areas.

Figure supplement 1. Temporal *Adgb* and *Sept10* expression profiles based on single-cell and bulk RNAseq datasets.

Figure supplement 2. *Adgb* and *Sept10* gene expression profiles based on single-cell RNAseq of mouse testes along the temporospatial axis of spermatid maturation.

Sept10, reciprocal co-immunoprecipitation (co-IP) experiments were performed (**Figure 4B and C**) on tissue extracts from wild-type and knockout testes (**Figure 4B**) and in HEK293 cells overexpressing full-length ADGB (**Figure 4C**) and SEPT10. The results demonstrate that Adgb and Sept10 interact both *in vivo* (**Figure 4B**) and *in vitro* (**Figure 4C**), whereas in testis lysates of Adgb-deficient mice, no Sept10 co-precipitation was observed (**Figure 4B**). Endogenous Sept10 protein levels were equal in testis lysates of Adgb-deficient and wild-type mice, as were endogenous levels of Sept11, Sept7, and Sept2, as well as other septins that are crucial for spermatogenesis, including Sept8, Sept9, and Sept14 (**Figure 4—figure supplement 1**). Reciprocal co-IP experiments between Adgb and other enriched septins (Sept2, Sept7, and Sept11) revealed no interaction, neither *in vivo* nor *in vitro* (**Figure 4—figure supplement 2**, **Figure 4—figure supplement 3**), whereas SEPT7 and SEPT10 interacted *in vitro* (**Figure 4—figure supplement 3**). We next investigated whether the interaction with SEPT10 occurs at the N-terminal or the C-terminal portion of ADGB (**Figure 4D and E**). Following co-overexpression of ADGB deletion constructs (**Figure 4—figure supplement 4**) with SEPT10 and subsequent co-IP, immunoblotting revealed that both parts of ADGB interact with SEPT10 (**Figure 4D and E**) and that this interaction remained intact also upon deletion of the coiled-coil domain of ADGB (**Figure 4—figure supplement 5**).

Consistent with a functional interaction, the temporal expression profiles of *SEPT10* and *ADGB* substantially overlap as illustrated by analysis of bulk and scRNAseq datasets of mouse and human RNA (**Figure 2—figure supplement 1**; **Figure 2—figure supplement 2**) as well as by RT-qPCR and immunoblotting of mouse tissue samples (**Figure 4—figure supplement 6**). The localization of Adgb and Sept10 was assessed in intact testis sections, microdissected tubules, and in epididymal sperm by immunofluorescence. Co-localization of Adgb and Sept10 was visible during spermatid maturation and in flagella in testis sections (**Figure 5A**) and at the level of the sperm annulus in S12 and S15 spermatids and mature wild-type sperm (**Figure 5B**, **Figure 5—figure supplement 1A**). A moderate Sept10 staining was also observed in the neck region of S15 spermatids and in mature sperm (**Figure 5B**, **Figure 5—figure supplement 1A**). The localization of Sept10 in knockout testis sections appeared overall fainter but within the same locations as in wild types, with the exception of sperm flagella (**Figure 5A**). However, in knockout S12 and S15 spermatids, as well as in epididymal sperm, only a single signal, likely corresponding to the annulus as verified by Sept7 staining of epididymal sperm (**Figure 5—figure supplement 1A**), was observed and displayed abnormal migration, indicating defective manchette or microtubule formation (**Figure 5B**, **Figure 5—figure supplement 1A**). The migration of the annulus drives mitochondrial placement along the forming mid-piece (Toure et al., 2011). Since knockout animals displayed abnormal ultrastructural mitochondria organization (**Figure 3**), CoxIV staining was performed on wild-type and knockout microdissected tubules and epididymal sperm. As expected, a robust staining was observed along the whole midpiece in wild-type spermatids and sperm, whereas mitochondria were barely visible and formed a cloudy structure around the neck region of knockout spermatids and sperm (**Figure 5—figure supplement 1B**).

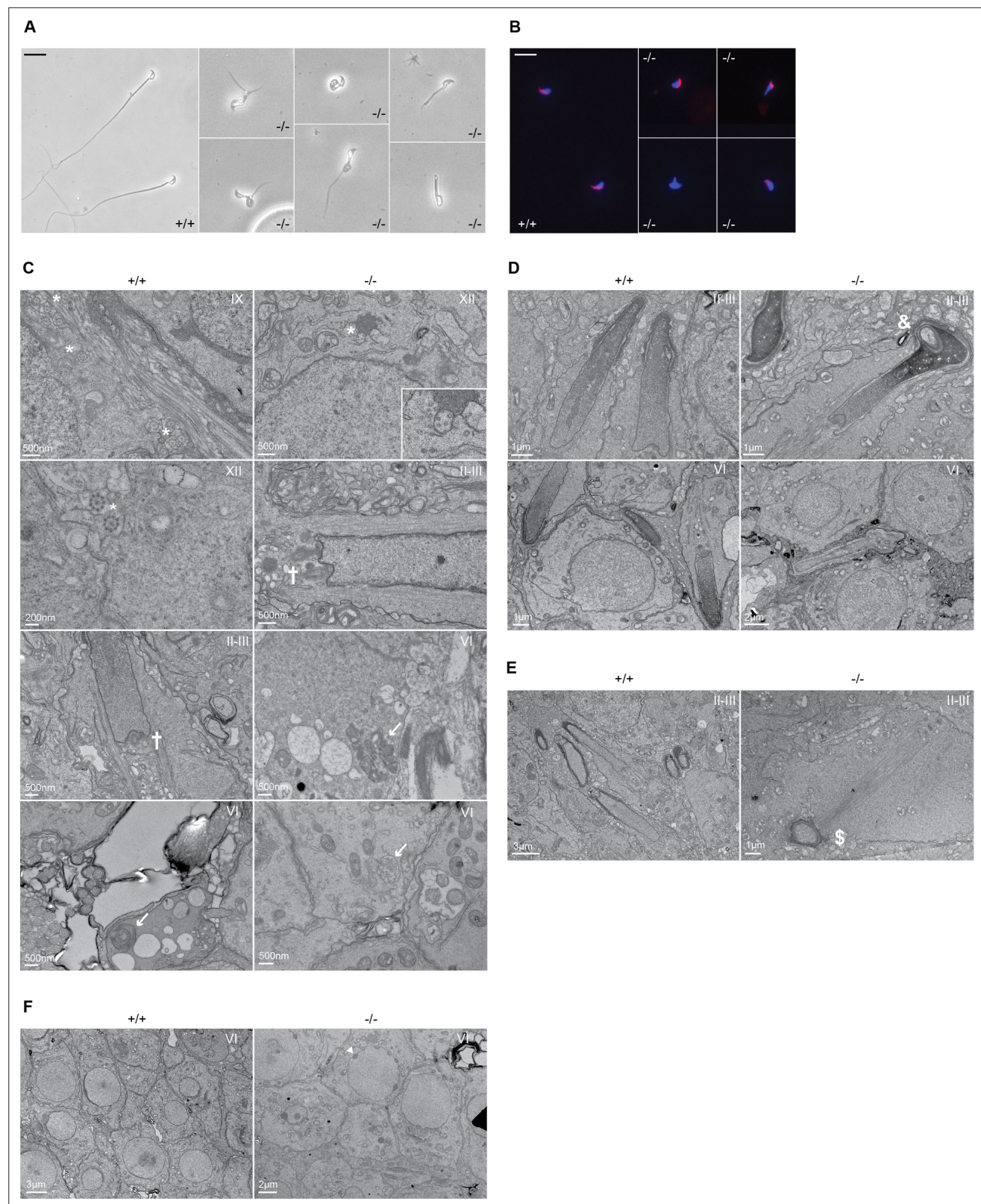


Figure 3. Defective spermatogenesis is associated with flagellar malformation in *Adgb* knockout mice. **(A)** Representative pictures of cauda epididymis sperm from wild-type (+/+) and *Adgb* knockout animals (-/-). Scale bar represents 20 μ m. **(B)** Representative pictures of peanut agglutinin-stained cauda epididymis sperm from wild-type (+/+) and *Adgb* knockout animals (-/-). Nuclei were stained with DAPI. Scale bar represents 20 μ m. **(C-F)** Representative transmission electron microscope (TEM) pictures from wild-type (+/+, left panels) and knockout (-/-, right panels) testes at various stages (IX, XII, II-III, VI). Scale bars represent 500 nm, 200 nm, 500 nm, and 500 nm respectively.

Figure 3 continued on next page

Figure 3 continued

stages of the first wave, tubular stages are indicated. (C) 9+2 microtubular structure (asterisks), forming sperm flagella (crosses), and (impaired) outer dense fibers (arrows) are shown. (D) Misshaped sperm heads with nuclear inclusions (ampersand), (E) defective manchette elongation (dollar), and (F) abnormal acrosomes (arrowheads) are shown. Scale bar lengths are indicated on each picture.

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. Differentially regulated genes in wild-type vs. *Adgb* knockout testes.

Figure supplement 1. Volcano plot and gene ontology (GO) term analysis of differentially expressed genes in *Adgb* knockout mice testis samples.

ADGB contributes to SEPT10 proteolytic cleavage *in vitro*

To analyze the functional consequences of the SEPT10-ADGB interaction, we transiently co-overexpressed both proteins in HEK293 cells. Intriguingly, apart from the intact form of overexpressed SEPT10 at 60 kDa, increased levels of a lower band of 37 kDa were consistently detected in the presence of co-overexpressed ADGB in a dose-dependent manner (**Figure 6A**). Immunoblotting with a V5-antibody upon co-overexpression of a C-terminal V5-tagged SEPT10 with ADGB (**Figure 6B**) as well as the presence of this band upon SEPT10/ADGB co-IP (**Figure 6B**, **Figure 6—figure supplement 1**) further supports its origin as proteolytic SEPT10 product. To investigate a potential oxygen-dependent influence of the globin domain, this experiment was repeated under normoxic and hypoxic conditions (0.2% O₂) but no differences were observed upon exposure to hypoxic conditions (**Figure 6C**). To determine a potential role of CaM, we constructed a deletion mutant lacking the IQ domain. Notably, transient overexpression of IQ-mutant ADGB resulted in considerably reduced appearance of the 37 kDa SEPT10 band relative to wild-type ADGB (**Figure 6D**), and the same was observed with an ADGB protease domain deletion mutant, further supporting a CaM-dependent proteolytic cleavage (**Figure 6D**). These findings prompted us to experimentally validate the suspected CaM-ADGB interaction. Whereas co-IP experiments following overexpression of full-length ADGB did not interact with CaM under basal experimental conditions in HEK293 cells (**Figure 6—figure supplement 2A**), a truncated construct covering the globin and IQ domains displayed robust ADGB-CaM interaction (**Figure 6E**). Consistently, MS analysis of proteins that were present in the IP of the overexpressed, isolated globin domain revealed a prominent enrichment of endogenous CaM (**Figure 6—figure supplement 3** and **Figure 6—source data 5**). Importantly, individual or double mutation of the proximal histidine (HisF8) or distal glutamine (GlnE7), critical residues in the globin domain for heme coordination, did not alter the interaction, suggesting that the ADGB-CaM interaction occurs independently of heme incorporation (**Figure 6F**, **Figure 6—figure supplement 4**). To fully exclude that the mutant globin domain might still be hemylated, we repeated these experiments under medium heme depletion and heme synthesis blocking conditions. Consistently, upon heme deprivation, robust interaction was observed between CaM and the isolated intact as well as double mutated (HisF8/GlnE7) globin domain. (**Figure 6—figure supplement 5**). As an additional layer of support for the ADGB-CaM interaction, chimeric Gal4 DNA-binding domain and VP16 transactivation domain-based fusion constructs were generated for mammalian 2-hybrid (M2H) luciferase reporter gene assays (**Figure 6G and H**). These M2H assays were performed in two different cell lines, HEK293 and A375, and revealed up to 7.5-fold increase in luciferase activity upon co-transfection of both chimeric proteins compared to single construct transfections, providing independent evidence that ADGB interacts with CaM. However, when full-length ADGB was employed in M2H assays with interchanged Gal4 and VP16 domains, no interaction with CaM was observed, corroborating the co-IP data (**Figure 6—figure supplement 2B**). To further assess the potential O₂-dependency of the CaM-ADGB interaction, we repeated the M2H assays in A375 cells under hypoxic conditions. Exposure to hypoxic conditions (0.2% O₂) did not alter luciferase activity (**Figure 6H**, left panel), while luciferase activity of a 5'/3'-hypoxia response element-dependent *EPO* promoter-driven reporter gene increased (**Figure 6H**, right panel), suggesting that the ADGB-CaM interaction is O₂ independent. Of note, while maintaining the O₂-independent interaction between Gal4-CaM and VP16-globin, the single Gal4-CaM construct control was increased under hypoxic conditions, albeit to a much lesser extent than the *EPO* control, possibly related to increases of cytoplasmic Ca²⁺ mobilized from intracellular stores or by extracellular influx under hypoxic conditions and according activation of CaM (Yuan et al., 2005). This result is consistent with the maintained ADGB-CaM interaction following mutation of critical residues in the globin domain in co-IP experiments and unchanged ADGB-dependent

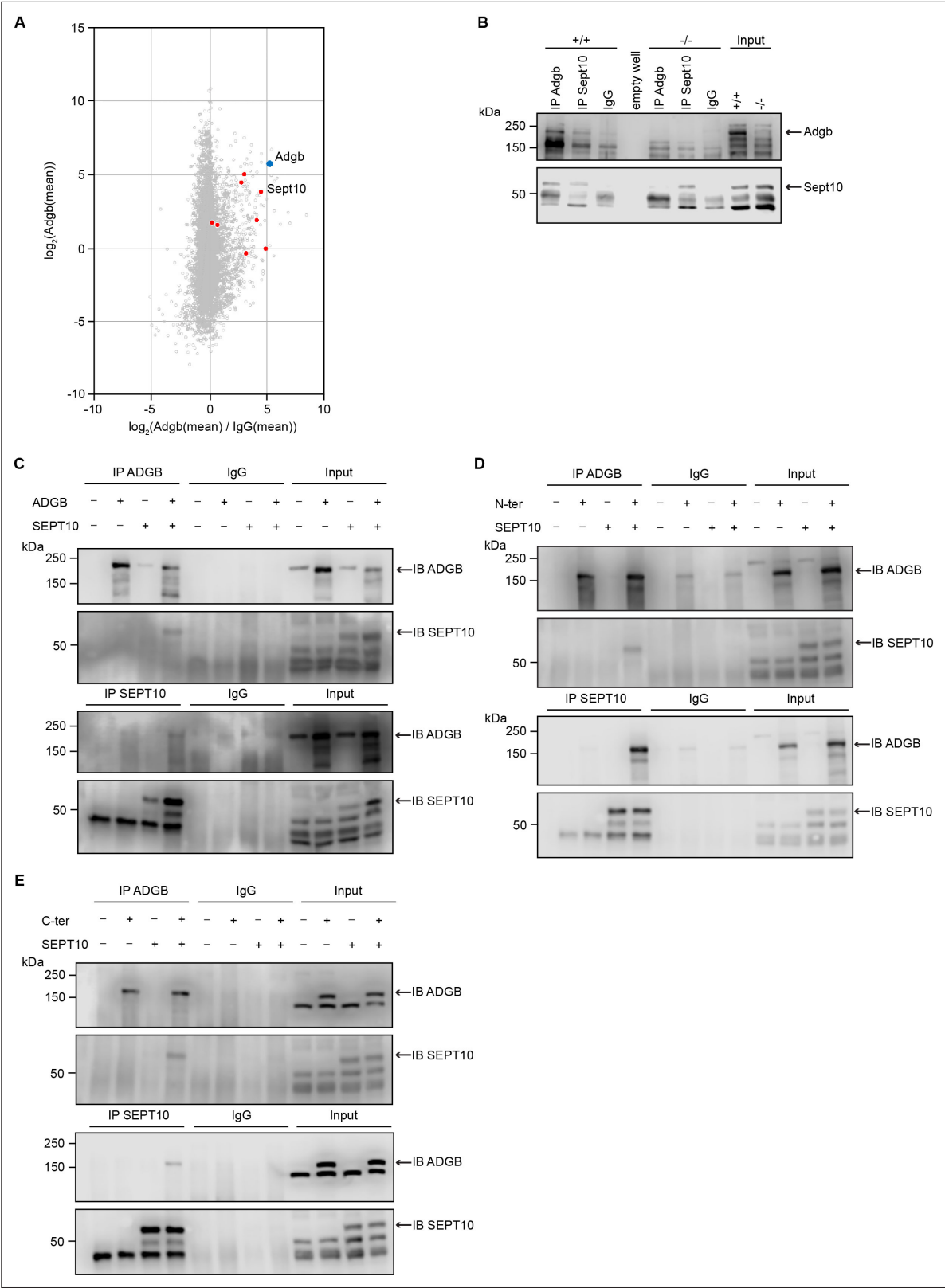


Figure 4. Adgb and Sept10 interact *in vivo* and *in vitro*. **(A)** Proteins of the septin family are specifically enriched in the Adgb immunoprecipitation (IP). The iBAQ (intensity-based absolute quantification) values of each Adgb IP (triplicate) and IgG control IP (duplicate) were log2 transformed and normalized against the median value. Missing values were imputed before the mean values of the Adgb and IgG control IPs were calculated. The normalized abundance of each protein detected in the Adgb IP (log2 Adgb [mean]) is plotted against its specific enrichment compared to the IgG

Figure 4 continued on next page

Figure 4 continued

control IP log2 (Adgb [mean]/IgG [mean]). Adgb and septins are highlighted as blue and red dots, respectively, in the christmas tree plot representation. (B) Representative immunoblot of Adgb and Sept10 in testis lysates from wild-type (+/+) and knockout (-/-) mice following co-IP of Adgb and Sept10. (C–E) Representative immunoblots of ADGB and SEPT10 in protein lysates of HEK293 cells (co-)transfected with full-length ADGB (C), N-ter ADGB (D) and C-ter ADGB (E), and SEPT10 following co-IP of ADGB and SEPT10. Schematic representation of deletion constructs is provided in **Figure 4—figure supplement 4**.

The online version of this article includes the following source data and figure supplement(s) for figure 4:

Source data 1. Original uncropped immunoblots of **Figure 4B and C** with indication of the cropped areas.

Source data 2. Original uncropped immunoblots of **Figure 4D** with indication of the cropped areas.

Source data 3. Original uncropped immunoblots of **Figure 4E** with indication of the cropped areas.

Source data 4. Raw mass spectrometry (MS) data of the Adgb immunoprecipitation (IP) vs. IgG control IP.

Figure supplement 1. The protein expression levels of Sept2, Sept7, Sept8, Sept9, Sept10, Sept11, and Sept14 are unaffected in Adgb knockout testis.

Figure supplement 1—source data 1. Original uncropped immunoblots of **Figure 4—figure supplement 1A–C** with indication of the cropped areas.

Figure supplement 1—source data 2. Original uncropped immunoblots of **Figure 4—figure supplement 1D–F** with indication of the cropped areas.

Figure supplement 1—source data 3. Original uncropped immunoblots of **Figure 4—figure supplement 1G** with indication of the cropped areas.

Figure supplement 2. Adgb does not interact with other septin family members *in vivo*.

Figure supplement 2—source data 1. Original uncropped immunoblots of **Figure 4—figure supplement 2A,B,C** with indication of the cropped areas.

Figure supplement 3. ADGB does not interact with other septin family members *in vitro*.

Figure supplement 3—source data 1. Original uncropped immunoblots of **Figure 4—figure supplement 3A** with indication of the cropped areas.

Figure supplement 3—source data 2. Original uncropped immunoblots of **Figure 4—figure supplement 3B** with indication of the cropped areas.

Figure supplement 3—source data 3. Original uncropped immunoblots of **Figure 4—figure supplement 3C** with indication of the cropped areas.

Figure supplement 3—source data 4. Original uncropped immunoblots of **Figure 4—figure supplement 3D** with indication of the cropped areas.

Figure supplement 3—source data 5. Original uncropped immunoblots of **Figure 4—figure supplement 3E** with indication of the cropped areas.

Figure supplement 4. ADGB constructs used throughout the study.

Figure supplement 5. The interaction between ADGB and SEPT10 is maintained despite mutation of the coiled-coil domains.

Figure supplement 5—source data 1. Original uncropped immunoblots of **Figure 4—figure supplement 5** with indication of the cropped areas.

Figure supplement 6. Sept10 temporal expression profile on mRNA and protein levels.

Figure supplement 6—source data 1. Original uncropped immunoblots of **Figure 4—figure supplement 6** with indication of the cropped areas.

cleavage of SEPT10 under hypoxic conditions. Collectively, these *in vitro* data suggest a scenario in which overexpressed ADGB proteolytically contributes to cleavage of overexpressed SEPT10 in an O₂-independent but CaM-dependent manner.

Discussion

In the present study, we explored the testicular function of Adgb by generating and analyzing Adgb constitutive knockout mice. Our results demonstrate that Adgb is indispensable for proper spermatogenesis and male fertility. The absence of Adgb has profound deleterious effects on the elongation of spermatids, leading to multiple malformations of the sperm flagellum and head, as evidenced by the presence of various ultrastructural defects, including aberrant microtubule formation, misshaped heads with nuclear inclusions, defective manchette formation, and abnormal acrosomes. Notably, Adgb could be localized not only to the midpiece and sperm flagellum but also in the neck region (likely the centriole) and within the sperm annulus.

In principle, the observed male infertility could be the result of defects in the hypothalamic-pituitary-gonadal (HPG) axis. Unchanged LSH and LH levels indicate an intact HPG signaling and suggest the phenotypes represent direct (primary) effects of Adgb deficiency in testis. However, indirect effects cannot be entirely excluded despite normal serum levels of gonadotropins. Unchanged testicular testosterone levels are consistent with the total absence of Adgb expression in Leydig cells of wild-type mice.

Strikingly, upon LC-MS/MS interactome analysis after Adgb immunoprecipitation in testicular lysates, several members of the septin family of proteins ranked among the top hits. Septins are

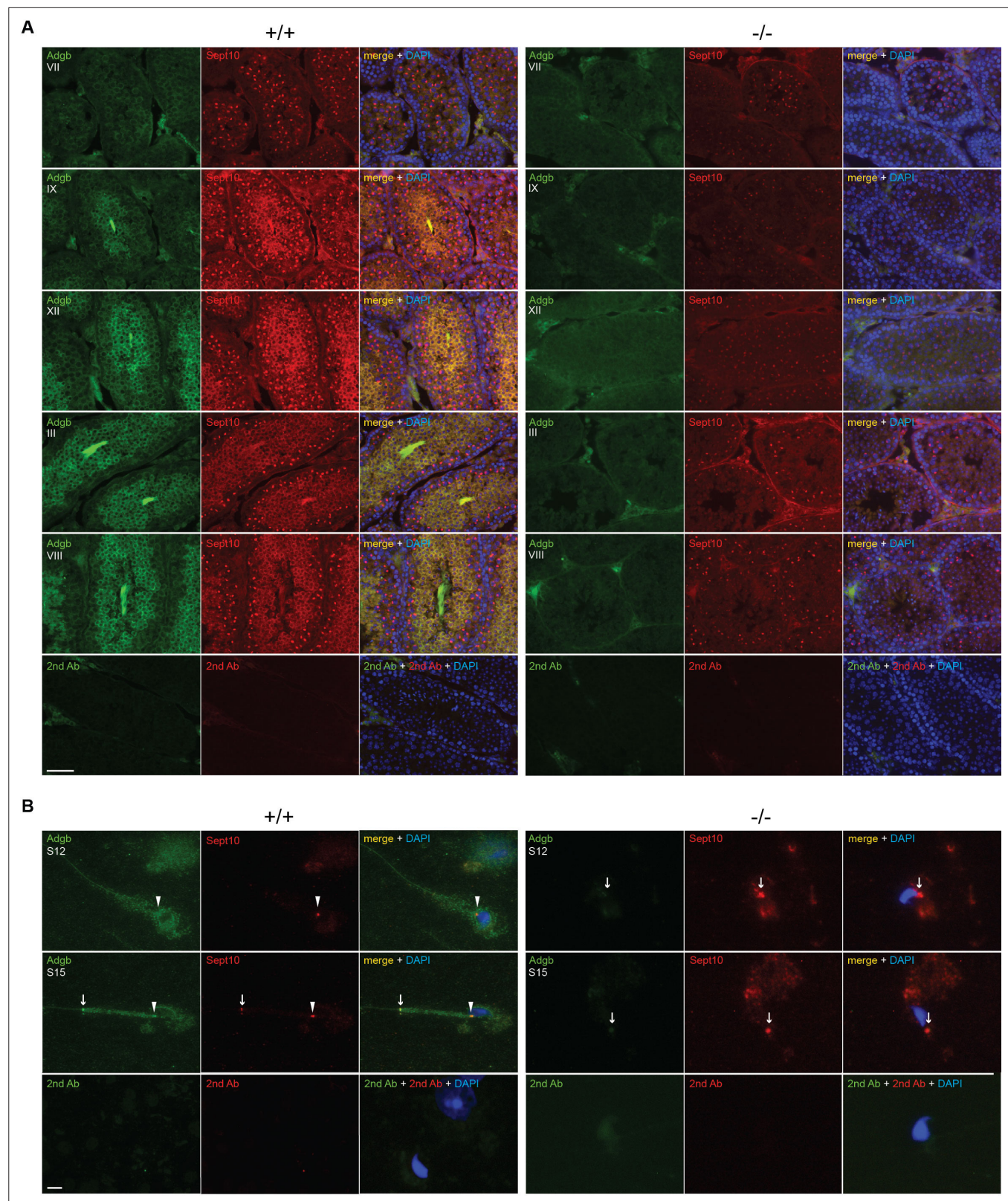


Figure 5. Adgb and Sept10 co-localize in the sperm neck and annulus. **(A)** Representative pictures of Adgb protein (green fluorescence) and Sept10 (red fluorescence) in testis sections from wild-type (+/+, left panels) and knockout (-/-, right panels) mice at various stages of the first wave, tubular stages are indicated. Sections were counterstained with DAPI. Negative control (secondary antibodies only) is shown (lower panels). Scale bar represents 50 μ m. **(B)** Representative pictures of Adgb (green fluorescence) and Sept10 (red fluorescence) in elongating spermatids (stage 12 [S12] upper panels

Figure 5 continued on next page

Figure 5 continued

and stage 15 [S15] middle panels) after stage-specific tubule dissection of wild-type (+/+) and knockout (-/-) testes. Nuclei were stained with DAPI. Negative controls (secondary antibodies only) are shown (lower panels). Scale bar represents 10 μ m. Sperm neck (arrowhead) and annulus (arrow) are highlighted.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Absence of Adgb leads to abnormal annulus migration and mitochondrial disorganization.

conserved GTPases that have the ability to form large oligomers and filamentous polymers and which associate with cell membranes and with the cytoskeleton. They serve as scaffolds for the proper localization of intracellular proteins via their diffusion barrier-forming characteristics (Dolat et al., 2014). In sperm, various septins (including Sept1, Sept4, Sept6, Sept7, and Sept12) have been localized to the sperm annulus, where they polymerize to a filamentous structure called the septin ring, forming a barrier between the midpiece and the principal piece of the spermatozoon (Toure et al., 2011). Interestingly, despite several members of the septin family being present in the interactome, only Sept10 was found to interact with Adgb in co-IP experiments. Furthermore, the interaction of the two proteins could also be localized within the connecting piece and within the annulus in wild-type sperm, whereas in knockout sperm, only one signal was detected, suggesting the absence of the annulus. Indeed, when staining against Sept7 which interacts with Sept10, the annulus was either missing or failed to migrate properly in knockout sperm, supporting a defect in the annulus formation and migration, thereby also leading to misalignment of the mitochondria in knockout sperm. In accordance with this observation, two main functions have been proposed for the annulus: (1) a diffusion barrier function to compartmentalize different proteins to various locations in the sperm tail and (2) a growth guide function for the sperm flagellum and for aligning the mitochondria along the axoneme (Avidor-Reiss et al., 2020; Avidor-Reiss et al., 2017; Toure et al., 2011). Supportive of this function, Sept4^{-/-} and Sept12^{+/-} mice are infertile and display disorganized sperm mitochondria (Ihara et al., 2005; Kissel et al., 2005; Lin et al., 2009). Moreover, Sept4^{-/-} mice display a bent sperm tail and absence of annulus (Ihara et al., 2005; Kissel et al., 2005), whereas Sept12^{+/-} mice exhibit broken acrosomes, misshaped nuclei, and increased apoptosis of germ cells (Lin et al., 2009). Correspondingly, SEPT12 mutations have been described in infertile men displaying abnormal sperm including defective annulus with a bent tail (Kuo et al., 2012). Additionally, defective sperm head morphology and DNA integrity have recently been reported for two different SEPT14 missense mutations (Lin et al., 2020; Wang et al., 2019). Interestingly, another ring-like septin structure was recently described at the sperm neck, composed of Sept12 which complexes together with Sept1, Sept2, Sept10, and Sept11. Two mutations of Sept12 identified in patients disrupted the complex, leading to unstable head-tail junctions and defective connecting piece formation. Strikingly, the mutation of Sept12 and the subsequent disruption of the complex led to loss of Sept10 signal in the annulus (Shen et al., 2020) as also observed in Adgb-deficient mice. Accordingly, our data suggest an interdependence between Adgb and Sept10, which is required for the maintenance of the annulus, head shaping, and proper mitochondrial localization.

Sperm flagella and motile cilia, which are hair-like microtubular protrusions at the surface of various cell types, share numerous characteristics, not only in their structure but also in their regulation and growth, whereby septins have also been identified as components of cilia. Sept2 forms a diffusion barrier at the base of the cilium, impeding ciliary formation through loss of Sonic Hedgehog signaling when depleted (Hu et al., 2010). Sept2/7/9 form a complex that associates with the ciliary axoneme, thereby regulating ciliary length (Ghossoub et al., 2013). Accordingly, septin association with the cytoskeleton and particularly with microtubular structures has been extensively studied (Spiliotis and Nakos, 2021), and numerous other cilia-related proteins participate in sperm flagellum formation. Furthermore, most ciliopathies include male infertility and immotile sperm due to defective axonemal organization (Brown and Witman, 2014). Likewise, Adgb knockout mice display aberrant microtubule arrangements, thus it cannot be excluded that a scaffolding and simultaneous regulatory action between Adgb and Sept10 is necessary to support microtubular structure. Moreover, a recent study reported the consistent presence of Adgb in the ciliomes of three distinct evolutionary ancestral taxa, further suggesting a conserved function related to microtubular organization and likely flagella formation (Sigg et al., 2017). In line with this study, our recent *in vitro* investigations have demonstrated

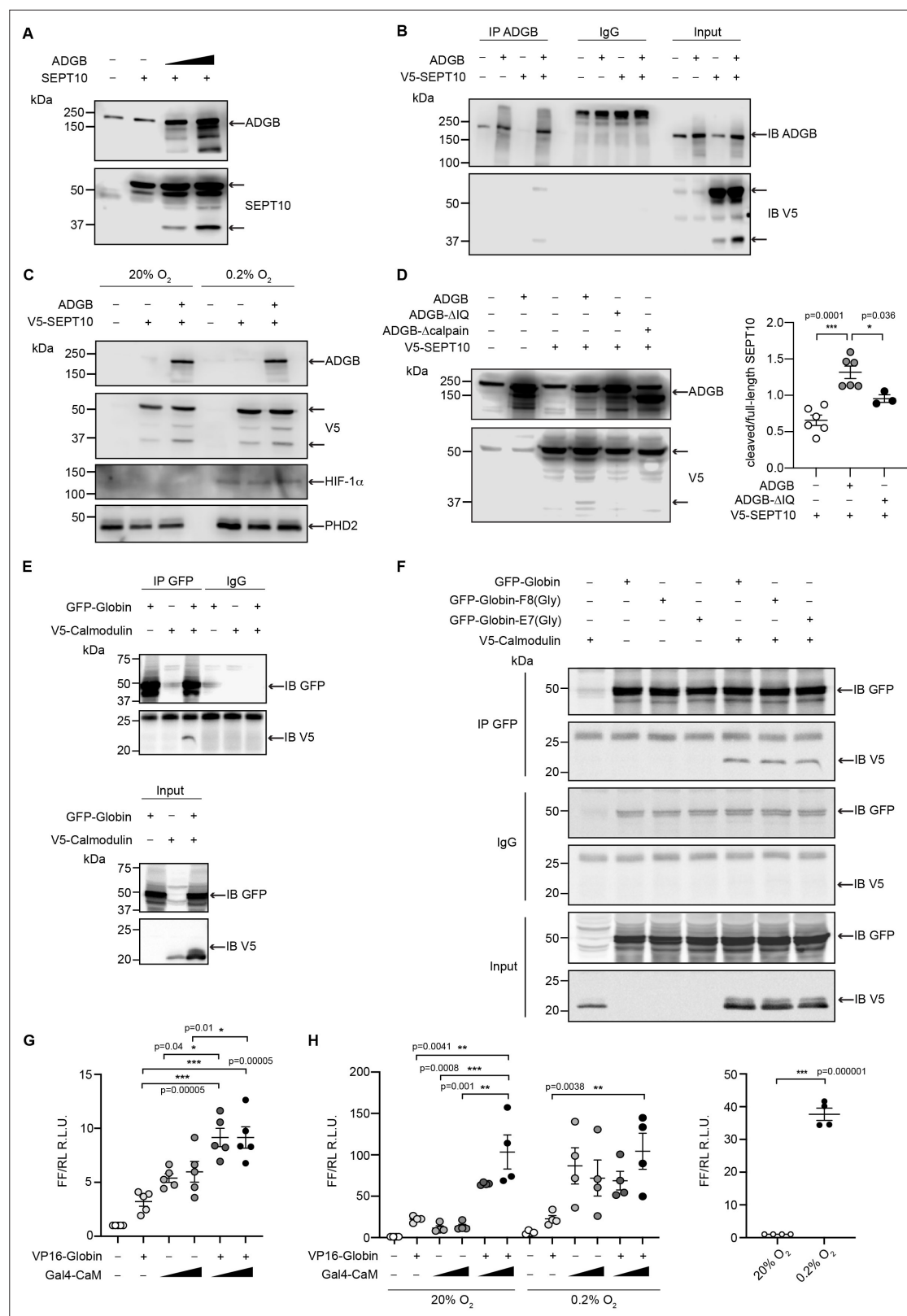


Figure 6. ADGB contributes to *in vitro* CaM-dependent SEPT10 cleavage. **(A)** Representative immunoblots of ADGB and SEPT10 in protein lysates of HEK293 cells co-transfected with plasmids encoding SEPT10 and two dose-dependent amounts of full-length ADGB. **(B)** Representative immunoblot of ADGB and V5 in protein lysates of HEK293 cells (co-)transfected with full-length ADGB and a C-terminally V5-tagged SEPT10 construct following co-immunoprecipitation (co-IP) of ADGB and V5-SEPT10. **(C)** Representative immunoblots of ADGB, V5, HIF-1α, and PHD2 in protein lysates of HEK293

Figure 6 continued on next page

Figure 6 continued

cells (co-)transfected with full-length ADGB and V5-SEPT10 following exposure to normoxic (20% O₂) and hypoxic conditions (0.2% O₂) for 24 hr. HIF-1 α and PHD2 were used as positive controls for hypoxia. **(D)** Representative immunoblot of ADGB and V5 in protein lysates of HEK293 cells (co-)transfected with full-length ADGB, V5-SEPT10, an ADGB-IQ deletion mutant, and an ADGB-calpain protease domain deletion mutant and corresponding protein quantification of cleaved/full-size SEPT10 ratio (n=3–6 independent experiments). Ponceau S protein staining was used as loading control. Schematic representation of deletion constructs is provided in **Figure 4—figure supplement 4**. **(E)** Representative immunoblot of GFP and V5 in protein lysates of HEK293 cells (co-)transfected with a truncated construct of the globin domain of ADGB (spanning the IQ domain) (GFP-Globin) and a V5-tagged CaM (V5-Calmodulin) following IP of GFP. **(F)** Representative immunoblot of GFP and V5 in protein lysates of HEK293 cells (co-)transfected with GFP-Globin, a GFP-Globin construct with mutation of the proximal heme-binding histidine (GFP-Globin-F8[Gly]), a GFP-Globin construct with mutation of the distal glutamine (GFP-Globin-E7[Gly]), and V5-CaM following IP of GFP. **(G, H)** Mammalian 2-hybrid assays in HEK293 cells under normoxic conditions (G) and A375 cells under normoxic and hypoxic (0.2% O₂) conditions (H) (n=3–5 independent experiments). HEK293 and A375 cells were transiently transfected with fusion protein vectors based on a Gal4 DNA-binding domain fused to calmodulin (Gal4-CaM) and a VP16 activation domain fused to the ADGB globin domain comprising the IQ domain (VP16-Globin), a Gal4 response element-driven firefly luciferase reporter, and a *Renilla* luciferase control vector. Increasing transfection amounts for the Gal4-CaM fusion protein were employed. Following transfection, A375 cells were incubated under normoxic (20% O₂) or hypoxic (0.2% O₂) conditions, and luciferase reporter gene activities were determined 24 hr later. Single construct transfections served as negative controls, whereby hypoxic regulation of CaM has been described previously (Yuan et al., 2005). A 5'/3'-hypoxia response element-dependent *EPO* promoter-driven firefly luciferase construct served as hypoxic control. * p<0.05, ** p<0.01, and *** p<0.001.

The online version of this article includes the following source data and figure supplement(s) for figure 6:

Source data 1. Original uncropped immunoblots of **Figure 6A and B** with indication of the cropped areas.

Source data 2. Original uncropped immunoblots of **Figure 6C** with indication of the cropped areas.

Source data 3. Original uncropped immunoblots of **Figure 6E** with indication of the cropped areas.

Source data 4. Original uncropped immunoblots of **Figure 6F** with indication of the cropped areas.

Source data 5. Raw mass spectrometry (MS) data of the ADGB globin IP vs. GFP control IP.

Figure supplement 1. Reciprocal co-immunoprecipitation (co-IP) of ADGB and V5-SEPT10 from **Figure 6B**.

Figure supplement 1—source data 1. Original uncropped immunoblots of **Figure 6—figure supplement 1** with indication of the cropped areas.

Figure supplement 2. Full-length ADGB does not interact with CaM.

Figure supplement 2—source data 1. Original uncropped immunoblots of **Figure 6—figure supplement 2** with indication of the cropped areas.

Figure supplement 3. ADGB globin immunoprecipitation (IP) vs. GFP control IP.

Figure supplement 4. Double mutation of the key heme-binding residues does not abrogate interaction of ADGB and CaM.

Figure supplement 4—source data 1. Original uncropped immunoblots of **Figure 6—figure supplement 4** with indication of the cropped areas.

Figure supplement 5. Heme depletion does not impact the interaction between GFP-Globin and CaM.

Figure supplement 5—source data 1. Original uncropped immunoblots of **Figure 6—figure supplement 5D** with indication of the cropped areas.

Figure supplement 5—source data 2. Original uncropped immunoblots of **Figure 6—figure supplement 5E** with indication of the cropped areas.

that Adgb is transcriptionally regulated by FoxJ1, a master regulator of ciliogenesis (Koay et al., 2021).

A robust ADGB-SEPT10 interaction was also detected by co-IP experiments in transiently co-transfected HEK293 cells. The interaction persisted despite various mutations of putative binding sites within ADGB, suggesting that the interaction occurs at different locations along the two proteins and that the proteins may entangle around each other. Indeed, it is not uncommon that protein-protein interactions involve multiple interaction surfaces, as referred to in a recent study (Egri et al., 2022). Since ADGB contains multiple larger domains of uncharacterized function, it is conceivable that various sites could contribute to interaction with proteins. The close ADGB-SEPT10 interaction may serve as targeting mechanism for the proteolytic processing of SEPT10 that we could observe upon ectopic expression of both proteins. The unique chimeric domain structure of Adgb with an N-terminal calpain-like protease domain and the presence of an IQ motif suggests a CaM-mediated regulation. CaM-bound calcium represents a crucial activator of proteolytic activity of Ca²⁺-dependent calpain proteases (Bähler and Rhoads, 2002; Villalobo et al., 2019). An interesting observation in our study is the interaction of CaM and ADGB upon isolation of the globin domain but not with the full-length ADGB protein. The lack of interaction upon overexpression of a full-length ADGB might be due to interference by the larger protein (particularly the 700 amino acid long C-terminal region of unidentified origin). As such the globin domain might be liberated possibly by proteolytic cleavage. This hypothesis is consistent with our previous observations of time-dependent truncation

of full-length ADGB from baculovirally infected *Spodoptera frugiperda* (Sf9) insect cells (Bracke et al., 2018). Similar truncations are detected upon overexpression of full-length ADGB in HEK293 cells in the current study (e.g. Figure 6A). Furthermore, this observation is consistent with the nature of Ca²⁺-dependent calpains, displaying frequent autolytic activation by auto-proteolysis (Ermolova et al., 2015; Osako et al., 2010). Additional experiments would be required to prove these hypotheses. Importantly, the interaction between CaM and the ADGB IQ-binding motif, as well as presence of the calpain protease domain, seems pivotal in the observed proteolytic cleavage of SEPT10 as the cleavage product completely disappeared upon overexpression of either IQ or protease deletion constructs. Calpain-dependent proteolytic cleavage of septins is not unprecedented; it was shown that Sept5 is a substrate of both calpain-1 and calpain-2 in platelets, where the cleavage triggers secretion of chemokine-containing granules (Randriamboavonjy et al., 2012). Septins are sensitive to molecular modifications, such as SUMOylation which impacts Sept6, 7, and 11-dependent filament formation (Ribet et al., 2017), or to PKA-dependent phosphorylation as reported for Sept12, leading to its dissociation from the septin complex and disruption of filament formation (Shen et al., 2017). It is therefore conceivable that CaM-dependent Adgb-mediated proteolytic cleavage of Sept10 may be a prerequisite for proper Sept10 function or localization within the sperm neck or annulus.

In conclusion, our study is the first to demonstrate a functional role for Adgb, the fifth mammalian globin. We present convincing *in vivo* evidence that Adgb is required for murine spermatogenesis. Adgb is necessary for sperm head shaping and for proper microtubule and flagellum formation. Our *in vitro* data illustrate CaM binding to ADGB and suggest that ADGB contributes to proteolytical cleavage of SEPT10 in a CaM-dependent manner. Our work provides a crucial contribution to the characterization of the physiological role of this novel enigmatic chimeric globin type.

Materials and methods

Animals, ethics statement, and genotyping

All experimental procedures and animal maintenance followed Swiss federal guidelines, and the study was revised and approved by the 'Service de la sécurité alimentaire et des affaires vétérinaires' (SAAV) of the canton of Fribourg, Switzerland (license number 2017_16_FR). Animals were housed in rooms with a 12 hr/12 hr light/dark cycle, controlled temperature and humidity levels and had free access to food and water. Interbreeding of heterozygous animals was performed to obtain wild-type (+/+), heterozygous (Tg/+ for tm1a and tm1b or +/- for tm1d), and homozygous/knockout (Tg/Tg for tm1a and tm1b or -/- for tm1d) littermates that were experimentally used, if not otherwise stated, between 3 and 9 months of age. Genotyping of tm1a and tm1b animals was performed using the following primers (Figure 1—figure supplement 1): F1 5'-CCGTGCCCAGCTATATGAGT-3'; R1 5'-CACAACGG GTTCTTCTGTAGTCC-3'; R2 5'-CCAGCGGTGTTCTTTCTTA-3'. Primers for tm1d genotyping were the following (Figure 1—figure supplement 1): F1, R2, and R4 5'-ACTGATGGCGAGCTCAGACC-3'. PCR amplification was performed for 36 cycles of 1 min at 95°C, 1 min at 56°C, and 1 min at 72°C. The PCR products were separated by electrophoresis on 2% agarose gels and visualized by ethidium bromide staining.

Gene targeting and knockout mouse generation

The *Adgb*^{tm1a(KOMP)Wtsi} (tm1a) strain was generated by blastocyst microinjection of ESC clone EPD0707_3_H06, provided by the KOMP (Skarnes et al., 2011). Correct targeting of the *Adgb* locus was verified prior to microinjection by long-range PCR using primers 5'S 5'-CTGTACTGGTTGTACTACTGGTCAACTG-3'; 5'AS 5'-GGACTAACAGAAGAACCCGTTGTG-3'; 3'S 5'-CACACCTCCCCCTGAACCTGAAAC-3'; 3'AS 5'-GTACTTGATTGGACGATGATCCAAG-3' (Figure 1—figure supplement 1), generating a band of 6.7 kb for 5' primers and 5.1 kb for 3' primers (Figure 1—figure supplement 1). Targeted clones were confirmed by Southern blot analysis using a hybridization probe that targets exon 13 (Figure 1—figure supplement 1) revealing a band of 4.1 kb (wild type) or 3.2 kb (tm1a allele) following digestion of genomic DNA with *PvuII* and a band of 2.8 kb (wild type) or 2.4 kb (tm1a allele) following digestion of genomic DNA with *PstI* (Figure 1—figure supplement 1). Chimeric mice were bred with C57BL/6-Tyr^{c-Brd} mice, and germline transmission in the F1 offspring was verified by PCR using primers F1, R1, and R2 (Figure 1—figure supplement 1). The mice were further bred to C57BL/6N-Hprt^{Tg(CMV-cre)Brd/Wtsi} transgenic mice expressing the Cre allele to delete exons 13 and 14 and

the neo cassette to generate the *Adgb*^{tm1b(KOMP)Wtsi} strain (tm1b), or to C57BL/6N-Gt(ROSA)26Sor^{tm1(FLP1)}_{Dym/Wtsi} transgenic mice expressing the Flp recombinase to delete the whole transgene cassette, thereby generating the *Adgb*^{tm1c(KOMP)Wtsi} strain. The latter were further crossed with C57BL/6N-Hprt^{Tg(CMV-cre)}_{Brd/Wtsi} transgenic mice to delete exons 13 and 14, thereby generating the *Adgb*^{tm1d(KOMP)Wtsi} (tm1d, knockout) strain. The Cre recombinase allele was bred out before any experiments were performed.

RNA extraction and RT-qPCR

Testes were frozen in liquid nitrogen and stored at −80°C. Tissues were homogenized using a Tissue-Lyser (Qiagen, Valencia, CA, USA). Subsequent RNA isolation and cDNA synthesis were performed as described previously (Keppner et al., 2019). In brief, RNA was extracted using an RNeasy Mini Kit (Qiagen) and reverse transcription was performed with 1.5 µg of total RNA and PrimeScript reverse transcriptase (Takara Bio Inc, Kusatsu, Japan). RT-qPCR was performed on a CFX96 C1000 real-time PCR cycler (Bio-Rad Laboratories, Hercules, CA) using SYBRgreen PCR master mix (Kapa Biosystems, London, UK). 21.5 ng of cDNA were loaded, and each sample was run as duplicate. mRNA levels were normalized to β-actin as previously described (De Backer et al., 2021). Primer sequences are displayed in **Supplementary file 1**.

Cell culture and transfection

HEK293 and A375 (ATCC CRL-1619) cells were maintained in Dulbecco's Minimum Essential Media (DMEM) (Gibco, Life Technologies, Carlsbad, CA, USA), containing L-glutamine, supplemented with 10% heat-inactivated fetal bovine serum (FBS; PAN Biotech, Aidenbach, Germany) and 100 Units/mL penicillin/100 µg/mL streptomycin (Gibco, Life Technologies, Carlsbad, CA, USA). Both cell lines were incubated in a humidified 5% CO₂ atmosphere at 37°C and were routinely subcultured after trypsinization. For hypoxic experiments, cells were seeded out in six-well plates or 100 mm culture dishes. On the subsequent day, hypoxia experiments were carried out at 0.2% O₂ and 5% CO₂ in a gas-controlled glove box (InviO2 400, Baker Ruskinn, Bridgend, UK) for 24 hr. Transfection of HEK293 cells was performed using calcium-phosphate (Jordan et al., 1996), with 2 µg of plasmid DNA for regular immunoblotting experiments and 5 µg of plasmid DNA for immunoprecipitation experiments. Briefly, the DNA was diluted in sterile water and mixed with 250 mM CaCl₂. 25 µM chloroquine was added to the cells and allowed to incubate for a minimum of 20 min. Prewarmed 37°C HBS buffer pH 7.05 (NaCl [280 mM], KCl [10 mM], Na₂HPO₄ [1.5 mM], D-glucose [12 mM], and HEPES [50 mM]) was added to the DNA solution (50% v/v), and the transfection mixture was added dropwise to the cells. The medium was replaced after 6 hr. Transfection of A375 cells was performed using JetOptimus (Polyplus-transfection SA, Illkirch-Graffenstaden, France) according to the manufacturer's instructions.

Heme depletion

To deprive HEK293T cells of intra- and extracellular heme, transfected HEK293T cells were cultured in DMEM medium supplemented with heme-depleted FBS along with 3 mM succinylacetone to inhibit intracellular heme synthesis. Heme depletion in FBS was carried out by incubation with 10 mM ascorbic acid at 37°C for 7 hr, with subsequent dialysis in PBS for three times (Chen et al., 2012). This treatment was carried out for 24 hr prior to cell lysis and co-IP.

SDS-PAGE and immunoblotting

Tissues were homogenized, and proteins for immunoblotting, immunoprecipitation (IP), and proteomics were extracted as described (Keppner et al., 2019). Briefly, tissues were harvested and snap-frozen in liquid nitrogen. The tissues were homogenized using a TissueLyser (Qiagen Valencia, CA, USA) in 1 mL protein extraction buffer (Tris-HCl [50 mM], EDTA [1 mM], EGTA [1 mM], sucrose [0.27 mM], leupeptin [2 µg/mL], aprotinin [2 µg/mL], pepstatin [2 µg/mL], PMSF [1 mM], NaF [50 mM], Na-pyrophosphate [5 mM], and Na₃VO₄ [1 mM]). The samples were centrifuged, the supernatant collected, and proteins quantified by Bradford assay. Cells were lysed in triton buffer (Tris-HCl [20 mM, pH 7.4], NaCl [150 mM], and triton X-100 [1%]), left on ice for 15 min, and centrifuged, and the proteins were quantified by Bradford assay.

25 µg of proteins were separated by SDS-PAGE on 10% gels, and proteins were electrotransferred to nitrocellulose membranes (Amersham Hybond-ECL, GE Healthcare, Chicago, IL, USA). The membranes were incubated overnight at 4°C with primary antibody (**Supplementary file 2**) and for

1 hr with donkey anti-rabbit or anti-mouse IgG HRP-conjugated secondary antibody (1:5000, Amersham, Buckinghamshire, UK). All antibodies were diluted in TBS-tween (1%) and dried milk (1%). The signal was revealed using ECL Prime (Amersham, Buckinghamshire, UK) on a C-DiGit Western blot scanner (LI-COR Biosciences) and quantified using ImageStudio program (LI-COR Biosciences, Lincoln, NE, USA). The polyclonal anti-Adgb antibody was custom-made (Proteintech Group Inc, Rosemont, IL, USA). A fusion protein immunogen raised against the 409–745 amino acid region of mouse Adgb was used for the immunization of two rabbits over a period of 102 days. The antibodies in immune sera were affinity purified. Pre-bleeds, test bleeds, and purified antibodies were tested and validated by immunoblotting on wild-type and knockout testis extracts.

Immunoprecipitation

For immunoprecipitation (IP) of subsequent LC-MS/MS and immunoblotting analyses, 4 and 2 mg of proteins were used, respectively. The protein lysates were first pre-cleared for 24 hr at 4°C with protein G-sepharose beads (GE Healthcare, Chicago, IL, USA) coupled to rabbit IgG (Bethyl Laboratories Inc, Montgomery, TX, USA). Samples were incubated overnight at 4°C with 2 µg primary antibody (**Supplementary file 2**) or 2 µg rabbit IgG, followed by 4 hr with protein G-sepharose beads, then washed two times with wash buffer (Tris-HCl [20 mM, pH 7.4], NaCl [300 mM], and LAP [1 mM]) and three times with equilibration buffer (Tris-HCl [20 mM, pH 7.4], NaCl [150 mM], and LAP [1 mM]). Samples were eluted by boiling for 5 min at 95°C in 2× sample buffer and separated from the beads by centrifugation.

LC-MS/MS analysis

Washed IP beads were incubated with Laemmli sample buffer (*Laemmli, 1970*), and proteins were reduced with 1 mM DTT for 10 min at 75°C and alkylated using 5.5 mM iodoacetamide for 10 min at room temperature (RT). Protein samples were separated by SDS-PAGE on 4–12% gradient gels (ExpressPlus, Genscript, New Jersey, NJ, USA). Each gel lane was cut into six equal slices, the proteins were in-gel digested with trypsin (Promega, Madison, WI, USA), and the resulting peptide mixtures were processed on StageTips (*Rappsilber et al., 2007; Shevchenko et al., 2006*).

LC-MS/MS measurements were performed on a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) coupled to an EASY-nLC 1000 nanoflow HPLC (Thermo Fisher Scientific). HPLC-column tips (fused silica) with 75 µm inner diameter were packed with Reprosil-Pur 120 C18-AQ, 1.9 µm (Dr. Maisch GmbH, Ammerbuch, Germany) to a length of 20 cm. A gradient of solvents A (0.1% formic acid in water) and B (0.1% formic acid in 80% acetonitrile in water) with increasing organic proportion was used for peptide separation (loading of sample with 0% B; separation ramp: from 5 to 30% B within 85 min). The flow rate was 250 nL/min and for sample application 650 nL/min. The mass spectrometer was operated in the data-dependent mode and switched automatically between MS (max. of 1×10^6 ions) and MS/MS. Each MS scan was followed by a maximum of 10 MS/MS scans using normalized collision energy of 25% and a target value of 1000. Parent ions with a charge state form $z=1$, and unassigned charge states were excluded from fragmentation. The mass range for MS was $m/z=370$ –1750. The resolution for MS was set to 70,000 and for MS/MS to 17,500. MS parameters were as follows: spray voltage 2.3 kV; no sheath and auxiliary gas flow; ion-transfer tube temperature 250°C. The MS raw data files were uploaded into the MaxQuant software version 1.6.2.10 for peak detection, generation of peak lists of mass error corrected peptides, and for database searches (*Tyanova et al., 2016*). A full-length UniProt mouse (based on UniProt FASTA version April 2016) or human database (UniProt FASTA version March 2016) additionally containing common contaminants, such as keratins and enzymes used for in-gel digestion, was used as reference. Carbamidomethylcysteine was set as fixed modification and protein amino-terminal acetylation and oxidation of methionine were set as variable modifications. Three missed cleavages were allowed, enzyme specificity was trypsin/P, and the MS/MS tolerance was set to 20 ppm. The average mass precision of identified peptides was in general less than 1 ppm after recalibration. Peptide lists were further used by MaxQuant to identify and relatively quantify proteins using the following parameters: peptide and protein false discovery rates (FDRs), based on a forward-reverse database, were set to 0.01, minimum peptide length was set to 7, minimum number of peptides for identification and quantitation of proteins was set to one which must be unique. The ‘match-between-run’ option (0.7 min) was used.

Propidium iodide staining and flow cytometry

Preparation of germ cell suspensions was achieved as described (Jeyaraj *et al.*, 2003). Briefly, decapsulated testes were incubated in 0.5 mg/mL collagenase type IV in PBS, washed with PBS, and incubated in 1 µg/mL DNase and 1 µg/mL trypsin. Soybean trypsin inhibitor was added, the suspension was filtered, washed in PBS, fixed with 70% ethanol, and stored at 4°C. DNA staining using propidium iodide was performed as described (Krishnamurthy *et al.*, 2000). Propidium iodide-stained cells were analyzed in a FACScan flow cytometer (Becton-Dickinson Immunocytometry, San Jose, CA, USA). The fluorescent signals of propidium iodide-stained cells were recorded, and a cytogram of DNA area vs. cell count was used to select cell populations on the basis of their DNA content. A total of 10,000 events was recorded for each histogram. Cell populations were selected based on their DNA content, and their relative numbers were calculated using Summit (Cytomation, CO, USA). Gating was performed as such that quantification represents the ratio of cell size to propidium iodide stainable nuclear DNA content.

Histological analyses and immunofluorescence

Testes were fixed in 4% paraformaldehyde (PFA) and embedded in paraffin. Preparation of sections and H&E staining was performed as described (Keppner *et al.*, 2015). Pictures were taken using a Nikon Eclipse microscope (Nikon Corporation, Tokyo, Japan). For immunofluorescence, testes were fixed in 4% PFA for at least 1 week and incubated in 30% sucrose for another week. The testes were embedded in Optimal Cutting Temperature compound (O.C.T. Tissue-Tek, Sakura Finetek, Tokyo, Japan), and 5 µm thick sections were cut using a cryotome. For seminiferous tubule dissections and stainings, slides were prepared as previously described (Kotaja *et al.*, 2004). For sperm stainings, cauda epididymal sperm was retrieved and diluted in PBS. A drop of the suspension was smeared on glass slides and fixed by drying for 15 min and by 4% PFA for 20 min. The slides were blocked in 10% normal goat serum and 0.5% triton X-100 for 1 hr. Testis sections and sperm slides were incubated overnight at 4°C with primary antibodies (Supplementary file 2) in 5% normal goat serum and 0.25% triton X-100. The slides were washed with PBS (3 × 10 min), incubated with secondary Alexa Fluor 488 or 594 coupled goat anti-mouse or anti-rabbit IgG (1:300, Invitrogen, Waltham, MA, USA) for 1 hr, washed again with PBS (3 × 10 min), and counterstained with Sudan Black (0.1%) for autofluorescence quenching. Slides were mounted with fluoromount mounting medium containing DAPI (Southern-Biotech, Birmingham, AL, USA) and visualized using a Nikon Eclipse fluorescent microscope (Nikon Corporation).

In situ mRNA hybridization

RNAscope *in situ* hybridization was performed using BaseScope Detection Reagent Kit v2 RED (Advanced Cell Diagnostics Inc, Newark, CA, USA, Cat. No. 323900) according to the manufacturer's instructions. H₂O₂ treatment, antigen retrieval, and protease treatment were performed on 5 µm-thick sections prior to hybridization with probes for Adgb (BA-Mm-Adgb-3zz-st, Advanced Cell Diagnostics, Cat. No. 862141), DapB as negative control (BA-DapB-3zz, Advanced Cell Diagnostics, Cat. No. 701011), and Ppib as positive control (Ba-Mm-Ppib-3zz, Advanced Cell Diagnostics, Cat. No. 701071) at 40°C for 2 hr followed by eight amplification steps. The signal was revealed with Fast Red, and the sections were counterstained with Gill's hematoxylin no. 1 and mounted with VectaMount (Vector Laboratories, Burlingame, CA, USA). The sections were visualized using a light microscope.

Electron microscopy

All electron microscopy experiments were performed at the Electron Microscopy Platform of the University of Lausanne, Switzerland. Mouse testes were fixed in 2.5% glutaraldehyde solution (EMS, Hatfield, PA, USA) in phosphate buffer (PB 0.1 M [pH 7.4]) for 1 hr at RT and post-fixed in a fresh mixture of osmium tetroxide 1% (EMS) with 1.5% of potassium ferrocyanide (Sigma, St. Louis, MO, USA) in PB buffer for 1 hr at RT. The samples were then washed twice in distilled water and dehydrated in acetone solution (Sigma) at graded concentrations (30%–40 min; 50%–40 min; 70%–40 min; 100%–2 × 1 hr). This was followed by infiltration in Epon resin (EMS, Hatfield, PA, USA) at graded concentrations (Epon 33% in acetone-4 hr; Epon 66% in acetone-4 hr; Epon 100%–2 × 8 hr) and finally polymerized for 48 hr at 60°C in an oven. Ultrathin sections of 50 nm were cut using a Leica Ultracut (Leica Mikrosysteme GmbH, Vienna, Austria), picked up on a copper slot grid of 2 × 1 mm (EMS,

Hatfield, PA, USA), and coated with a polystyrene film (Sigma, St Louis, MO, USA). Sections were post-stained with uranyl acetate (Sigma, St. Louis, MO, USA) 4% in H₂O for 10 min, rinsed several times with H₂O followed by Reynolds lead citrate in H₂O (Sigma, St Louis, MO, USA) for 10 min, and rinsed several times with H₂O. Micrographs were taken with a TEM FEI CM100 (FEI, Eindhoven, The Netherlands) at an acceleration voltage of 80 kV with a TVIPS TemCamF416 digital camera (TVIPS GmbH, Gauting, Germany).

RNAseq library preparation and transcriptome sequencing

Total RNA from two independent samples of wild-type and *Adgb*^{-/-} testis was extracted using the mirVana miRNA-Kit according to manufacturer's instructions (Life Technologies, Carlsbad, USA). Prior to library construction, RNA quality was assessed using an Agilent 2100 Bioanalyzer and the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA, USA). RNA was quantified using Qubit RNA BR Assay Kit (Invitrogen, Waltham, MA, USA). Libraries were prepared starting from 1000 ng of total RNA using the RNA Sample Prep Kit v2 (Illumina Inc, San Diego, CA, USA) including a poly-A selection step following the manufacturer's instructions and sequenced as 2 × 100 nt paired-end reads using an Illumina HiSeq 2500. Library preparation and sequencing were performed by the NGS Core Facility of the Department of Biology, Johannes-Gutenberg University (Mainz, Germany). RNAseq data are available from the European Nucleotide Archive under accession number PRJEB46499.

Differential gene expression, GO term annotation, and pathway enrichment analyses

Raw sequences were pre-processed to remove low quality reads and residual Illumina adapter sequences using BBduk from the BBtools suite (<https://sourceforge.net/projects/bbmap/>). The overall sequencing quality and the absence of adapter contamination were evaluated with FastQC. Mapping was performed with HISAT2, and quantification of gene expression was done using StringTie. Differentially expressed genes were determined using DESeq2. Genes were considered differentially expressed when presenting |fold change| > 2 and FDR-corrected p-value ≤ 0.1. GO term enrichment analyses were performed using WebGestalt 2019 using the Overrepresentation Enrichment Analysis method, requiring a BH-corrected p-value ≤ 0.05 and a minimum enrichment of four genes for term/pathway. Enrichment in canonical pathways were performed with Qiagen's Ingenuity Pathway Analysis (IPA, Qiagen, Hilden, Germany), Core analysis tool using bias-corrected z-score (when applicable), and BH-corrected p-values ≤ 0.05.

Cloning and construction of expression plasmids

Generation of pLenti6-ADGB was described before (Bracke et al., 2018), pLenti6-SEPT10-V5 was purchased from DNASU (clone ID HsCD00943271, DNASU Plasmid Repository, Arizona State University, AZ, USA). All additional recombinant genes were cloned into pFLAG-CMVTM-6a expression vector (Sigma) unless otherwise specified. All coding sequences were amplified by PCR using Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific). Recombinant ADGB, SEPT10, and SEPT7 with N-terminal FLAG tag and C-terminal myc tag were constructed by amplifying and ligating their respective coding sequence with in-primer designed myc tag (EQKLISEEDL) into the expression vector, in-frame with the N-terminal FLAG tag. A glycine-serine (GSG) linker was added between the last codon of SEPT10 and the first codon of the myc tag. SEPT2, SEPT11, and SEPT12 expression vectors were cloned by amplifying and ligating their respective coding sequence into the pFLAG-CMV-6a expression vector containing an N-terminal FLAG tag. Truncated ADGB proteins consisting of the calpain-like domain, 350-residue uncharacterized domain and globin domain (N-terminal mutant), or the 700-residue uncharacterized region (C-terminal mutant) domain were designed with GFP tags at both N- and C-termini. The N-terminal mutant ADGB was amplified between codons of residues 58 and 968 in ADGB, while the C-terminal mutant ADGB was amplified between codons of residues 969 and 1667. Amplicons were designed with 5'- and 3'- overhangs compatible with two customized GFP amplicons designed to anneal at the 5'- and 3'-ends of the genes. GSG linkers (GGSGGGGSGG) were added to bridge the GFP tags and the truncated ADGB proteins. Similarly, N-terminally GFP-tagged isolated ADGB globin domain was cloned by amplifying and ligating amplicons of the ADGB globin coding sequence downstream to a GFP coding sequence with complementary overhangs, with the GSG linker added between the two proteins. With the same construction, GFP-tagged ADGB

globin domains with a single mutation on the proximal histidine in helix F codon 8 (H824G) or the distal glutamine in helix E codon 7 (Q792G), or both (H824G/Q792G) were cloned by amplifying and ligating the globin domain coding sequence using primers designed to carry the mutated codon sequence. ADGB- Δ IQ, ADGB- Δ calpain, and ADGB- Δ CCD were constructed by amplifying designed ADGB amplicons with compatible overhangs and were ligated in-frame to generate an ADGB coding sequence with deletions in the desired domains. For M2H assays, Gal4-CaM, Gal4-ADGB, VP16-globin domain, VP16-CaM, or VP16-ADGB were cloned into a pcDNA3.0 expression vector. Gal4 DNA-binding domain or VP16 transactivation domain sequences were amplified with the in-primer designed GSG linker at the 3'-end of the amplicons. The coding sequence of *CALM3*, ADGB globin domain, and ADGB full length were amplified with complementary 5'-end and ligated to the Gal4 and VP16 sequences to generate the fusion genes.

Reporter gene assays

For M2H assays, 2.15×10^5 HEK293 or 4×10^5 A375 cells were transiently transfected with 1 μ g firefly luciferase reporter plasmid (5xGAL4-TATA-luciferase, Addgene, 46756) (Sun et al., 1994) and 500 ng or 200 and 300 ng chimeric Gal4 and VP16 fusion protein vectors, respectively, in 12- or 6-well format using CaCl_2 or JetOptimus. To control for differences in transfection efficiency and extract preparation, 25 ng or 50 ng pRL-SV40 *Renilla* luciferase reporter vector (Promega, Madison, WI, USA) was co-transfected, respectively for HEK293 and A375 cells. Cultures were evenly split onto 12-well plates 24 hr after transfection for A375 cells. For hypoxia control experiments, 4×10^5 A375 cells were transiently co-transfected with 500 ng firefly 5'/3'-hypoxia response element-dependent *EPO* promoter-driven luciferase reporter plasmid (Storti et al., 2014) and 50 ng pRL-SV40 *Renilla* luciferase reporter vector. Luciferase activities of duplicate wells were determined using the Dual Luciferase Reporter Assay System (Promega) as described before (Schörg et al., 2015). Reporter activities were expressed as relative firefly/*Renilla* luciferase activities. All reporter gene assays were performed at least three times independently.

Testosterone quantification

For intra-testis testosterone quantification, testis samples were homogenized twice using the Precellys 24 tissue homogenizer (Bertin Instruments; Rockville, MD, USA) (4°C, 3 $\times$, 30 s at 6500 rpm, cycle break 30 s) in chloroform-isopropanol (1 mL, 50/50%; v/v) containing ISTD. Combined supernatants were centrifuged (10 min, RT, 16,000 $\times$ g) and evaporated using a Genevac EZ-2 evaporator (Stepbios, Muttentz, Switzerland) (3 hr, 35°C). Samples were reconstituted in methanol (50 μ L, 10 min, RT, 1300 rpm) and sonicated (10 min, RT). Reconstituted samples were centrifuged (10 min, RT, 16,000 $\times$ g), and supernatants were transferred to LC-MS vials. Testosterone content was analyzed by ultra-performance liquid chromatography-MS/MS (UPLC-MS/MS) using an Agilent 1290 Infinity II UPLC coupled to an Agilent 6495 triple quadrupole mass spectrometer equipped with a jet-stream electrospray ionization interface (Agilent Technologies). Analyte separation was achieved using a reverse-phase column (1.7 μ m, 2.1 mm $\times$ 150 mm; Acquity UPLC BEH C18; Waters). Data acquisition and quantitative analysis were performed by MassHunter (Version B.10.0. Build 10.0.27, Agilent Technologies).

TUNEL assay

TUNEL assay was performed on paraffin-embedded testis sections using the DeadEnd Fluorometric TUNEL System according to manufacturer's instructions (Promega, Madison, WI, USA). The positive control sections were pre-incubated with DNase I solution (Qiagen, Valencia, CA, USA), while negative control sections were incubated in incubation buffer without recombinant terminal deoxynucleotidyl transferase (rTdT enzyme). Slides were mounted with fluoromount mounting medium containing DAPI (SouthernBiotech, Birmingham, AL, USA) and visualized using a Nikon Eclipse fluorescent microscope (Nikon Corporation). 10 pictures were taken for each animal and TUNEL-positive cells were counted using FIJI software (Schindelin et al., 2012).

ELISA quantification of LH and FSH

Serum LH and FSH were quantified by ELISA according to manufacturer's instructions using 50 μ L of sample (Cloud-Clone Corp., Houston, TX, USA).

Statistical analysis

All values were presented as mean $\pm$ SEM. Differences in means between two groups were analyzed with unpaired two-tailed Student's t-test (**Figure 1A, C, D and E; Figure 6H** right graph; **Figure 1—figure supplement 2A,B Figure 1—figure supplement 3A–G,I Figure 4—figure supplement 1H–N Figure 6—figure supplement 5A,B,C**) and those among multiple groups with one-way ANOVA followed by Tukey post-hoc test (**Figure 2D; Figure 6D, G and H** left graph, **Figure 6—figure supplement 2B**). All statistics were performed with GraphPad Prism software 7.05. Values of $p \leq 0.05$ were considered statistically significant.

Acknowledgements

We thank Christine Roulin for technical assistance. We thank Damien De Bellis from the Electron Microscopy Platform of the University of Lausanne for EM section preparation and image acquisition. This work was supported by the Swiss National Science Foundation to DH (grants 31,003 A_173000 and 310030_207460) and the German Research Foundation to DH (HO 5837/1–1) and TH (HA 2103/9–1).

Additional information

Funding

Funder	Grant reference number	Author
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung	31003A_173000	David Hoogewijs
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung	310030_207460	David Hoogewijs
Deutsche Forschungsgemeinschaft	HO 5837/1-1	David Hoogewijs
Deutsche Forschungsgemeinschaft	HA 2103/9-1	Thomas Hankeln

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Anna Keppner, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft; Miguel Correia, Formal analysis, Investigation, Visualization; Sara Santambrogio, Teng Wei Koay, Darko Maric, Denise V Winter, Frédéric Chalmel, Formal analysis, Investigation, Methodology; Carina Osterhof, Formal analysis, Investigation, Methodology, Visualization; Angèle Clerc, Investigation; Michael Stumpe, Dieter Kressler, Data curation, Formal analysis, Methodology; Sylvia Dewilde, Formal analysis; Alex Odermatt, Formal analysis, Methodology, Resources; Thomas Hankeln, Data curation, Formal analysis, Funding acquisition, Methodology; Roland H Wenger, Formal analysis, Investigation, Resources, Writing - review and editing; David Hoogewijs, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft

Author ORCIDs

Carina Osterhof  <http://orcid.org/0000-0002-1699-7410>

Michael Stumpe  <http://orcid.org/0000-0002-9443-9326>

Dieter Kressler  <http://orcid.org/0000-0003-4855-3563>

David Hoogewijs  <http://orcid.org/0000-0001-5547-6004>

Ethics

All experimental procedures and animal maintenance followed Swiss federal guidelines and the study was revised and approved by the "Service de la sécurité alimentaire et des affaires vétérinaires" (SAAV) of the canton of Fribourg, Switzerland (license number 2017_16_FR).

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.72374.sa1>

Author response <https://doi.org/10.7554/eLife.72374.sa2>

Additional files

Supplementary files

- Transparent reporting form
- Supplementary file 1. List of primers used for RT-qPCR.
- Supplementary file 2. List of antibodies used throughout the study.

Data availability

RNA-sequencing data have been submitted to ENA with accession number PRJEB46499 and is also available as supplemental dataset 1 (excel table). All data generated or analysed during this study are included in the manuscript and supporting files. Source data files are provided for Figures 1, 2, 4B, 4C, 4D, 4E, 6A, 6B, 6C, 6D, 6E, 6F, Fig. 1-fig. suppl. 1, Fig. 4-fig. suppl. 1A, B, C, D, E, F, G, Fig. 4-fig. suppl. 2A, B, C, Fig. 4-fig. suppl. 3A, B, C, D, E, Fig. 4-fig. suppl. 5, Fig. 4-fig. suppl. 6B, Fig. 6-fig. suppl. 1A, Fig. 6-fig. suppl. 4, Fig. 6-fig. suppl. 5D, E.

References

- Andrews C, Xu Y, Kirberger M, Yang JJ. 2020. Structural Aspects and Prediction of Calmodulin-Binding Proteins. *International Journal of Molecular Sciences* **22**:E308. DOI: <https://doi.org/10.3390/ijms22010308>, PMID: 33396740
- Avidor-Reiss T, Ha A, Basiri ML. 2017. Transition Zone Migration: A Mechanism for Cytoplasmic Ciliogenesis and Postaxonemal Centriole Elongation. *Cold Spring Harbor Perspectives in Biology* **9**:a028142. DOI: <https://doi.org/10.1101/cshperspect.a028142>, PMID: 28108487
- Avidor-Reiss T, Carr A, Fishman EL. 2020. The sperm centrioles. *Molecular and Cellular Endocrinology* **518**:110987. DOI: <https://doi.org/10.1016/j.mce.2020.110987>, PMID: 32810575
- Bähler M, Rhoads A. 2002. Calmodulin signaling via the IQ motif. *FEBS Letters* **513**:107–113. DOI: [https://doi.org/10.1016/s0014-5793\(01\)03239-2](https://doi.org/10.1016/s0014-5793(01)03239-2), PMID: 11911888
- Bracke A, Hoogewijs D, Dewilde S. 2018. Exploring three different expression systems for recombinant expression of globins: *Escherichia coli*, *Pichia pastoris* and *Spodoptera frugiperda*. *Analytical Biochemistry* **543**:62–70. DOI: <https://doi.org/10.1016/j.ab.2017.11.027>, PMID: 29203135
- Brown JM, Witman GB. 2014. Cilia and Diseases. *Bioscience* **64**:1126–1137. DOI: <https://doi.org/10.1093/biosci/biu174>, PMID: 25960570
- Chen CY, Samuel TK, Krause M, Dailey HA, Hamza I. 2012. Heme utilization in the *Caenorhabditis elegans* hypodermal cells is facilitated by heme-responsive gene-2. *The Journal of Biological Chemistry* **287**:9601–9612. DOI: <https://doi.org/10.1074/jbc.M111.307694>, PMID: 22303006
- Darde TA, Sallou O, Becker E, Evrard B, Monjeaud C, Le Bras Y, Jégou B, Collin O, Rolland AD, Chalmel F. 2015. The ReproGenomics Viewer: an integrative cross-species toolbox for the reproductive science community. *Nucleic Acids Research* **43**:W109–W116. DOI: <https://doi.org/10.1093/nar/gkv345>, PMID: 25883147
- Darde TA, Lecluze E, Lardenois A, Stévant I, Alary N, Tüttelmann F, Collin O, Nef S, Jégou B, Rolland AD, Chalmel F, Kelso J. 2019. The ReproGenomics Viewer: a multi-omics and cross-species resource compatible with single-cell studies for the reproductive science community. *Bioinformatics (Oxford, England)* **35**:3133–3139. DOI: <https://doi.org/10.1093/bioinformatics/btz047>, PMID: 30668675
- Darszon A, Nishigaki T, Beltran C, Treviño CL. 2011. Calcium channels in the development, maturation, and function of spermatozoa. *Physiological Reviews* **91**:1305–1355. DOI: <https://doi.org/10.1152/physrev.00028.2010>, PMID: 22013213
- De Backer J, Maric D, Bosman M, Dewilde S, Hoogewijs D. 2021. A reliable set of reference genes to normalize oxygen-dependent cytoglobin gene expression levels in melanoma. *Scientific Reports* **11**:10879. DOI: <https://doi.org/10.1038/s41598-021-90284-6>, PMID: 34035373
- Dolat L, Hu Q, Spiliotis ET. 2014. Septin functions in organ system physiology and pathology. *Biological Chemistry* **395**:123–141. DOI: <https://doi.org/10.1515/hsz-2013-0233>, PMID: 24114910
- Egri SB, Ouch C, Chou HT, Yu Z, Song K, Xu C, Shen K. 2022. Cryo-EM structures of the human GATOR1-Rag-Regulator complex reveal a spatial-constraint regulated GAP mechanism. *Molecular Cell* **82**:1836–1849. DOI: <https://doi.org/10.1016/j.molcel.2022.03.002>, PMID: 35338845

- Ermolova N**, Kramerova I, Spencer MJ. 2015. Autolytic activation of calpain 3 proteinase is facilitated by calmodulin protein. *The Journal of Biological Chemistry* **290**:996–1004. DOI: <https://doi.org/10.1074/jbc.M114.588780>, PMID: 25389288
- Fainberg J**, Kashanian JA. 2019. Recent advances in understanding and managing male infertility. *F1000Research* **8**:F1000. DOI: <https://doi.org/10.12688/f1000research.17076.1>, PMID: 31143441
- Ghossoub R**, Hu Q, Failler M, Rouyez M-C, Spitzbarth B, Mostowy S, Wolfrum U, Saunier S, Cossart P, Jameson W, Benmerah A. 2013. Septins 2, 7 and 9 and MAP4 colocalize along the axoneme in the primary cilium and control ciliary length. *Journal of Cell Science* **126**:2583–2594. DOI: <https://doi.org/10.1242/jcs.111377>, PMID: 23572511
- Green CD**, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ, Moore BB, Tallquist MD, Li JZ, Hammoud SS. 2018. A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Developmental Cell* **46**:651–667. DOI: <https://doi.org/10.1016/j.devcel.2018.07.025>, PMID: 30146481
- Hermo L**, Pelletier RM, Cyr DG, Smith CE, Pelletier RM. 2010a. Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. Part 1: background to spermatogenesis, spermatogonia, and spermatocytes. *Microscopy Research and Technique* **73**:241–278. DOI: <https://doi.org/10.1002/jemt.20783>, PMID: 19941293
- Hermo L**, Pelletier RM, Cyr DG, Smith CE, Pelletier RM. 2010b. Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. Part 2: changes in spermatid organelles associated with development of spermatozoa. *Microscopy Research and Technique* **73**:279–319. DOI: <https://doi.org/10.1002/jemt.20787>, PMID: 19941292
- Hermo L**, Pelletier RM, Cyr DG, Smith CE, Pelletier RM. 2010c. Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. Part 3: developmental changes in spermatid flagellum and cytoplasmic droplet and interaction of sperm with the zona pellucida and egg plasma membrane. *Microscopy Research and Technique* **73**:320–363. DOI: <https://doi.org/10.1002/jemt.20784>, PMID: 19941287
- Hermo L**, Pelletier RM, Cyr DG, Smith CE, Pelletier RM. 2010d. Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. Part 4: intercellular bridges, mitochondria, nuclear envelope, apoptosis, ubiquitination, membrane/voltage-gated channels, methylation/acetylation, and transcription factors. *Microscopy Research and Technique* **73**:364–408. DOI: <https://doi.org/10.1002/jemt.20785>, PMID: 19941288
- Hermo L**, Pelletier RM, Cyr DG, Smith CE, Pelletier RM. 2010e. Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. Part 5: intercellular junctions and contacts between germ cells and Sertoli cells and their regulatory interactions, testicular cholesterol, and genes/proteins associated with more than one germ cell generation. *Microscopy Research and Technique* **73**:409–494. DOI: <https://doi.org/10.1002/jemt.20786>, PMID: 19941291
- Hoogewijs D**, Ebner B, Germani F, Hoffmann FG, Fabrizio A, Moens L, Burmester T, Dewilde S, Storz JF, Vinogradov SN, Hankeln T. 2012. Androglobin: A Chimeric Globin in Metazoans That Is Preferentially Expressed in Mammalian Testes. *Molecular Biology and Evolution* **29**:1105–1114. DOI: <https://doi.org/10.1093/molbev/msr246>, PMID: 22115833
- Hou S**, Freitas T, Larsen RW, Piatibratov M, Sivozhlezov V, Yamamoto A, Meleshkevitch EA, Zimmer M, Ordal GW, Alam M. 2001. Globin-coupled sensors: A class of heme-containing sensors in Archaea and Bacteria. *PNAS* **98**:9353–9358. DOI: <https://doi.org/10.1073/pnas.161185598>, PMID: 11481493
- Hu Q**, Milenkovic L, Jin H, Scott MP, Nachury MV, Spiliotis ET, Nelson WJ. 2010. A septin diffusion barrier at the base of the primary cilium maintains ciliary membrane protein distribution. *Science (New York, N.Y.)* **329**:436–439. DOI: <https://doi.org/10.1126/science.1191054>, PMID: 20558667
- Ihara M**, Kinoshita A, Yamada S, Tanaka H, Tanigaki A, Kitano A, Goto M, Okubo K, Nishiyama H, Ogawa O, Takahashi C, Itohara S, Nishimune Y, Noda M, Kinoshita M. 2005. Cortical organization by the septin cytoskeleton is essential for structural and mechanical integrity of mammalian spermatozoa. *Developmental Cell* **8**:343–352. DOI: <https://doi.org/10.1016/j.devcel.2004.12.005>, PMID: 15737930
- Jan SZ**, Vorrer TL, Jongejan A, Röling M, Silber SJ, de Rooij DG, Hamer G, Repping S, van Pelt AMM. 2017. Unraveling transcriptome dynamics in human spermatogenesis. *Development (Cambridge, England)* **144**:3659–3673. DOI: <https://doi.org/10.1242/dev.152413>, PMID: 28935708
- Jégou B**, Sankararaman S, Rolland AD, Reich D, Chalmers F. 2017. Meiotic Genes Are Enriched in Regions of Reduced Archaic Ancestry. *Molecular Biology and Evolution* **34**:1974–1980. DOI: <https://doi.org/10.1093/molbev/msx141>, PMID: 28444387
- Jeyaraj DA**, Grossman G, Petrusz P. 2003. Dynamics of testicular germ cell apoptosis in normal mice and transgenic mice overexpressing rat androgen-binding protein. *Reproductive Biology and Endocrinology* **1**:48. DOI: <https://doi.org/10.1186/1477-7827-1-48>, PMID: 12831397
- Jordan M**, Schallhorn A, Wurm FM. 1996. Transfecting Mammalian Cells: Optimization of Critical Parameters Affecting Calcium-Phosphate Precipitate Formation. *Nucleic Acids Research* **24**:596–601. DOI: <https://doi.org/10.1093/nar/24.4.596>, PMID: 8604299
- Kaur P**, Nagar S, Bhagwat M, Uddin M, Zhu Y, Vancurova I, Vancura A, Singh PK. 2021. Activated heme synthesis regulates glycolysis and oxidative metabolism in breast and ovarian cancer cells. *PLOS ONE* **16**:e0260400. DOI: <https://doi.org/10.1371/journal.pone.0260400>
- Keppner Anna**, Andreasen D, Méritat A-M, Bapst J, Ansermet C, Wang Q, Maillard M, Malsure S, Nobile A, Hummler E, Cheng X. 2015. Epithelial Sodium Channel-Mediated Sodium Transport Is Not Dependent on the

- Membrane-Bound Serine Protease CAP2/Tmprss4. *PLOS ONE* **10**:e0135224. DOI: <https://doi.org/10.1371/journal.pone.0135224>
- Keppner A, Maric D, Sergi C, Ansermet C, De Bellis D, Kratschmar DV, Canonica J, Klusonova P, Fenton RA, Odermatt A, Crambert G, Hoogewijs D, Hummler E. 2019. Deletion of the serine protease CAP2/Tmprss4 leads to dysregulated renal water handling upon dietary potassium depletion. *Scientific Reports* **9**:19540. DOI: <https://doi.org/10.1038/s41598-019-55995-x>, PMID: 31863073
- Keppner A, Maric D, Correia M, Koay TW, Orlando IMC, Vinogradov SN, Hoogewijs D. 2020. Lessons from the post-genomic era: Globin diversity beyond oxygen binding and transport. *Redox Biology* **37**:101687. DOI: <https://doi.org/10.1016/j.redox.2020.101687>
- Khawar MB, Gao H, Li W. 2019. Mechanism of Acrosome Biogenesis in Mammals. *Frontiers in Cell and Developmental Biology* **7**:195. DOI: <https://doi.org/10.3389/fcell.2019.00195>, PMID: 31620437
- Kissel H, Georgescu MM, Larisch S, Manova K, Hunnicutt GR, Steller H. 2005. The Sept4 septin locus is required for sperm terminal differentiation in mice. *Developmental Cell* **8**:353–364. DOI: <https://doi.org/10.1016/j.devcel.2005.01.021>, PMID: 15737931
- Koay TW, Osterhof C, Orlando IMC, Keppner A, Andre D, Yousefian S, Suárez Alonso M, Correia M, Markworth R, Schödel J, Hankeln T, Hoogewijs D. 2021. Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis. *The Journal of Biological Chemistry* **296**:100291. DOI: <https://doi.org/10.1016/j.jbc.2021.100291>, PMID: 33453283
- Kotaja N, Kimmins S, Brancorsini S, Hentsch D, Vonesch JL, Davidson I, Parvinen M, Sassone-Corsi P. 2004. Preparation, isolation and characterization of stage-specific spermatogenic cells for cellular and molecular analysis. *Nature Methods* **1**:249–254. DOI: <https://doi.org/10.1038/nmeth1204-249>, PMID: 16144087
- Krishnamurthy H, Danilovich N, Morales CR, Sairam MR. 2000. Qualitative and quantitative decline in spermatogenesis of the follicle-stimulating hormone receptor knockout (FORKO) mouse. *Biology of Reproduction* **62**:1146–1159. DOI: <https://doi.org/10.1095/biolreprod62.5.1146>, PMID: 10775161
- Kuo YC, Lin YH, Chen HI, Wang YY, Chiou YW, Lin HH, Pan HA, Wu CM, Su SM, Hsu CC, Kuo PL. 2012. SEPT12 mutations cause male infertility with defective sperm annulus. *Human Mutation* **33**:710–719. DOI: <https://doi.org/10.1002/humu.22028>, PMID: 22275165
- Kwak S, Jung MW. 2019. Distinct roles of striatal direct and indirect pathways in value-based decision making. *eLife* **8**:e46050. DOI: <https://doi.org/10.7554/eLife.46050>, PMID: 31310237
- Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685. DOI: <https://doi.org/10.1038/227680a0>, PMID: 5432063
- Lin YH, Lin YM, Wang YY, Yu IS, Lin YW, Wang YH, Wu CM, Pan HA, Chao SC, Yen PH, Lin SW, Kuo PL. 2009. The expression level of septin12 is critical for spermiogenesis. *The American Journal of Pathology* **174**:1857–1868. DOI: <https://doi.org/10.2353/ajpath.2009.080955>, PMID: 19359518
- Lin YH, Huang CY, Ke CC, Wang YY, Lai TH, Liu HC, Ku WC, Chan CC, Lin YH. 2020. ACTN4 Mediates SEPT14 Mutation-Induced Sperm Head Defects. *Biomedicines* **8**:E518. DOI: <https://doi.org/10.3390/biomedicines8110518>, PMID: 33228246
- Lukassen S, Bosch E, Ekici AB, Winterpacht A. 2018. Characterization of germ cell differentiation in the male mouse through single-cell RNA sequencing. *Scientific Reports* **8**:6521. DOI: <https://doi.org/10.1038/s41598-018-24725-0>, PMID: 29695820
- Mecklenburg JM, Hermann BP. 2016. Mechanisms Regulating Spermatogonial Differentiation. *Results and Problems in Cell Differentiation* **58**:253–287. DOI: https://doi.org/10.1007/978-3-319-31973-5_10, PMID: 27300182
- Neto FTL, Bach PV, Najari BB, Li PS, Goldstein M. 2016. Spermatogenesis in humans and its affecting factors. *Seminars in Cell & Developmental Biology* **59**:10–26. DOI: <https://doi.org/10.1016/j.semcdb.2016.04.009>, PMID: 27143445
- Osako Y, Maemoto Y, Tanaka R, Suzuki H, Shibata H, Maki M. 2010. Autolytic activity of human calpain 7 is enhanced by ESCRT-III-related protein IST1 through MIT-MIM interaction. *The FEBS Journal* **277**:4412–4426. DOI: <https://doi.org/10.1111/j.1742-4658.2010.07822.x>, PMID: 20849418
- Platts AE, Dix DJ, Chemes HE, Thompson KE, Goodrich R, Rockett JC, Rawe VY, Quintana S, Diamond MP, Strader LF, Krawetz SA. 2007. Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. *Human Molecular Genetics* **16**:763–773. DOI: <https://doi.org/10.1093/hmg/ddm012>, PMID: 17327269
- Randriamboavonjy V, Isaak J, Elghezawy A, Pistrosch F, Frömel T, Yin X, Badenhop K, Heide H, Mayr M, Fleming I. 2012. Calpain inhibition stabilizes the platelet proteome and reactivity in diabetes. *Blood* **120**:415–423. DOI: <https://doi.org/10.1182/blood-2011-12-399980>, PMID: 22665935
- Rappsilber J, Mann M, Ishihama Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* **2**:1896–1906. DOI: <https://doi.org/10.1038/nprot.2007.261>, PMID: 17703201
- Ribet D, Boscaini S, Cauvin C, Siguier M, Mostowy S, Echard A, Cossart P. 2017. SUMOylation of human septins is critical for septin filament bundling and cytokinesis. *The Journal of Cell Biology* **216**:4041–4052. DOI: <https://doi.org/10.1083/jcb.201703096>, PMID: 29051266
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**:676–682. DOI: <https://doi.org/10.1038/nmeth.2019>, PMID: 22743772
- Schörg A, Santambrogio S, Platt JL, Schödel J, Lindenmeyer MT, Cohen CD, Schrödter K, Mole DR, Wenger RH, Hoogewijs D. 2015. Destruction of a distal hypoxia response element abolishes trans-activation of the PAG1

- gene mediated by HIF-independent chromatin looping. *Nucleic Acids Research* **43**:5810–5823. DOI: <https://doi.org/10.1093/nar/gkv506>, PMID: 26007655
- Shen YR**, Wang HY, Kuo YC, Shih SC, Hsu CH, Chen YR, Wu SR, Wang CY, Kuo PL. 2017. SEPT12 phosphorylation results in loss of the septin ring/sperm annulus, defective sperm motility and poor male fertility. *PLOS Genetics* **13**:e1006631. DOI: <https://doi.org/10.1371/journal.pgen.1006631>, PMID: 28346465
- Shen YR**, Wang HY, Tsai YC, Kuo YC, Wu SR, Wang CY, Kuo PL. 2020. The SEPT12 complex is required for the establishment of a functional sperm head-tail junction. *Molecular Human Reproduction* **26**:402–412. DOI: <https://doi.org/10.1093/molehr/gaaa031>, PMID: 32392324
- Shevchenko A**, Tomas H, Havlis J, Olsen JV, Mann M. 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* **1**:2856–2860. DOI: <https://doi.org/10.1038/nprot.2006.468>, PMID: 17406544
- Sigg MA**, Menchen T, Lee C, Johnson J, Jungnickel MK, Choksi SP, Garcia G, Busengdal H, Dougherty GW, Pennekamp P, Werner C, Rentzsch F, Florman HM, Krogan N, Wallingford JB, Omran H, Reiter JF. 2017. Evolutionary Proteomics Uncovers Ancient Associations of Cilia with Signaling Pathways. *Developmental Cell* **43**:744–762. DOI: <https://doi.org/10.1016/j.devcel.2017.11.014>, PMID: 29257953
- Skarnes WC**, Rosen B, West AP, Koutourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, Bradley A. 2011. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**:337–342. DOI: <https://doi.org/10.1038/nature10163>, PMID: 21677750
- Spiliotis ET**, Nakos K. 2021. Cellular functions of actin- and microtubule-associated septins. *Current Biology* **31**:R651–R666. DOI: <https://doi.org/10.1016/j.cub.2021.03.064>, PMID: 34033796
- Storti F**, Santambrogio S, Crowther LM, Otto T, Abreu-Rodríguez I, Kaufmann M, Hu CJ, Dame C, Fandrey J, Wenger RH, Hoogewijs D. 2014. A novel distal upstream hypoxia response element regulating oxygen-dependent erythropoietin gene expression. *Haematologica* **99**:e45–e48. DOI: <https://doi.org/10.3324/haematol.2013.102707>, PMID: 24510339
- Sun P**, Enslen H, Myung PS, Maurer RA. 1994. Differential activation of CREB by Ca²⁺/calmodulin-dependent protein kinases type II and type IV involves phosphorylation of a site that negatively regulates activity. *Genes & Development* **8**:2527–2539. DOI: <https://doi.org/10.1101/gad.8.21.2527>, PMID: 7958915
- Thijs L**, Vinck E, Bolli A, Trandafir F, Wan X, Hoogewijs D, Coletta M, Fago A, Weber RE, Van Doorslaer S, Ascenzi P, Alam M, Moens L, Dewilde S. 2007. Characterization of a globin-coupled oxygen sensor with a gene-regulating function. *The Journal of Biological Chemistry* **282**:37325–37340. DOI: <https://doi.org/10.1074/jbc.M705541200>, PMID: 17925395
- Toure A**, Rode B, Hunnicutt GR, Escalier D, Gacon G. 2011. Septins at the annulus of mammalian sperm. *Biological Chemistry* **392**:799–803. DOI: <https://doi.org/10.1515/BC.2011.074>, PMID: 21740329
- Tüttelmann F**, Ruckert C, Röpke A. 2018. Disorders of spermatogenesis: Perspectives for novel genetic diagnostics after 20 years of unchanged routine. *Medizinische Genetik* **30**:12–20. DOI: <https://doi.org/10.1007/s11825-018-0181-7>, PMID: 29527098
- Tyanova S**, Temu T, Cox J. 2016. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols* **11**:2301–2319. DOI: <https://doi.org/10.1038/nprot.2016.136>, PMID: 27809316
- Villalobo A**, González-Muñoz M, Berchtold MW. 2019. Proteins with calmodulin-like domains: structures and functional roles. *Cellular and Molecular Life Sciences* **76**:2299–2328. DOI: <https://doi.org/10.1007/s00018-019-03062-z>
- Wang M**, Liu X, Chang G, Chen Y, An G, Yan L, Gao S, Xu Y, Cui Y, Dong J, Chen Y, Fan X, Hu Y, Song K, Zhu X, Gao Y, Yao Z, Bian S, Hou Y, Lu J, et al. 2018. Single-Cell RNA Sequencing Analysis Reveals Sequential Cell Fate Transition during Human Spermatogenesis. *Cell Stem Cell* **23**:599–614. DOI: <https://doi.org/10.1016/j.stem.2018.08.007>, PMID: 30174296
- Wang YY**, Lai TH, Chen MF, Lee HL, Kuo PL, Lin YH. 2019. SEPT14 Mutations and Teratozoospermia: Genetic Effects on Sperm Head Morphology and DNA Integrity. *Journal of Clinical Medicine* **8**:E1297. DOI: <https://doi.org/10.3390/jcm8091297>, PMID: 31450874
- Yuan G**, Nanduri J, Bhasker CR, Semenza GL, Prabhakar NR. 2005. Ca²⁺/calmodulin kinase-dependent activation of hypoxia inducible factor 1 transcriptional activity in cells subjected to intermittent hypoxia. *The Journal of Biological Chemistry* **280**:4321–4328. DOI: <https://doi.org/10.1074/jbc.M407706200>, PMID: 15569687

2.3 A role of Androglobin in cancer? – insights from an overexpression system and transcriptome data mining

Carina Osterhof, [REDACTED]

[REDACTED]: The association of Androglobin with cancer – insights from an overexpression system. (unpublished)

Supplementary files: electronic Appendix

Own contributions to this publication:

- Planning and initial tests for generation of an A549 Adgb overexpression system. Stable cell lines were generated and analysed via RT-PCR and immunofluorescence under my supervision by MSc student [REDACTED] (2021).
- Cultivation of A549 cells. RNA isolation, quantification, and quality control from 3 Adgb OE samples as well as 3 empty control lines.
- Quality processing and mapping of mRNA-Seq data from OE system; quantification and subsequent differential gene expression analysis; interpretation of gene lists via gene ontology analysis and upstream regulator prediction.
- Phenotypic analysis of Adgb OE and empty control lines via migration assays. Images of migration assays were processed with help of [REDACTED].
- Reanalysis of interpretation of qRT-PCR data (carried out by [REDACTED]) from clinical lung and testis samples and comparison to reanalysed, publicly available transcriptomic data on cancer progression.
- Mining of public databases containing transcriptomic information on cell lines (CCLE, HPA) and clinical tumour samples from various tissues (TCGA); Correlation analysis with known regulatory factors.

Planning of experiments, analysis and interpretation of data as well as drafting of the manuscript were realized together with [REDACTED]. The project was managed by Prof. Dr. T. Hankeln.

A role of Andoglobin in cancer? – insights from an overexpression system and transcriptome data mining



Abstract

Androglobin is a phylogenetically ancient, multimeric protein with an embedded globin domain. Although this globin domain is circularly permuted and interrupted by a calmodulin IQ domain, it still displays heme binding properties and may therefore interact with gaseous signaling molecules. For other members of the globin gene family, such as NGB, CYGB and MB, either tumor suppressor or oncogenic functions have been proposed – depending on the tissue the malignancy originated from. Also for ADGB, a potential oncogenic effect was suggested derived from data generated in prostate and brain cancer cell models, but it remained unclear whether ADGB is also expressed in real tumor entities. We therefore investigated the expression pattern of ADGB in the context of *in vivo* cancer samples and its potential influence on tumor malignancy in a lung cancer model.

We analyzed a wide variety of transcriptomic datasets from cancer cell lines, tumors and studies on tumor progression and found that ADGB is systematically downregulated in both immortalized cell lines as well as cancer entities derived from endogenous ADGB expression sites. In the rare cases ADGB was expressed in cell lines, only few exons were transcribed or the ADGB gene locus was multiplied. To study the effect of ADGB on immortalized cells, we established a stable overexpression system in A549 lung cancer cells. Surprisingly, transcriptomic analysis of ADGB+ lung cancer cells showed higher motility and restructuring of the extracellular matrix – both hallmarks of elevated malignancy in cancer entities. Phenotypically, ADGB+ cells had longer cellular appendages, possibly reflecting the higher migratory capacity, but the effect was not pronounced enough to be observable in a migration assay. Thus, ADGB has oncogenic potential *in vitro*. These findings are in line with previous studies describing an oncogenic effect of ADGB in a glioma and pancreatic cancer cell model. However, our expression analysis on *in vivo* cancer entities shows that ADGB mRNA is simply not expressed or even actively silenced in malignancies. We therefore conclude that despite its potential oncogenic properties, ADGB is unsuitable as prognostic marker in cancer patients.

Introduction

Globins are an evolutionary old family of proteins, which form a hydrophobic pocket (the so-called globin-fold) that allows the binding of a heme prosthetic group. Vertebrate globins, such as Hemoglobin (Hb) or Myoglobin (Mb), are well known for their vital role in oxygen supply and transport (Wittenberg & Wittenberg, 2003). In addition, both have the ability to either produce or scavenge reactive oxygen species (ROS) depending on the concentration of oxygen in their surroundings (Flögel et al., 2001; Hendgen-Cotta et al., 2008). By this property, globins have been reported to promote or inhibit tumor growth. Several members of the globin family are in fact deregulated in human cancer entities and may even have prognostic value. The expression of Mb, for example, is upregulated in colon and breast cancer (Flonta et al., 2009; Gorr et al., 2011; Kristiansen et al., 2010). Depending on the tissue of origin, the presence of Mb is either associated with a favorable outcome (Breast cancer, Kristiansen et al., 2010; Head and neck squamous cell carcinoma, Meller et al., 2016) or with a more aggressive tumor phenotype (Glioma, Elsherbiny et al., 2021). Cytoglobin (Cygb), a vertebrate globin type mostly expressed in connective tissue cells of many organs, is mostly downregulated in tumors via hypermethylation of its promoters, e.g. in esophageal cancer, non-small cell lung cancer and head and neck squamous carcinoma (as reviewed in: Oleksiewicz et al., 2011). Knockout of Cygb in two independent mouse models showed a higher susceptibility for neoplasia in liver, lung, and colon (Thi Thanh Thuy et al., 2011; Yassin et al., 2018). In line with these findings, decreased Cygb expression was correlated with more frequent tumor recurrence and poor prognosis in glioma as well as in pancreatic ductal adenocarcinoma patients (Kono et al., 2021; Xu et al., 2013). For Neuroglobin (Ngb), a globin variant mostly expressed in neurons of the nervous system (Burmester et al., 2000), a dualistic function in cancer progression has been proposed. While it may act as a tumor suppressor in hepatocellular carcinoma (J. Zhang et al., 2013), its expression is upregulated in a variety of cancers (Emara et al., 2010). However, antibody free hybridization analysis could not reproduce this upregulation on the mRNA level (Gorr et al., 2011). In glioma and breast cancer models, endogenous Ngb has shown an anti-apoptotic function and promotes proliferation (Fiocchetti et al., 2018; B. Zhang et al., 2018). Together, these results suggest that globins may present promising candidates not only for cardiovascular related, but also for cancer research.

The most recently reported “novel” member of the vertebrate globin family is Androglobin (ADGB) (Hoogewijs et al., 2012). ADGB is structurally very different from the other family members, since its globin domain is embedded in a larger protein which carries high resemblance with a calpain protease. In addition, the ADGB globin domain is also permuted and interrupted by a putative calmodulin binding site. In analogy to other family members, the name was designed to hint at the predominant expression site of the gene, since ADGB was initially reported to be mainly expressed in later stages of spermatogenesis in mammalian testes (Hoogewijs et al., 2012). Subsequently, ADGB expression was found to be a hallmark of multi-ciliated cells from lung, brain and also the female reproductive tract, where its expression is probably regulated by FoxJ1, the master transcription factor of ciliogenesis (Koay et al., 2021). This strong association with ciliated cells suggests a direct function in ciliogenesis, however, ADGB’s exact mode of action is unclear. ADGB has also been studied in the context of cancer. An *in vitro* cell culture study suggested that ADGB could act as an oncogene in glioma formation (Huang et al., 2014). Another study proposed that epigenetic silencing of ADGB via the lncRNA STXBP5-AS1 reduces cell stemness in pancreatic cancer (Chen et al., 2020). Besides this, however, the expression pattern and functional role of ADGB in cancer entities remains elusive.

We therefore aimed at characterizing the expression intensity of ADGB in different human cancer entities and during cancer progression to analyze whether ADGB could influence tumorigenesis *in vivo*.

Material and Methods

RNA-sequencing analysis

RNA-seq libraries were prepared from 1000 ng of total RNA with Illumina's *TruSeq RNA Sample Prep Kit v2*, including poly-A selection, and sequenced as 70 nt single-end reads on an Illumina HiSeq2500. Library preparation and sequencing was performed by the NGS Core Facility of the Department of Biology, Johannes-Gutenberg-University (Mainz, Germany). Raw sequencing data are available at the European Nucleotide Archive under accession number XXXXX.

Processing of RNA-seq data

Raw transcriptome data were downloaded either from NCBI or ENA web servers (<https://www.ncbi.nlm.nih.gov/sra>; <https://www.ebi.ac.uk/ena>). Trimming parameters were assessed for each dataset via inspection with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter and quality trimming were performed with BBDuk (<https://sourceforge.net/projects/bbmap/>). Minimum quality cutoff of all datasets was phred 20, minimal read length was 20bp. After processing, the reads were mapped against the reference genome of *Homo sapiens* (Hsa 38) with HISAT2 under default parameters (Kim et al., 2019). Calculation of *transcripts-per-million* values (TPM) was performed with StringTie (Pertea et al., 2015) in guided mode.

Differential expression and GO term analysis

Differential gene expression analysis was conducted in R (<https://www.r-project.org/>) with Bioconductor package DESeq2 (Love et al., 2014). Genes were considered significant with p -adjusted ≤ 0.1 . No LFC cutoff was applied. GO term enrichment analysis on gene lists with subsequent term reduction and visualization was performed with Metascape (Zhou et al., 2019) and g:profiler (Raudvere et al., 2019).

Prediction of upstream regulators

Upstream regulators were predicted on lists of differentially expressed genes with the Core Ingenuity pathway analysis (IPA, Qiagen). Chemical regulators were excluded from the analysis. Upstream regulators with a Z-Score $> |2|$ and a BH-corrected p -value > 0.05 were considered significant.

Analysis of TCGA datasets

TCGA datasets were filtered, analysed and visualized with the GEPIA 2 tool (Tang et al., 2019). Parameters for differential testing of ADGb expression were $|LFC| \geq 0.5$ and p-value cutoff = 0.01.

Generation of a stable overexpression system

A549 cells were provided by Prof. Dr. Susanne Strand (University Medical Center, Mainz). Cells were cultured in DMEM supplemented with 10 % FBS and 1 % Pen/Strep (PAN Biotech) at room air with 5 % CO₂ and 37°C. pcDNA3.1 Vector with either full length ADGB or an empty backbone as control were transfected using Lipofectamine 3000 (Invitrogen). Cells were selected with Geneticin (Roth) over 2 weeks and clones were obtained through single cell dilution. For each condition, several clones of independent integration events were used for further analysis.

RNA extraction

RNA extraction of tumor tissues was performed from snap-frozen samples with the RNeasy Plus Universal Mini kit (Qiagen) according to the manufacturer's instruction. ~50 mg of tissues was grinded and homogenized with a MiniLys (Precellys) system using mixed ceramic beads (Precellys Lysing CKMix, Precellys). RNA was eluted in nuclease-free water. RNA extraction from cell culture was performed with RNeasy Mini kit (Qiagen) according to the instructions of the manufacturer.

RNA quality was assessed via a Bioanalyzer chip (Agilent), and only samples with RIN > 7 (for tissues) or > 9 (for cultured cells) were used for further analysis. RNA was quantified via Qubit measurement using the Broad Range RNA Assay Kit (Thermo Fisher) and was stored at -80°C until further use.

qRT-PCR

To validate the bioinformatical findings, we performed quantitative reverse-transcription PCR (qRT-PCR) on tissues as described previously (Koay et al. 2021). 1000 ng of total RNA per sample was used for reverse transcription with the SuperScript III enzyme (10000 units per assay; Invitrogen) using an Oligo-dT primer. In the absence of validated reference genes, the amount of mRNA expression was normalized on the adjusted total amount of RNA. To

additionally control for differences in cDNA synthesis, 100 ng of *Drosophila* total RNA was added to the reaction as a spike-in control. qRT-PCR was carried out using GoTaq qPCR Master Mix (Promega) on the ABI Prism 7500 Fast Detection System (SDS, Applied Biosystems, Carlsbad, USA) and interpreted using 7500 Software Version 2.3. Quantification of ADGB cDNA molecules (fwd primer: 5'-CGGAAGGAAAACATTCAAACAGG-3'; rev Primer: 5'-CGAAACTGATGAATTTCTTCCGC-3') was done in absolute numbers applying a calibration standard curve with known amounts of target PCR product, previously cloned into the pGEM T-easy vector system (Promega). Copies of the *Drosophila* Globin 1 (Glob1) cDNA (fwd primer: 5'-GGAGCTAAGTGGAATGCTCG-3'; rev primer: 5'-GATGATTCCGTAGACATGGTC-3') of the internal control were measured in parallel to identify samples with substandard reverse transcription.

Immunofluorescence

A549 cells were seeded in 24 well plates on uncoated glass cover slips. After 24-48h, cells were fixed with 4% PFA, washed, blocked in blocking buffer (1xPBS-TritonX with 5% Horse serum and 1% BSA) and incubated with corresponding primary antibody (α -acetylated tubulin; custom made α -ADGB (Charles River laboratories)) 1:500 in blocking buffer overnight. After several washing steps, cells were incubated with secondary antibody (CF 488, CF 555 or CF 633 conjugated anti-mouse/anti-rabbit F(ab)'-Fragment (Sigma Aldrich)) and counterstained with DAPI. Cover slips were mounted in RotiMount Fluorcare antifading solution (Roth) and sealed with Roti-Mount (Roth). Image acquisition was done with a Leica SP5. Cellular appendages were measured with ImageJ.

Migration assay

Cells were seeded in 4-well culture dishes (ibidi) with silicone insert, which allowed comparison of multiple cell population at once with a defined gap size. The silicon insert was removed after cells reached confluency (~24h). Every 4h, cells were observed under an inverse microscope and pictures were taken. Gap area was segmented, measured and plotted with the ImageJ MiToBo ScratchAssayAnalyzer Toolbox (<https://mitobo.informatik.uni-halle.de/index.php>), with $\sigma=2$ and an entropy filter size of 25.

Results

ADGB is downregulated in tumorous tissues and cell lines

Advances in NGS techniques have led to a plethora of publicly available data on transcriptomic expression changes associated with tumorigenesis. We therefore mined mRNA sequencing data from several big consortia that analysed large numbers of transcriptomic data from cancerous tissues as well as commonly used cell lines.

Since lung is an endogenous expression site of ADGB and also a very frequent cancer type, we firstly reanalyzed two RNA-Seq datasets on lung cancer progression in human. We found a significantly lower amount of ADGB expression in fully developed squamous cell carcinoma (Ooi et al., 2014) in comparison to normal basal cells (Figure 1A). There was also a tendency towards lower levels of ADGB mRNA already in premalignant cell lesions, however, this was not statistically significant. The second study (Morton et al., 2014) showed a similar result, with significantly lower ADGB levels in invasive lung carcinomas (Figure 1B). Stratified by patients, the study also showed that ADGB mRNA was downregulated in every patient analysed (n=6, Figure 1C).

Next, we investigated the extensive TCGA datasets (<https://portal.gdc.cancer.gov/>; Figure 2A) with a focus on tissues with endogenous ADGB expression (Koay et al., 2021). The downregulation of ADGB mRNA expression could again be verified in different tumor entities from lung and testis. We validated these findings via qPCR in paired normal and tumour tissue samples from human lung and testis of an independent cohort (Figure 1 D,E). Absolute expression of ADGB mRNA was lower in every tumour sample compared to its healthy counterpart. The TCGA data also showed a (non-significant) tendency towards higher levels of ADGB mRNA in two types of uterine cancer (Figure 2A). However, there was no difference of ADGB mRNA levels between glioblastoma multiforme, brain lower grade glioma, pancreatic adenocarcinomas and their respective healthy tissues (Figure 2B).

In summary, ADGB mRNA expression decreases during cancer progression in testis and lung tissue, but the starting amount of ADGB varies greatly in accordance with our earlier findings. This is probably due to ADGBs strict cell type specific expression pattern and different proportions of ciliated cells in the analysed samples (Koay et al., 2021).

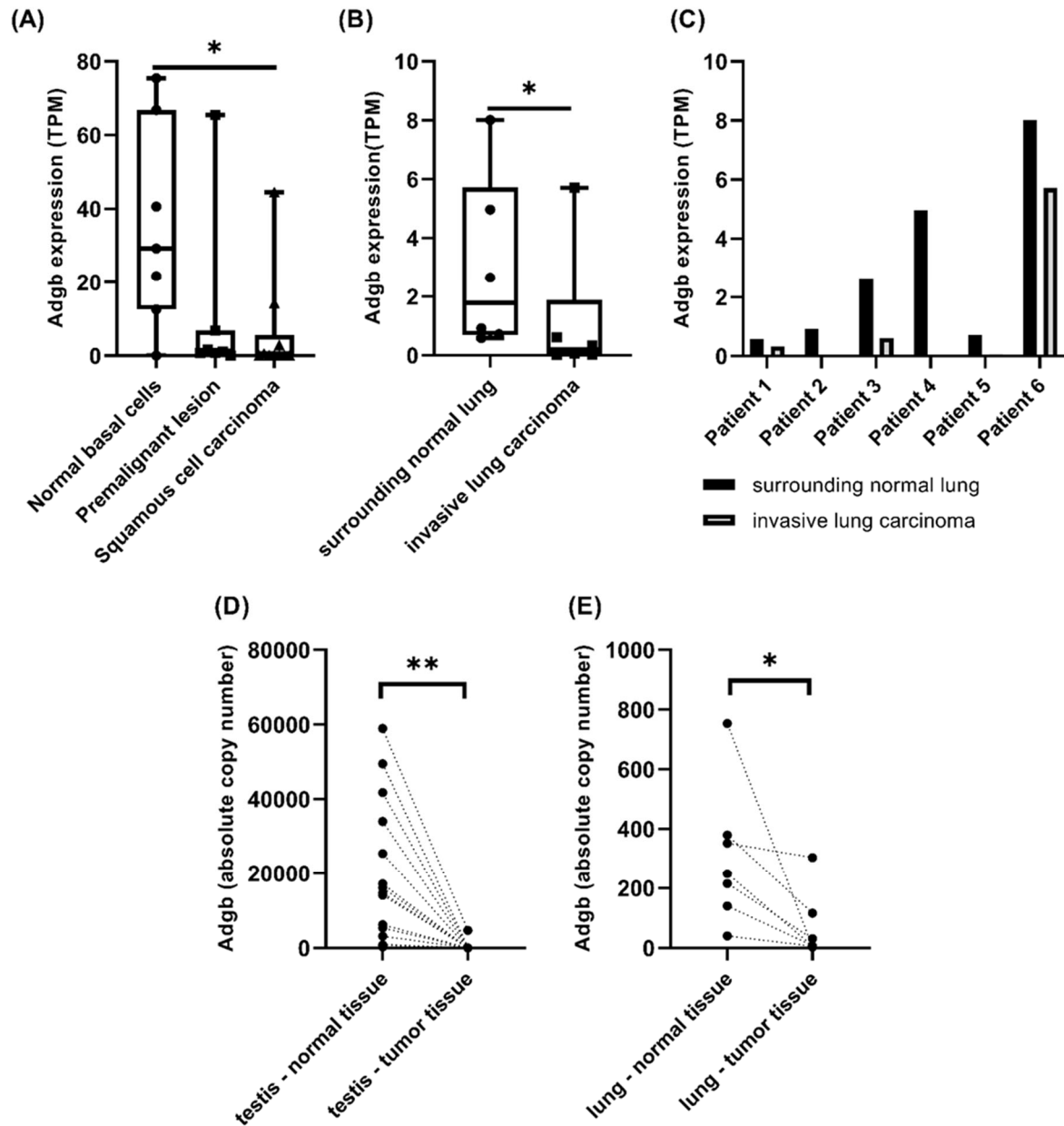


Figure 1: Expression of ADGB mRNA in human tumor tissue. (A) ADGB mRNA expression decreases significantly during progression from normal basal cells to squamous cell carcinoma (RNA-Seq data set from Ooi et al. 2014) (B) + (C) ADGB mRNA expression is also significantly lower in invasive lung carcinoma (data set from Morton et al. 2014). Although there is high variability in the amount of ADGB mRNA in normal tissue samples, it decreases in every patient analysed during tumour progression. (D) + (E) qPCR analysis of human ADGB mRNA levels in paired normal and tumour tissue samples of testis and lung cancer. ADGB mRNA expression is significantly lower in both types of tumour tissue biopsies. * = $p < 0.05$, ** = $p < 0.01$

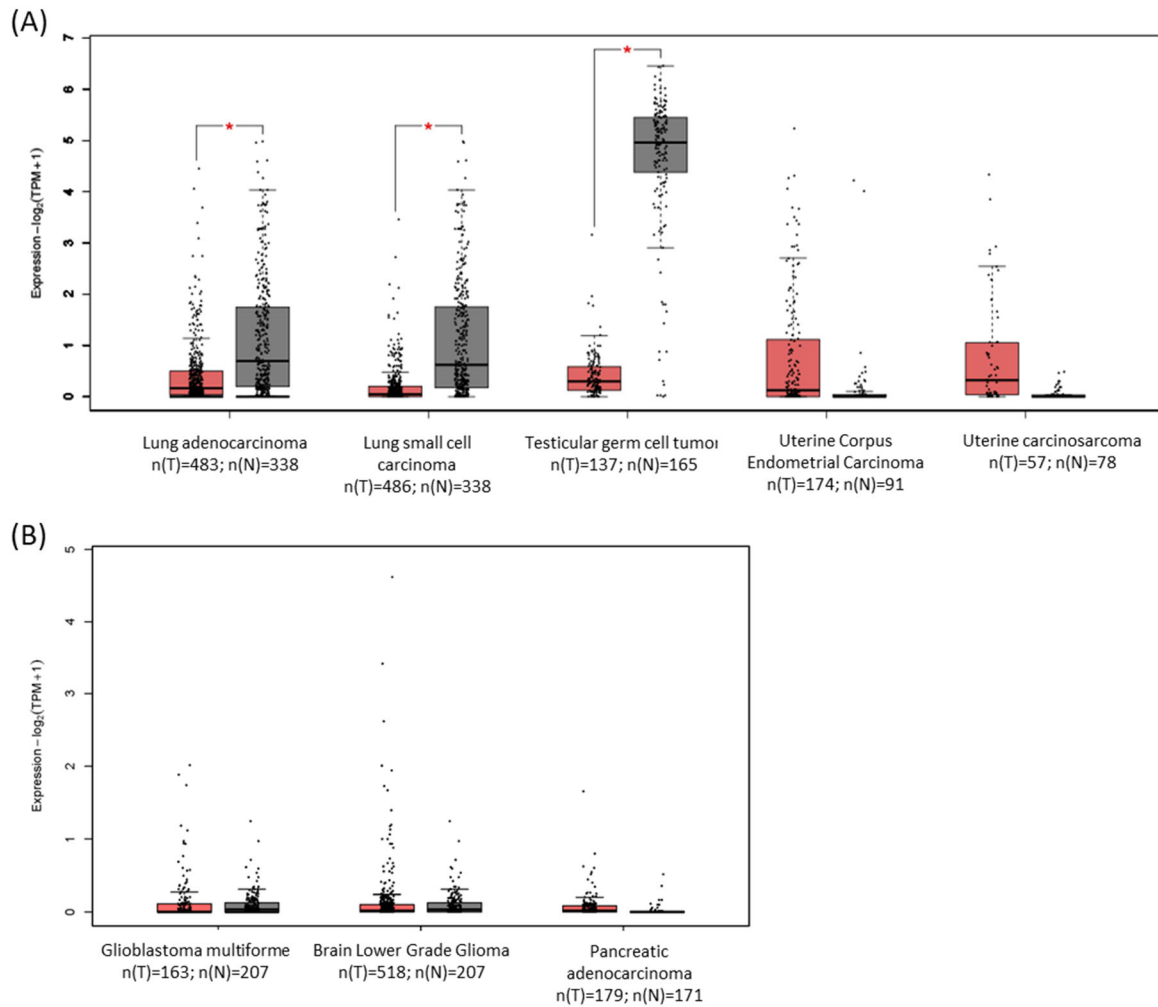


Figure 2: ADGB expression in a subset of tumours sequenced by the TCGA consortium. Red = tumorous tissue; grey = normal tissue. Number in brackets indicate the number of tumour or healthy tissue entities that were used for the analysis. (A) Expression of ADGB in tumours which resulted from tissues with endogenous expression. ADGB is significantly downregulated in lung adenocarcinoma, lung small cell carcinoma as well as testicular germ cell tumours. In contrast, ADGB mRNA expression is higher in tumorous uterine tissues, but compared to normal tissue this difference is not significant. (B) ADGB expression in glioblastoma multiforme, brain lower grade glioma, pancreatic adenocarcinoma compared to the corresponding healthy tissue. Only low levels of ADGB mRNA can be detected in all these groups.

In light of other studies which suggested an oncogenic function of ADGB in vitro (Chen et al., 2020; Huang et al., 2014), we also screened a variety of publicly available databases with transcriptomic information on cell lines. We found that of the 961 cell lines analyzed by the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012; Ghandi et al., 2019), only four showed a low amount of ADGB mRNA expression: lung cancer NCIH2106 (TPM 12.09); liposarcoma KMLS1 (TPM 11.93), Hodgkin lymphoma HDLM2 (TPM 4.33) and colon adenocarcinoma C75 (TPM 2.68) (Figure 3A). Intriguingly, genomic resequencing data and subsequent gene copy number analysis derived from the DepMap portal suggested that elevated ADGB expression in NCIH2106 as well as KMLS1 is not due to modulation of ADGB regulation, but rather a consequence of an increased number of gene copies in these cell lines (Supplementary Figure 1A). Of the 63 cell lines analyzed by the Human Protein Atlas consortium, again only HDLM-2 expressed low amounts of ADGB (Figure 3B). We reanalysed transcriptomic data of HDLM-2 from 2 independent studies. ADGB was indeed expressed in both (TPM 7.63, 6.48, 9.91 (PRJNA668062) and 1.85 (PRJEB30312)). Visual inspection of the alignments revealed that the cDNA reads covered only the last 4 exons from the ADGB gene locus (Supplementary Figure 1B). This indicates that the HDLM-2 cell line may express only a truncated version of the gene which neither encompasses the calpain, nor the globin and calmodulin-binding domains of ADGB. We found no bioinformatical evidence for endogenous ADGB mRNA expression in glioblastoma cell lines U87 and U251 in contrast to what was described previously (Huang et al., 2014) (Figure 3A, “Brain cancer”).

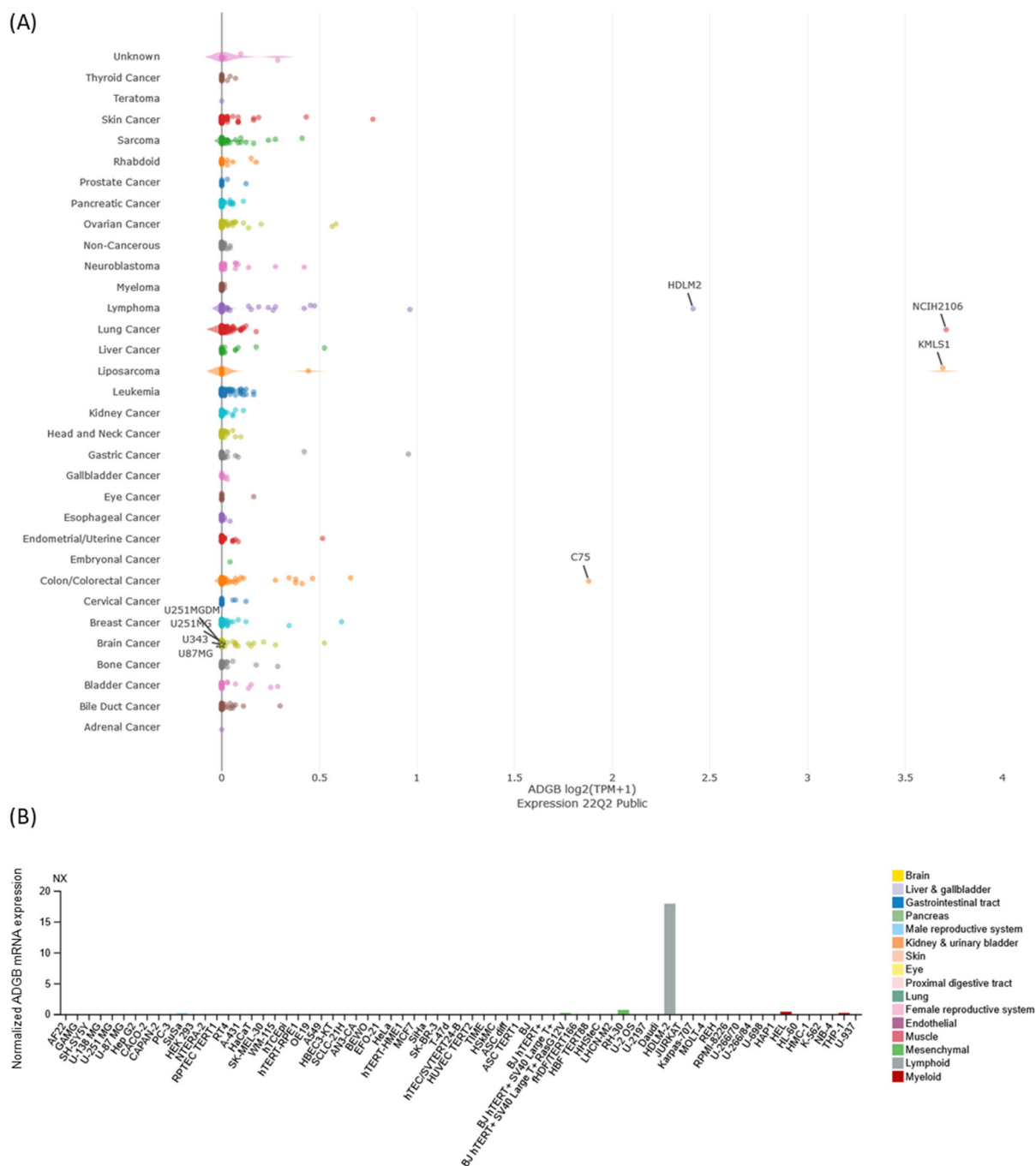


Figure 3: ADGB mRNA expression in cancer cell lines analysed by the CCLE (A) and Human Protein Atlas consortia (B). Only a small subset of cancer cell lines shows endogenous expression of ADGB mRNA.

Correlation analysis implies complex regulatory mechanisms of ADGB expression

Recently, a possible mechanism for the observed downregulation of ADGB mRNA was proposed. A study suggested that the expression of LncRNA STXBP5-AS1 and subsequent recruitment of transcription factor EZH2 lead to heavy methylation and thus silencing of the ADGB promoter (Chen et al., 2020). As a consequence, tissues with high levels of STXBP5-AS1 should show lower levels of ADGB mRNA. It seems, however, that the downregulation of ADGB in tumorous tissues might follow a different mechanism, since we did not find evidence for the expected inverse correlation of STXBP5-AS1 and ADGB in the TCGA data of tissues with endogenous ADGB expression (Supplementary figure 2A). On the contrary, both genes show a positive correlation in their expression pattern in this dataset. Since the STXBP5-AS1 gene is located directly downstream of the ADGB gene on chromosome 6, an open chromatin conformation could possibly affect both genes simultaneously. We also analysed the correlation coefficient of ADGB and STXBP5-AS1 in pancreatic tissue only, since this is the tissue for which the mechanism was initially proposed. Also in this sub-dataset, we found no evidence of a correlated expression pattern (Supplementary figure 2B).

Another study recently showed that overexpression of transcription factors FOXJ1 and RFX2 was able to induce the expression of ADGB in HEK293 cells (Koay et al., 2021). Indeed, we found a positive correlation between ADGB and expression of FOXJ1 as well as RFX2 in the TCGA datasets. Interestingly, ADGB mRNA expression can take place with virtually no expression of FOXJ1 at all (Supplementary figure 2C, red box). In contrast, there was no RFX2-negative sample that proved ADGB positive (Supplementary figure 2D). RFX2 should mechanistically stabilize FOXJ1 binding during motile ciliogenesis (Quigley & Kintner, 2017) and the cell culture experiments showed that RFX2 alone is not sufficient to induce ADGB expression (Koay et al., 2021). The results of the present study therefore suggest the existence of additional, FOXJ1-independent regulatory mechanisms for ADGB mRNA expression in tissue samples.

Overexpression of Androglobin induces morphological changes in A549 cells

As shown above, none of the analyzed cell lines seemed to be an adequate model of endogenous ADGB expression to study its function: expression levels were either extremely low, expression did not cover the complete ADGB gene locus, cell lines putatively carried duplications of the ADGB gene locus (complicating genetic KO experiments), or cell lines were derived from tissues which themselves are no native expression site of ADGB in healthy organisms.

To study which effect the expression of ADGB would have on a cancer cell line, we decided to stably overexpress ADGB in the widely used A549 lung adenocarcinoma cell line. Overall, 10 clonal cell lines were generated through G418 selection, 5 of which expressed human full length ADGB and another 5 contained an empty vector backbone as control. We validated the overexpression on mRNA level via qPCR as well as on protein level by immunofluorescent staining of fixed cell samples (Supplementary figure 3). ADGB antibody signal showed an even cytoplasmic distribution. Although ADGB carries a putative nuclear localisation signal, there was no evidence of ADGB transport to the nucleus (Figure 4, Supplementary figure 3). Intriguingly, ADGB positive A549 cells showed abnormally long cellular appendages (Figure 4), also including cytokinetic bridges, with a median length of 10.52 μm vs. 4.49 μm in empty control cells. Rare outliers were even as long as 50 μm (Figure 4D).

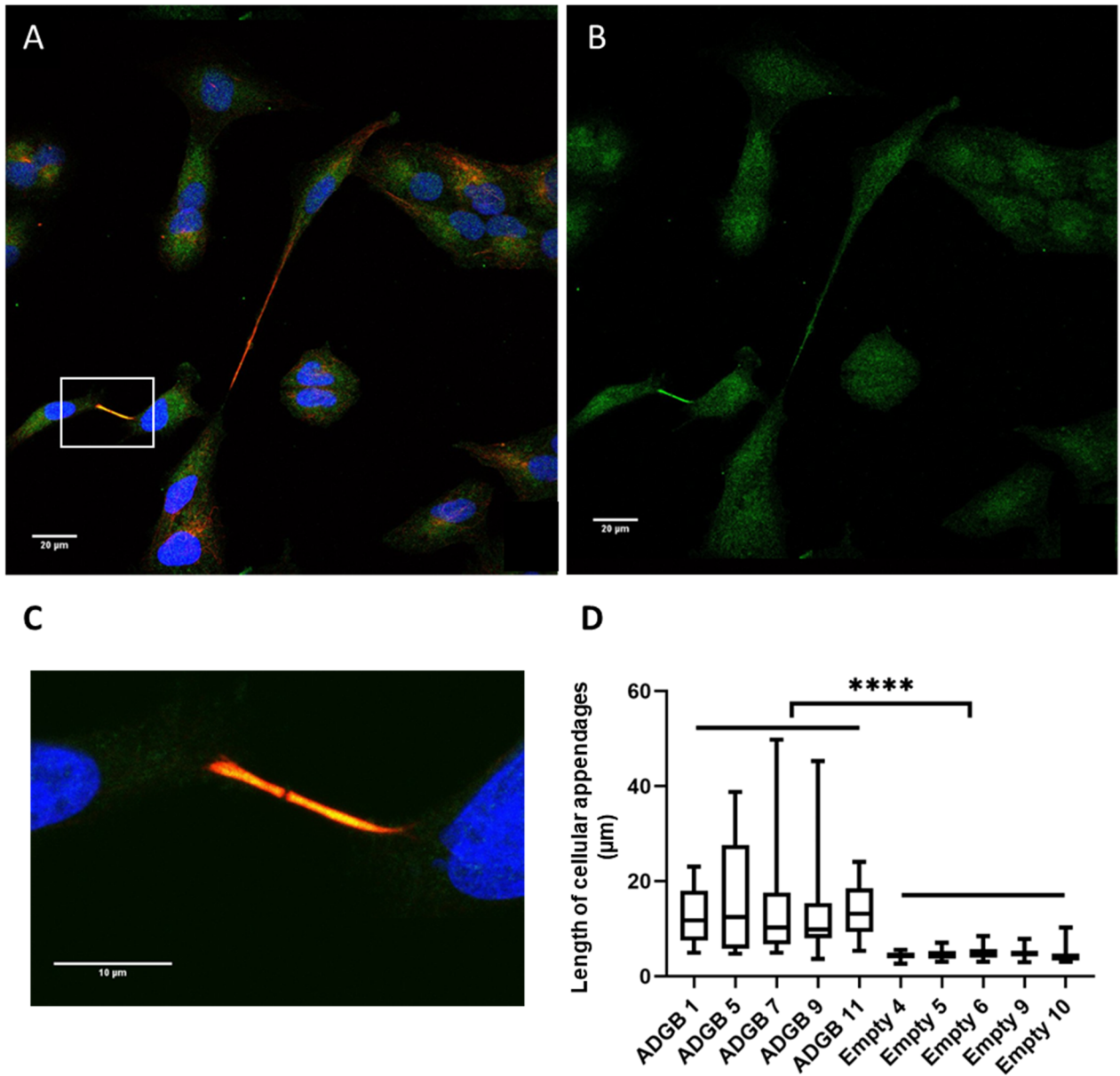


Figure 4: Immunofluorescent staining of ADGB-overexpressing A549 cells. ADGB = green, DAPI (DNA) = blue; acetylated α -Tubulin = red; ADGB (isolated channel in (B)) is localized in the cytoplasm and accumulates in cytokinetic bridges (C) which are also rich in Tubulin. (D) Ectopic overexpression of ADGB in A549 clones significantly increased the length of cellular appendages.

Transcriptomic changes in response to ADGB overexpression in A549 cells

To further characterize the molecular phenotype of ectopic ADGB mRNA overexpression, we performed RNA-seq analysis of our stably transfected A549 cell clones (n=3) in comparison to empty control cell lines (n=3). TPM values for ADGB mRNA were 4.01, 1.90 and 0.57 and thus roughly comparable to the amount of ADGB mRNA in healthy lung samples (see also qPCR results in Supplementary figure 3).

Differential gene expression analysis identified a set of 497 genes, 231 of which were upregulated in response to ADGB overexpression (Supplementary file 1). Among these, a distinct set of genes was either fully induced (counts in vector only cells: 0 in all three replicates) or downregulated (final counts: 0 in all three replicates) upon ADGB induction (Table 1). Noteworthy, many of the upregulated genes are expressed in tissues which are positive for endogenous ADGB mRNA expression. The expression of the lincRNA EWSAT1 for example is strongly induced upon overexpression of ADGB. It has been described as a potential oncogene in both colorectal cancer and osteosarcoma (Shen et al., 2021; R. Zhang et al., 2018), but in healthy tissue it is predominantly expressed in the testis. This gene and the others are potential candidates for further analyses of ADGB's gene regulatory network.

We then performed GO overrepresentation analysis on the list of differentially expressed genes (Figure 5; full list: Supplementary file 2). ADGB overexpression is strongly associated with transcriptional changes of genes associated with, among others, the extracellular matrix organisation (GO:0030198), cell adhesion (GO:0016477) and cell migration (GO:0016477). A similar picture was also seen when only upregulated genes were used for GO analysis (Supplementary file 3). Comparison of our gene list against databases dedicated to pathways resulted in hits for extracellular matrix organisation (REACTOME, Gillespie et al., 2022) and epithelial-to-mesenchymal transition in both thyroid cells and colorectal cancer (WikiPathways, Martens et al., 2021).

To identify potential regulators involved in the transcriptional changes induced by ADGB overexpression, we performed an upstream regulator analysis using the Ingenuity Pathway software suite (Qiagen). The resulting activated (blue) or deactivated (red) regulators are summarized in Supplementary table 1. We evaluated the literature on these regulators with regard to their influence on the motility of cells. Due to the proposed involvement of ADGB in cilia/flagella formation (Keppner et al., 2022; Koay et al., 2021) and its possible calcium, NO or

oxygen reactivity (Hoogewijs et al., 2012), we also investigated the upstream regulators with respect to these terms. Indeed, many of these candidate regulators turned out to be associated with ciliogenesis or were even known to influence the motility or length of cilia. Few, however, were connected with oxygen or NO metabolism, in literature. In line with the GO term analysis of the target genes, many of these regulators are involved in modulating the motility of cells or regulate the epithelial to mesenchymal transition.

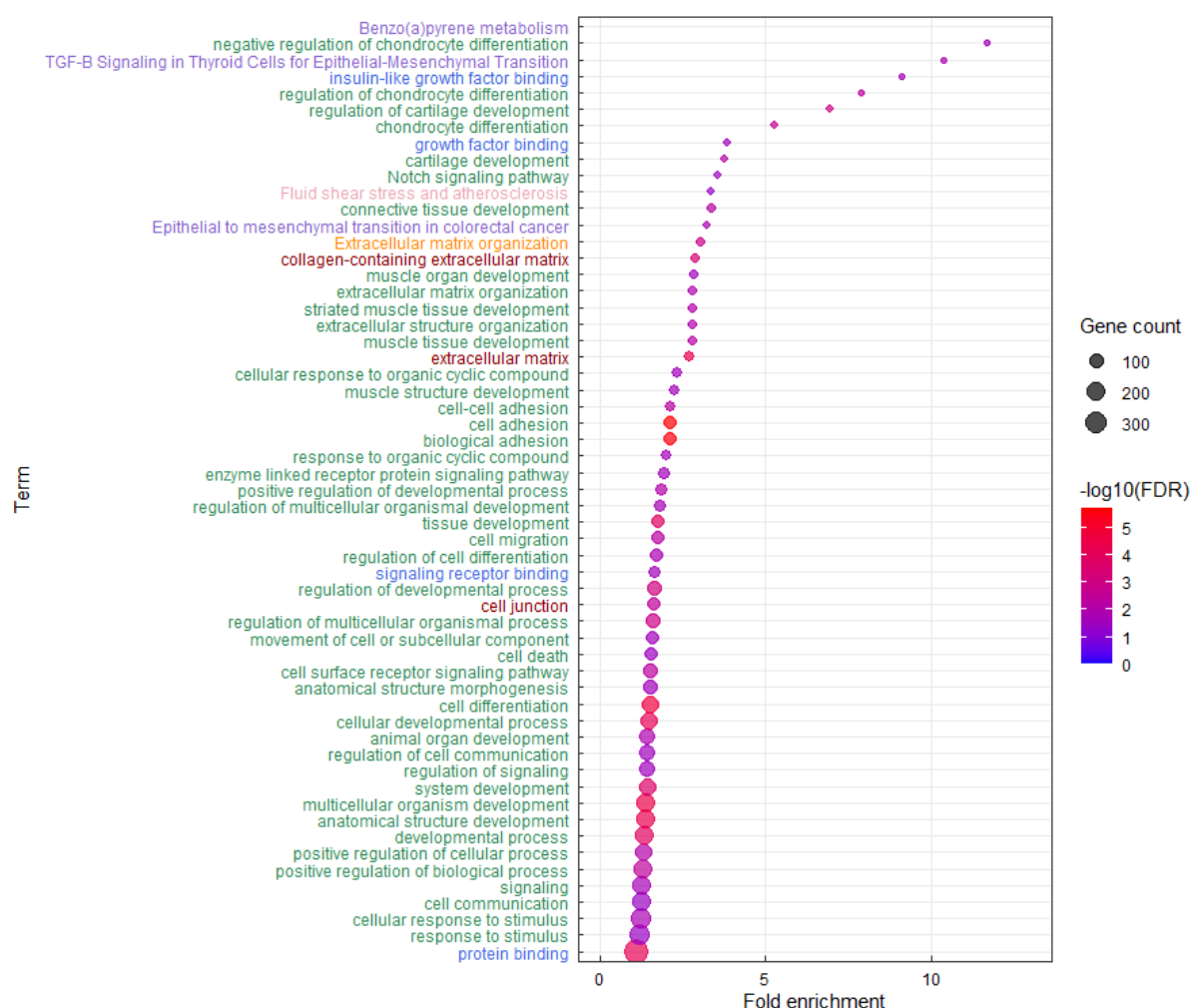


Figure 5: GO terms enriched in differentially expressed genes upon overexpression of ADGB in A549 lung cancer cells. Differentially expressed genes: LFC > 0.5; adjusted p-value < 0.05. Terms are strongly associated with changes in the extracellular matrix and motility as well as adhesion of cells. Terms are colour coded by their respective database: green = GO: Biological process; blue = GO: Molecular function; red = GO: Cellular component; rose = KEGG; Purple = WikiPathways; orange = REACTOME Pathways.

ADGB overexpressing cells do not exhibit higher motility

The bioinformatical findings predicted a molecular phenotype in motility and the composition of the extracellular matrix. Thus, we investigated the migratory capacity of our cell culture system. Although transcriptomic data suggested a more mesenchymal phenotype in ADGB+ vs. ADGB- cells, cell migration assays revealed no statistically significant difference in migratory capacity (Figure 6).

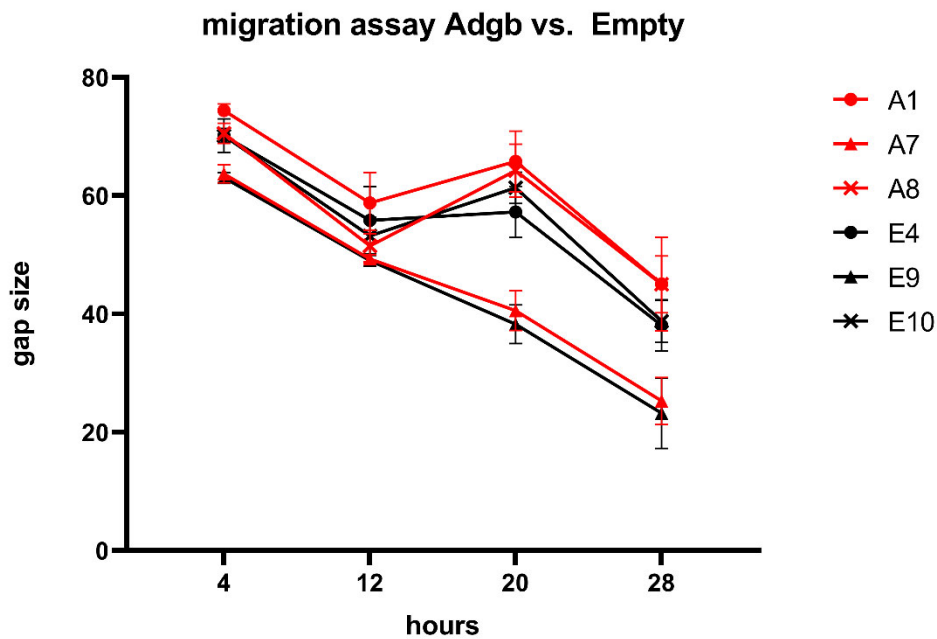


Figure 6: Migration assay of ADGB+ (red) and Empty control cells (black). Symbols indicate pairs of cell lines that were grown in the same culture vessel. In pair A7/E9, a constant decline in gap size could be measured, but no difference between the genotypes was observable. In the 2 other pairs, gap size increased unexpectedly after 12 hours, indicating technical difficulties with the experiment.

Discussion

Androglobin is a chimeric protein harbouring a globin domain (Hoogewijs et al., 2012). The other members of the globin protein family are known for differential regulation and both oncogenic or protective functions in tumour entities. We therefore studied the expression of ADGB in the context of cancer. Our analysis of a wide variety of transcriptome of cancer cell lines and different malignancies revealed that ADGB is not expressed in the vast majority of cell lines analysed. In addition, ADGB was found downregulated in tumours derived from tissues with endogenous ADGB expression. This could suggest that ADGB either has a tumour suppressor function or simply no influence at all upon tumorigenesis. Counterintuitively, our transcriptomic analysis of an ADGB overexpression system suggested a putative oncogenic role for ADGB, much in agreement with two previous studies, which showed that downregulation of ADGB resulted in slower proliferation and higher rate of apoptosis of glioblastoma cells (Huang et al., 2014) and compromised stemness and metastatic capacity in pancreatic cancer cell lines (Chen et al., 2020). Ectopic ADGB expression in A549 cells led to transcriptional changes which would promote cell motility and changes of the ECM, both processes relevant for the epithelial to mesenchymal transition and a hallmark of malignant cancer entities (e.g. Grelet et al., 2017; Thiery et al., 2009). However, this potential gain in motility was not pronounced enough to be observable in a classic migration assay. Nevertheless, ADGB overexpressing cells showed longer cellular appendages including cytokinetic bridges. These bridges are the final connection of two daughter cells which have completed mitosis. Since the process of final abscission is carefully timed (Nähse et al., 2017), longer bridges suggest a higher motility of the connected cells or a change in adhesive properties, which would be in line with the transcriptomic and regulator analysis of our cell model. ADGB+ tumour cells may thus get a selective advantage *in vitro* – but ADGB positivity simply does not happen in cancer *in vivo*.

In healthy tissues, ADGB is exclusively expressed in ciliated cells of lung, brain and the fallopian tubes as well as during late stages of spermatogenesis (Koay et al., 2021). The overexpression of FOXJ1 and RFX2, which are important ciliogenesis transcription factors, was capable of ADGB induction *in vitro* (Koay et al., 2021). We have found that tissues can be positive for ADGB mRNA without the expression of FOXJ1, possibly hinting at additional regulatory mechanisms for ADGB *in vivo*.

Ciliogenesis is a process which is tightly connected to the cell cycle: the centriole acts both as a initiation point for the spindle apparatus and as the basal body, which migrates to the cell borders and serves as an axoneme nucleation point to initiate cilium protrusion (Avasthi & Marshall, 2012). The cilium can then be absorbed upon re-entry into mitosis. In contrast to this, differentiation to a cell carrying motile cilia is not reversible any more. Overproliferative cells such as some cancer cell lines and also most malignancies in vivo notably do not carry cilia. It seems that loss of the differentiation status of cells and tissues that comes along with cancer progression concomitantly leads to the loss of ADGB expression, indicating that ADGB is specifically needed in fully differentiated ciliated cells and also tightly regulated with this process.

In conclusion, ADGB may have the capability to enhance malignancy in tumour entities, and possibly does so in artificial overexpressing cell systems, but since its expression is simply not induced in vivo in proliferative, cancerous tissues, this has no prognostic value for cancer patients.

Acknowledgements

The results reported here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. The authors thank [REDACTED] for help in analysing qRT-PCR data and [REDACTED] for critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) to T.H. (HA 2103/9-1).

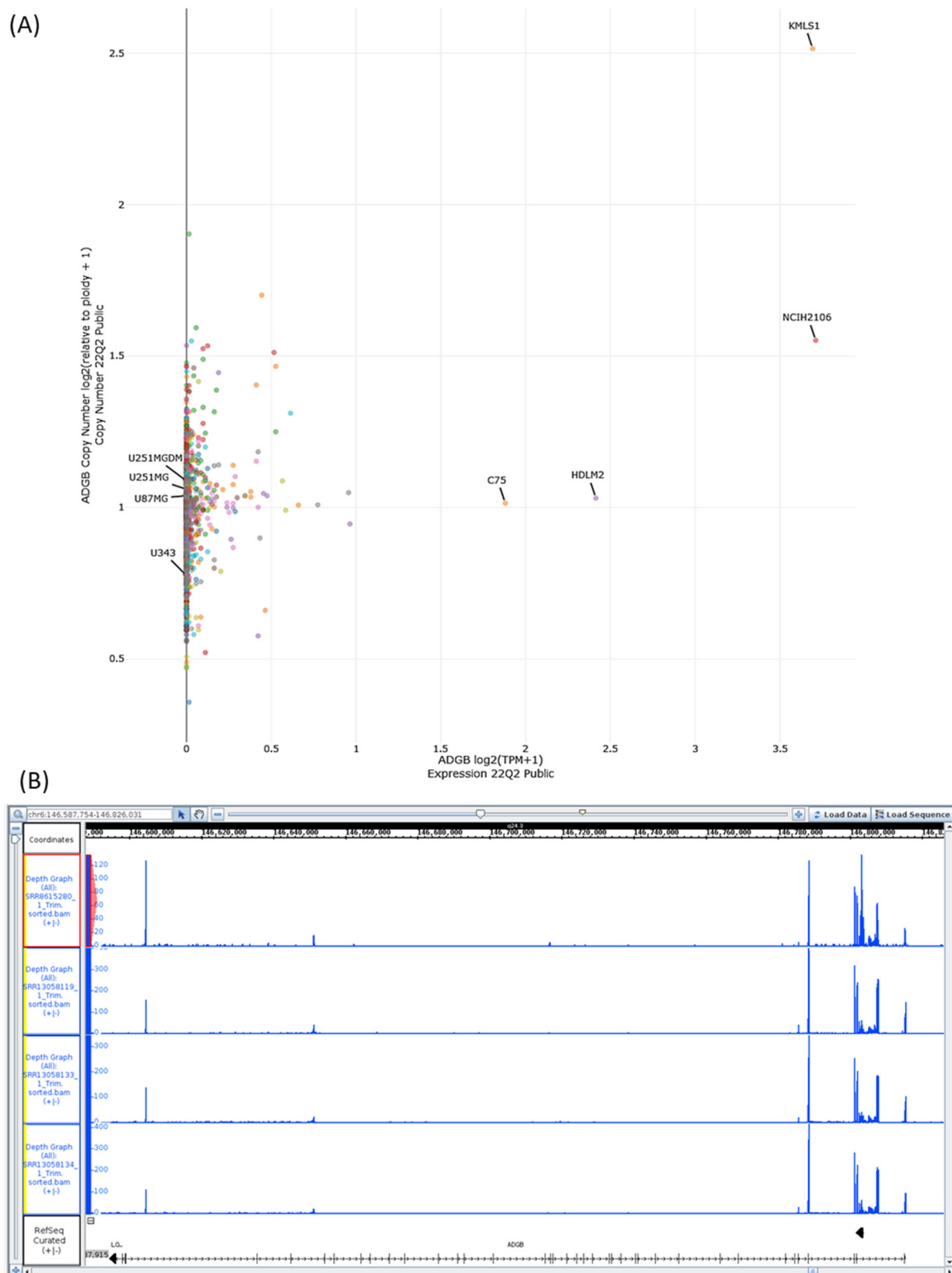
Table 1: Differentially expressed genes upon overexpression of ADGB in A%\$) lung cancer cells. All genes were either fully induced or downregulated in all replicates. LFC = LogFoldChange; p-adjusted = BH corrected p-value.

Gene Name	LFC	p-adjusted	RNA-type	main mRNA expression site (HPA + GTEx)
AC027088.2	10.99	8.71E-16	linc-RNA	testis, brain
ADGB	8.33	5.47E-07	protein-coding	fallopian tube, testis, lung, brain; ciliated cells
NMU	8.17	1.81E-07	protein-coding	esophagus, skin; squamous epithelial cells
LINC00648	7.69	4.39E-06	linc-RNA	broadly expressed
NELL2	7.69	8.23E-05	protein-coding	brains; neurons, T-cells
LINC02081	7.69	4.59E-06	linc-RNA	testis, lung
CTNND2	7.58	2.79E-05	protein-coding	brain
MAGEA1	7.38	1.82E-05	protein-coding	testis; spermatogonia, spermatocytes
JPH1	6.90	2.23E-03	protein-coding	broadly expressed; highest in Cardiomyocytes
GIMAP2	6.74	5.51E-03	protein-coding	lymphoid tissue; B-cells
GLI3	6.25	6.37E-02	protein-coding	broadly expressed; fibroblasts
SLC27A6	6.24	1.19E-02	protein-coding	broadly expressed, adrenal gland, brain, fallopian tube , heart muscle
RARRES3	4.48	5.32E-02	protein-coding	broadly expressed
WNT6	-4.00	5.63E-04	protein-coding	broadly expressed; Sertoli cells
OPRL1	-4.62	5.61E-04	protein-coding	testis; early & late spermatids
MUC5B	-4.98	6.64E-07	protein-coding	salivary gland; intestinal epithelial cells
ADGRL3	-5.62	1.91E-04	protein-coding	brain
AL645608.7	-5.99	1.45E-04	linc-RNA	broadly expressed, artery
S100P	-6.00	3.13E-04	protein-coding	bone marrow, stomach, urinary bladder
SELENBP1	-6.51	2.69E-05	protein-coding	intestine, liver
SLITRK4	-6.54	3.15E-15	protein-coding	adrenal gland, brain , skeletal muscle
LUM	-6.85	1.36E-06	protein-coding	gallbladder, placenta; fibroblasts
TNNC1	-7.20	1.47E-07	protein-coding	muscle; striated muscle cell
CCNYL2	-7.26	4.24E-05	pseudogene	testis
MYOCD	-7.72	1.15E-06	protein-coding	broadly expressed; fibroblasts
HOXA9	-8.06	6.29E-10	protein-coding	broadly expressed, highest in kidney; proximal tubular cells
ASAH2	-9.65	3.51E-15	protein-coding	intestine; enterocytes
LINC00839	-9.85	7.83E-15	linc-RNA	broadly expressed, kidney, spleen

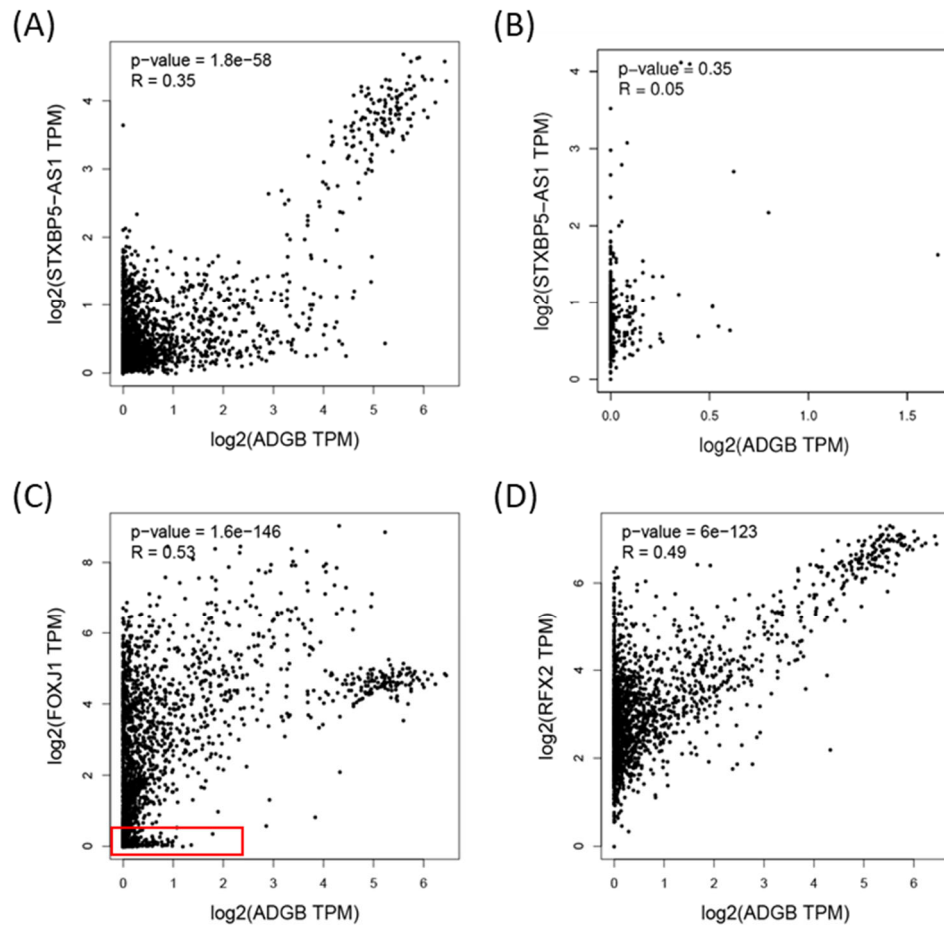
Supplementary table 1: IPA upstream regulator prediction in ADGB+ cells: activated (blue) and deactivated (red) transcription factors

Regulator	z-score	p-value	motility?	cilia?	calcium?	hypoxia/NO?
<i>TGFB3</i>	3,385	4,34E-06	↓ motility of cells (Han et al., 2012)	defect ↓ sperm motility (Drozdzik et al., 2015)		
<i>FGF2</i>	3,171	1,12E-04	↑ sperm motility (Saucedo et al., 2015)	controls ciliary length (Bosakova et al., 2019)	↑ intracellular Ca ²⁺ levels (Y. Liu & Schneider, 2014)	
<i>TNF</i>	2,978	2,04E-08	↑ ciliary beat frequency (dosage dependent) (Rhee et al., 1999; Weiterer et al., 2014)		↑ Ca ²⁺ levels (González et al., 2016)	
<i>EGF</i>	2,840	4,37E-07	↑ motility (Xie et al., 1998)	↓ ciliogenesis (Kasahara et al., 2018)		
<i>Tgf beta</i>	2,674	3,09E-05	regulates differentiation of motile cilia (Tözser et al., 2015)	controls ciliary length (Tözser et al., 2015)	controls ECM via Ca ²⁺ (Mukherjee et al., 2012)	
<i>STAT1</i>	2,599	1,29E-04	ROS mediated ↑ of motility (J. Liu et al., 2019)	destabilizes primary cilia (S. Wang et al., 2018)	regulated by calmodulin dependent kinase (L. Wang et al., 2008)	hypoxia ↓ STAT1 (Ivanov et al., 2007)
<i>MSTN</i>	2,412	2,50E-03	-			
<i>SNAI2</i>	2,395	1,71E-03				
<i>IFNB1</i>	2,385	7,78E-04				
<i>HGF</i>	2,244	2,48E-06	↑ epithelial cell motility (Spix et al., 2007)			
<i>Interferon alpha</i>	2,219	4,64E-07	cf. STAT1			
<i>TWIST2</i>	2,205	7,93E-06	↑ EMT (T. Wang et al., 2014)			
<i>PRMT1</i>	2,186	2,78E-04	↑ EMT (Gao et al., 2016; Jackson-Weaver et al., 2020)	localizes to Flagellum (Mizuno & Sloboda, 2017)	upregulated by Ca ²⁺ (M. Y. Liu et al., 2020)	hypoxia inducible in A549 (Lim et al., 2013)
<i>OSM</i>	2,180	9,39E-08				

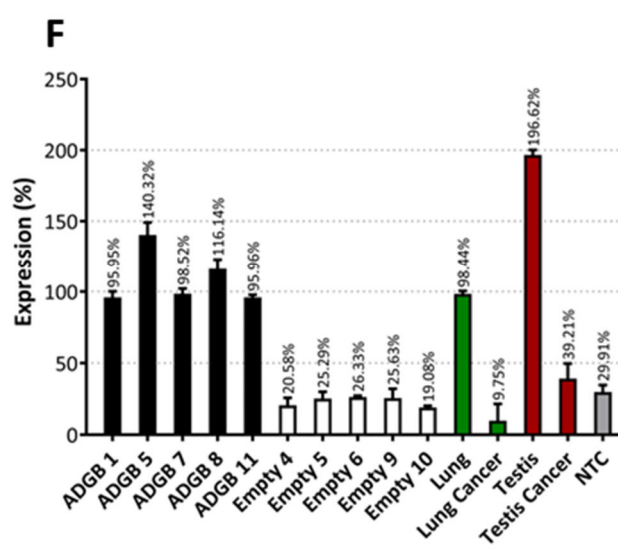
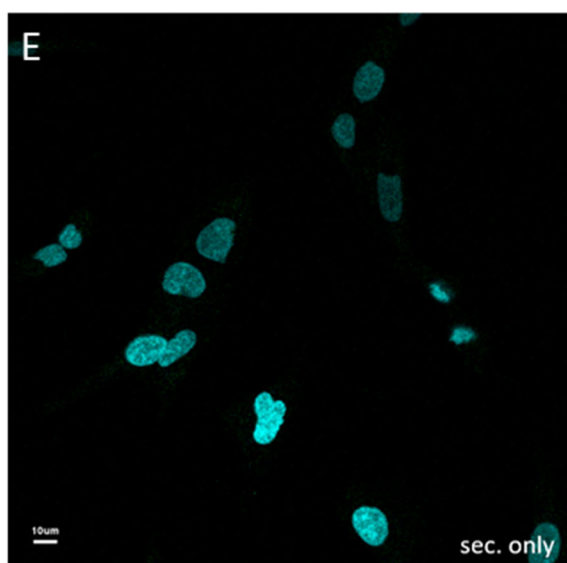
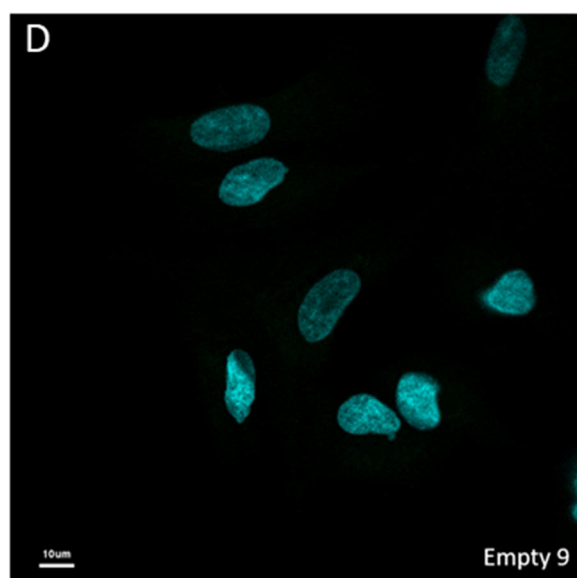
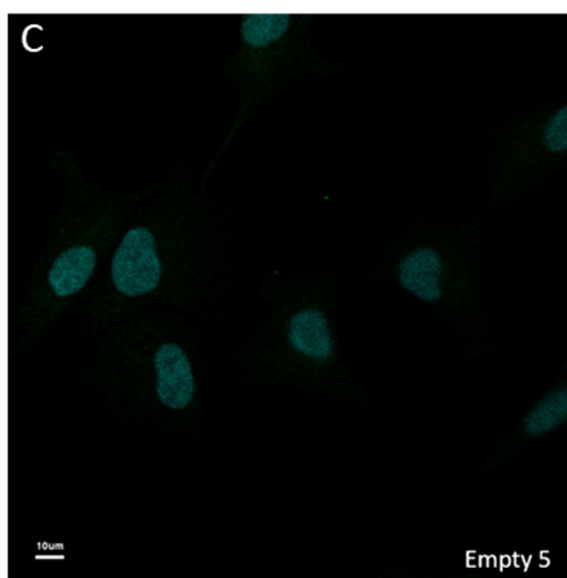
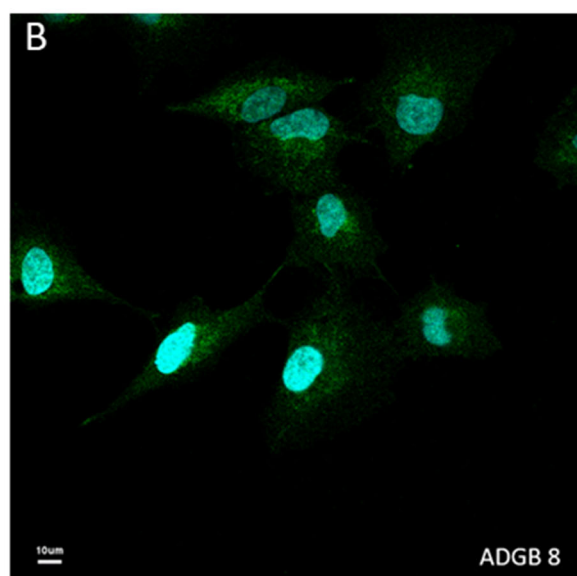
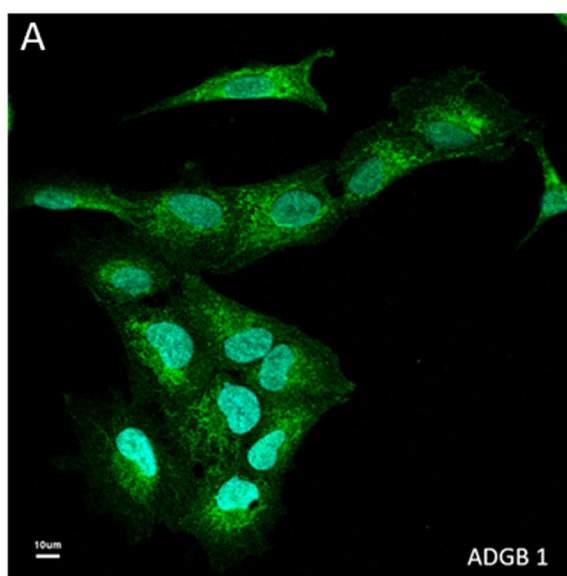
<i>TFAP4</i>	2,164	9,42E-04				
<i>SNCA</i>	2,138	2,66E-04				
<i>RASSF1</i>	-2,714	1,68E-06	isoform a ↓ motility isoform c ↑ motility (Vlahov et al., 2015)			
<i>AHR</i>	-2,599	2,23E-04	-	essential for ciliogenesis (embryo) (Villa et al., 2016)	regulates Ca-signalling (Brinchmann et al., 2018)	= oxygen sensor (Rothhammer & Quintana, 2019)
<i>BCL3</i>	-2,415	1,45E-03	↑(Wakefield et al., 2013)/↓(Wakefield et al., 2008) motility	-	-	
<i>TRIM24</i>	-2,170	1,40E-05	↓ motility (↑ adhesion) (Appikonda et al., 2018)	-	-	
<i>NEUROG1</i>	-2,121	3,89E-05	-	-	-	hypoxia ↑ NSC proliferation (Qi et al., 2017); NEUROG1 ↑ differentiation (Luo et al., 2010)
<i>MYOC</i>	-2,000	3,35E-04	↑ migration (Kwon & Tomarev, 2011)	localizes to cilia of photoreceptor cells (Kubota et al., 1997)	binds calcium (Donegan et al., 2012)	



Supplementary figure 1: (A) ADGB expression in cell lines from the CCLE consortium : TPM values and ADGB gene copy number appear to correlate. Cell lines NCIH2106 and KMLS1, which show the highest endogenous ADGB expression, also harbour multiple copies of the ADGB gene. (B) Graphical representation of ADGB read coverage in HDLM-2 RNA-Seq data. The 5'-end of the gene is free from reads (left) whereas the last 4 exons are expressed.



Supplementary figure 2: Correlation analysis of ADGB and potential transcription factors on mRNA level.



Supplementary figure 3: Validation of ectopic ADGB mRNA expression via qPCR and immunofluorescent staining. Green = α -ADGB antibody, blue = DAPI. (A) and (B) show staining of ADGB overexpressing lines 1 and 8, whereas (C) and (D) where stably transfected with an empty control (lines 5 and 9). (E) ADGB overexpressing cells stained only with a secondary antibody as control did not show any fluorescence in the green channel. (F) Validation of ectopic ADGB mRNA expression in transfected A549 cell clones (ADGB 1,5,7,8,11) via qPCR analysis. cDNA from lung, lung cancer, testis and testis cancer were used as references to assess intensity of overexpression. Expression intensity is shown in relation to the lung (Replicate 1, 100%). Weak PCR signals were obtained in vector transfected “Empty” A549 cells and the no-template control (NTC). Melting curve analysis showed that these signals were due to short amplicons originating from primer dimerization. No melting point corresponding to those artifactual amplicons could be detected in the other samples.

References

- Appikonda, S., Thakkar, K. N., Shah, P. K., Dent, S. Y. R., Andersen, J. N., & Barton, M. C. (2018). Cross-talk between chromatin acetylation and SUMOylation of tripartite motif-containing protein 24 (TRIM24) impacts cell adhesion. *Journal of Biological Chemistry*, 293(19), 7476–7485.
- Avasthi, P., & Marshall, W. F. (2012). Stages of Ciliogenesis and Regulation of Ciliary Length. *Differentiation; Research in Biological Diversity*, 83(2), S30.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M. F., Monahan, J. E., Morais, P., Meltzer, J., Korejwa, A., Jané-Valbuena, J., ... Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391), 603–607.
- Bosakova, M. K., Nita, A., Gregor, T., Varecha, M., Gudernova, I., Fafilek, B., Barta, T., Basheer, N., Abraham, S. P., Balek, L., Tomanova, M., Kuceroval, J. F., Bosak, J., Potesil, D., Zieba, J., Song, J., Konik, P., Park, S., Duran, I., ... Krejci, P. (2019). Fibroblast growth factor receptor influences primary cilium length through an interaction with intestinal cell kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 116(10), 4316–4325.
- Brinchmann, B. C., Le Ferrec, E., Bisson, W. H., Podechard, N., Huitfeldt, H. S., Gallais, I., Sergent, O., Holme, J. A., Lagadic-Gossmann, D., & Øvrevik, J. (2018). Evidence of selective activation of aryl hydrocarbon receptor nongenomic calcium signaling by pyrene. *Biochemical Pharmacology*, 158, 1–12.
- Burmester, T., Welch, B., Reinhardt, S., & Hankeln, T. (2000). A vertebrate globin expressed in the brain. *Nature*, 407(6803), 520–523.
- Chen, S., Huang, L., Li, G., Qiu, F., Wang, Y., Yang, C., Pan, J., Wu, Z., Chen, J., & Tian, Y. (2020). LncRNA STXBP5-AS1 suppresses stem cell-like properties of pancreatic cancer by epigenetically inhibiting neighboring androglobin gene expression. *Clinical Epigenetics*, 12(1), 1–12.
- Donegan, R. K., Hill, S. E., Turnage, K. C., Orwig, S. D., & Lieberman, R. L. (2012). The glaucoma-

- associated olfactomedin domain of myocilin is a novel calcium binding protein. *Journal of Biological Chemistry*, 287(52), 43370–43377.
- Drozdik, M., Kaczmarek, M., Malinowski, D., Bros, U., Kazienko, A., Kurzawa, R., & Kurzawski, M. (2015). TGFβ3 (TGFB3) polymorphism is associated with male infertility. *Scientific Reports*, 5, 71–74.
- Elsherbiny, M. E., Shaaban, M., El-Tohamy, R., Elkholi, I. E., Hammam, O. A., Magdy, M., Allalunis-Turner, J., & Emara, M. (2021). Expression of Myoglobin in Normal and Cancer Brain Tissues: Correlation With Hypoxia Markers. *Frontiers in Oncology*, 11, 1561.
- Emara, M., Turner, A. R., & Allalunis-Turner, J. (2010). Hypoxic regulation of cytoglobin and neuroglobin expression in human normal and tumor tissues. *Cancer Cell International*, 10(1), 1–16.
- Fiocchetti, M., Cipolletti, M., Ascenzi, P., & Marino, M. (2018). Dissecting the 17β-estradiol pathways necessary for neuroglobin anti-apoptotic activity in breast cancer. *Journal of Cellular Physiology*, 233(7), 5087–5103.
- Flögel, U., Merx, M. W., Gödecke, A., Decking, U. K. M., & Schrader, J. (2001). Myoglobin: A scavenger of bioactive NO. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 735–740.
- Flonta, S. E., Arena, S., Pisacane, A., Michieli, P., & Bardelli, A. (2009). Expression and functional regulation of myoglobin in epithelial cancers. *The American Journal of Pathology*, 175(1), 201–206.
- Gao, Y., Zhao, Y., Zhang, J., Lu, Y., Liu, X., Geng, P., Huang, B., Zhang, Y., & Lu, J. (2016). The dual function of PRMT1 in modulating epithelial-mesenchymal transition and cellular senescence in breast cancer cells through regulation of ZEB1. *Scientific Reports*, 6(December 2015), 1–13.
- Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., Barretina, J., Gelfand, E. T., Bielski, C. M., Li, H., Hu, K., Andreev-Drakhlin, A. Y., Kim, J., Hess, J. M., Haas, B. J., Aguet, F., Weir, B. A., Rothberg, M. V., Paoletta, B. R., ... Sellers, W. R. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, 569(7757), 503–508.

- Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., Deng, C., Varusai, T., Ragueneau, E., Haider, Y., May, B., Shamovsky, V., Weiser, J., Brunson, T., Sanati, N., ... D'Eustachio, P. (2022). The reactome pathway knowledgebase 2022. *Nucleic Acids Research*, 50(D1), D687–D692.
- González, C., Droguett, K., Rios, M., Cohen, N. A., & Villalón, M. (2016). TNF α Affects Ciliary Beat Response to Increased Viscosity in Human Pediatric Airway Epithelium. *BioMed Research International*, 2016, 1–9.
- Gorr, T. A., Wichmann, D., Pilarsky, C., Theurillat, J. P., Fabrizius, A., Laufs, T., Bauer, T., Koslowski, M., Horn, S., Burmester, T., Hankeln, T., & Kristiansen, G. (2011). Old proteins – new locations: myoglobin, haemoglobin, neuroglobin and cytoglobin in solid tumours and cancer cells. *Acta Physiologica*, 202(3), 563–581.
- Grelet, S., McShane, A., Geslain, R., & Howe, P. H. (2017). Pleiotropic Roles of Non-Coding RNAs in TGF- β -Mediated Epithelial-Mesenchymal Transition and Their Functions in Tumor Progression. *Cancers*, 9(7).
- Han, A., Bandyopadhyay, B., Jayaprakash, P., Lua, I., Sahu, D., Chen, M., Woodley, D. T., & Li, W. (2012). The anti-motility signaling mechanism of TGF β 3 that controls cell traffic during skin wound healing. *Biology Open*, 1(12), 1169–1177.
- Hendgen-Cotta, U. B., Merx, M. W., Shiva, S., Schmitz, J., Becher, S., Klare, J. P., Steinhoff, H. J., Goedecke, A., Schrader, J., Gladwin, M. T., Kelm, M., & Rassaf, T. (2008). Nitrite reductase activity of myoglobin regulates respiration and cellular viability in myocardial ischemia-reperfusion injury. *Proceedings of the National Academy of Sciences of the United States of America*, 105(29), 10256–10261.
- Hoogewijs, D., Ebner, B., Germani, F., Hoffmann, F. G., Fabrizius, A., Moens, L., Burmester, T., Dewilde, S., Storz, J. F., Vinogradov, S. N., & Hankeln, T. (2012). Androglobin: A Chimeric Globin in Metazoans That Is Preferentially Expressed in Mammalian Testes. *Molecular Biology and Evolution*, 29(4), 1105–1114.
- Huang, B., Lu, Y., Li, X., Zhu, Z., Li, K., Liu, J., Zheng, J., & Hu, Z. (2014). Androglobin knockdown inhibits growth of glioma cell lines. *Int J Clin Exp Pathol*, 7(5), 2179–2184.
- Ivanov, S. V., Salnikow, K., Ivanova, A. V., Bai, L., & Lerman, M. I. (2007). Hypoxic repression of

- STAT1 and its downstream genes by a pVHL/HIF-1 target DEC1/STRA13. *Oncogene*, 26(6), 802–812.
- Jackson-Weaver, O., Ungvijanpunya, N., Yuan, Y., Qian, J., Gou, Y., Wu, J., Shen, H., Chen, Y., Li, M., Richard, S., Chai, Y., Sucov, H. M., & Xu, J. (2020). PRMT1-p53 Pathway Controls Epicardial EMT and Invasion. *Cell Reports*, 31(10).
- Kasahara, K., Aoki, H., Kiyono, T., Wang, S., Kagiwada, H., Yuge, M., Tanaka, T., Nishimura, Y., Mizoguchi, A., Goshima, N., & Inagaki, M. (2018). EGF receptor kinase suppresses ciliogenesis through activation of USP8 deubiquitinase. *Nature Communications*, 9(1).
- Keppner, A., Correia, M., Santambrogio, S., Koay, T. W., Maric, D., Osterhof, C., Winter, D. V., Clerc, A., Stumpe, M., Chalmel, F., Dewilde, S., Odermatt, A., Kressler, D., Hankeln, T., Wenger, R. H., & Hoogewijs, D. (2022). Androglobin, a chimeric mammalian globin, is required for male fertility. *ELife*, 11.
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915.
- Koay, T. W., Osterhof, C., Orlando, I. M. C., Keppner, A., Andre, D., Yousefian, S., Alonso, M. S., Correia, M., Markworth, R., Schödel, J., Hankeln, T., & Hoogewijs, D. (2021). Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis. *Journal of Biological Chemistry*, 296.
- Kono, H., Hosomura, N., Amemiya, H., Kawaida, H., Furuya, S., Akaike, H., Kawaguchi, Y., Sudo, M., & Ichikawa, D. (2021). Cytoglobin as a Prognostic Factor for Pancreatic Ductal Adenocarcinoma. *Pancreas*, 50(7), 994–999.
- Kristiansen, G., Rose, M., Geisler, C., Fritzsche, F. R., Gerhardt, J., Lüke, C., Ladhoff, A. M., Knüchel, R., Dietel, M., Moch, H., Varga, Z., Theurillat, J. P., Gorr, T. A., & Dahl, E. (2010). Endogenous myoglobin in human breast cancer is a hallmark of luminal cancer phenotype. *British Journal of Cancer*, 102(12), 1736–1745.
- Kubota, R., Noda, S., Wang, Y., Minoshima, S., Asakawa, S., Kudoh, J., Mashima, Y., Oguchi, Y., & Shimizu, N. (1997). A novel myosin-like protein (myocilin) expressed in the connecting cilium of the photoreceptor: Molecular cloning, tissue expression, and chromosomal

- mapping. *Genomics*, 41(3), 360–369.
- Kwon, H. S., & Tomarev, S. I. (2011). Myocilin, a glaucoma-associated protein, promotes cell migration through activation of integrin-focal adhesion kinase-serine/threonine kinase signaling pathway. *Journal of Cellular Physiology*, 226(12), 3392–3402.
- Lim, S. K., Jeong, Y. W., Kim, D. Il, Park, M. J., Choi, J. H., Kim, S. U., Kang, S. S., Han, H. J., & Park, S. H. (2013). Activation of PRMT1 and PRMT5 mediates hypoxia- and ischemia-induced apoptosis in human lung epithelial cells and the lung of miniature pigs: The role of p38 and JNK mitogen-activated protein kinases. *Biochemical and Biophysical Research Communications*, 440(4), 707–713.
- Liu, J., Qu, L., Meng, L., & Shou, C. (2019). Topoisomerase inhibitors promote cancer cell motility via ROS-mediated activation of JAK2-STAT1-CXCL1 pathway. *Journal of Experimental and Clinical Cancer Research*, 38(1), 1–12.
- Liu, M. Y., Hua, W. K., Chiou, Y. Y., Chen, C. J., Yao, C. L., Lai, Y. T., Lin, C. H., & Lin, W. J. (2020). Calcium-dependent methylation by PRMT1 promotes erythroid differentiation through the p38 α MAPK pathway. *FEBS Letters*, 594(2), 301–316.
- Liu, Y., & Schneider, M. F. (2014). FGF2 activates TRPC and Ca²⁺ signaling leading to satellite cell activation. *Frontiers in Physiology*, 5 FEB(February), 1–10.
- Love, M. I., Anders, S., & Huber, W. (2014). Differential analysis of count data - the DESeq2 package. In *Genome Biology* (Vol. 15, Issue 12).
- Luo, Y., Shan, G., Guo, W., Smrt, R. D., Johnson, E. B., Li, X., Pfeiffer, R. L., Szulwach, K. E., Duan, R., Barkho, B. Z., Li, W., Liu, C., Jin, P., & Zhao, X. (2010). Fragile X mental retardation protein regulates proliferation and differentiation of adult neural stem/progenitor cells. *PLoS Genetics*, 6(4).
- Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D. N., Hanspers, K., Miller, R. A., Digles, D., Lopes, E. N., Ehrhart, F., Dupuis, L. J., Winckers, L. A., Coort, S. L., Willighagen, E. L., Evelo, C. T., Pico, A. R., & Kutmon, M. (2021). WikiPathways: connecting communities. *Nucleic Acids Research*, 49(D1), D613–D621.
- Meller, S., van Ellen, A., Gevensleben, H., Bicker, A., Hankeln, T., Gorr, T. A., Sailer, V., Dröge, F., Schröck, F., Bootz, F., Schröck, A., Perner, S., Dietrich, D., & Kristiansen, G. (2016).

Ectopic Myoglobin Expression Is Associated with a Favourable Outcome in Head and Neck Squamous Cell Carcinoma Patients. *Anticancer Research*, 36(12), 6235–6242.

Mizuno, K., & Sloboda, R. D. (2017). Protein arginine methyltransferases interact with intraflagellar transport particles and change location during flagellar growth and resorption. *Molecular Biology of the Cell*, 28(9), 1208–1222.

Morton, M. L., Bai, X., Merry, C. R., Linden, P. A., Khalil, A. M., Leidner, R. S., & Thompson, C. L. (2014). Identification of mRNAs and lincRNAs associated with lung cancer progression using next-generation RNA sequencing from laser micro-dissected archival FFPE tissue specimens. *Lung Cancer*, 85(1), 31–39.

Mukherjee, S., Kolb, M. R. J., Duan, F., & Janssen, L. J. (2012). Transforming growth factor- β evokes Ca^{2+} waves and enhances gene expression in human pulmonary fibroblasts. *American Journal of Respiratory Cell and Molecular Biology*, 46(6), 757–764.

Nähse, V., Christ, L., Stenmark, H., & Campsteijn, C. (2017). The Abscission Checkpoint: Making It to the Final Cut. *Trends in Cell Biology*, 27(1), 1–11.

Oleksiewicz, U., Liloglou, T., Field, J. K., & Xinarianos, G. (2011). Cytoglobin: biochemical, functional and clinical perspective of the newest member of the globin family. *Cellular and Molecular Life Sciences* 2011 68:23, 68(23), 3869–3883.

Ooi, A. T., Gower, A. C., Zhang, K. X., Vick, J. L., Hong, L., Nagao, B., Wallace, W. D., Elashoff, D. A., Walser, T. C., Dubinett, S. M., Pellegrini, M., Lenburg, M. E., Spira, A., & Gomperts, B. N. (2014). Molecular profiling of premalignant lesions in lung squamous cell carcinomas identifies mechanisms involved in stepwise carcinogenesis. *Cancer Prevention Research*, 7(5), 487–495.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3).

Qi, C., Zhang, J., Chen, X., Wan, J., Wang, J., Zhang, P., & Liu, Y. (2017). Hypoxia stimulates neural stem cell proliferation by increasing HIF-1 α expression and activating Wnt/ β -catenin signaling. *Cellular and Molecular Biology*, 63(7), 12.

Quigley, I. K., & Kintner, C. (2017). Rfx2 Stabilizes Foxj1 Binding at Chromatin Loops to Enable

- Multiciliated Cell Gene Expression. *PLoS Genetics*, 13(1), 1–29.
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, 1.
- Rhee, C.-S., Hong, S. K., Min, Y.-G., Lee, C. H., Lee, K. S., Ahn, S.-H., Park, K.-S., & Yi, W. J. (1999). Effects of IL-1 β , TNF- α , and TGF- β on Ciliary Beat Frequency of Human Nasal Ciliated Epithelial Cells in Vitro. *American Journal of Rhinology*, 13(1), 27–30.
- Rothhammer, V., & Quintana, F. J. (2019). The aryl hydrocarbon receptor: an environmental sensor integrating immune responses in health and disease. *Nature Reviews Immunology*, 19(3), 184–197.
- Saucedo, L., Buffa, G. N., Rosso, M., Guillardoy, T., Góngora, A., Munuce, M. J., Vazquez-Levin, M. H., & Marín-Briggiler, C. (2015). Fibroblast growth factor receptors (FGFRs) in human sperm: Expression, functionality and involvement in motility regulation. *PLoS ONE*, 10(5), 1–17.
- Shen, D., Liu, Y., Liu, Y., Wang, T., Yuan, L., Huang, X., & Wang, Y. (2021). Long non-coding RNA EWSAT1 promoted metastasis and actin cytoskeleton changes via miR-24-3p sponging in osteosarcoma. *Journal of Cellular and Molecular Medicine*, 25(2), 716–728.
- Spix, J. K., Chay, E. Y., Block, E. R., & Klarlund, J. K. (2007). Hepatocyte growth factor induces epithelial cell motility through transactivation of the epidermal growth factor receptor. *Experimental Cell Research*, 313(15), 3319–3325.
- Tang, Z., Kang, B., Li, C., Chen, T., & Zhang, Z. (2019). GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Research*, 47(W1), W556–W560.
- Thi Thanh Thuy, L., Morita, T., Yoshida, K., Wakasa, K., Iizuka, M., Ogawa, T., Mori, M., Sekiya, Y., Momen, S., Motoyama, H., Ikeda, K., Yoshizato, K., & Kawada, N. (2011). Promotion of liver and lung tumorigenesis in DEN-treated cytoglobin-deficient mice. *The American Journal of Pathology*, 179(2), 1050–1060.
- Thiery, J. P., Acloque, H., Huang, R. Y. J., & Nieto, M. A. (2009). Epithelial-Mesenchymal Transitions in Development and Disease. *Cell*, 139(5), 871–890.

- Tözser, J., Earwood, R., Kato, A., Brown, J., Tanaka, K., Didier, R., Megraw, T. L., Blum, M., & Kato, Y. (2015). TGF- β Signaling Regulates the Differentiation of Motile Cilia. *Cell Reports*, 11(7), 1000–1007.
- Villa, M., Crotta, S., Dingwell, K. S., Hirst, E. M. A., Gialitakis, M., Ahlfors, H., Smith, J. C., Stockinger, B., & Wack, A. (2016). The aryl hydrocarbon receptor controls cyclin O to promote epithelial multiciliogenesis. *Nature Communications*, 7, 1–11.
- Vlahov, N., Scrace, S., Soto, M. S., Grawenda, A. M., Bradley, L., Pankova, D., Papaspyropoulos, A., Yee, K. S., Buffa, F., Goding, C. R., Timpson, P., Sibson, N., & O'Neill, E. (2015). Alternate RASSF1 transcripts control SRC activity, E-cadherin contacts, and YAP-mediated invasion. *Current Biology*, 25(23), 3019–3034.
- Wakefield, A., Piggott, L., Croston, D., Jiang, W., & Clarkson, R. (2008). Suppression of the NF- κ B cofactor Bcl3 inhibits mammary epithelial cell apoptosis and, in breast tumours, correlates with poor prognosis. *Breast Cancer Research*, 10(S2), O4.
- Wakefield, A., Soukupova, J., Montagne, A., Ranger, J., French, R., Muller, W. J., & Clarkson, R. W. E. (2013). Bcl3 selectively promotes metastasis of ERBB2-driven mammary tumors. *Cancer Research*, 73(2), 745–755.
- Wang, L., Tassiulas, I., Park-Min, K. H., Reid, A. C., Gil-Henn, H., Schlessinger, J., Baron, R., Zhang, J. J., & Ivashkiv, L. B. (2008). “Tuning” of type I interferon-induced Jak-STAT1 signaling by calcium-dependent kinases in macrophages. *Nature Immunology*, 9(2), 186–193.
- Wang, S., Liu, A., Wu, G., Ding, H. F., Huang, S., Nahman, S., & Dong, Z. (2018). The CPLANE protein Intu protects kidneys from ischemia-reperfusion injury by targeting STAT1 for degradation. *Nature Communications*, 9(1), 1–11.
- Wang, T., Li, Y., Wang, W., Tuerhanjiang, A., Wu, Z., Yang, R., Yuan, M., Ma, D., Wang, W., & Wang, S. (2014). Twist2, the key Twist isoform related to prognosis, promotes invasion of cervical cancer by inducing epithelial-mesenchymal transition and blocking senescence. *Human Pathology*, 45(9), 1839–1846.
- Weiterer, S., Schulte, D., Müller, S., Kohlen, T., Uhle, F., Weigand, M. A., & Henrich, M. (2014). Tumor necrosis factor alpha induces a serotonin dependent early increase in ciliary beat

- frequency and epithelial transport velocity in murine tracheae. *PLoS ONE*, 9(3), 1–10.
- Wittenberg, J. B., & Wittenberg, B. A. (2003). Myoglobin function reassessed. *Journal of Experimental Biology*, 206(12), 2011–2020.
- Xie, H., Pallero, M. A., Gupta, K., Chang, P., Ware, M. F., Witke, W., Kwiatkowski, D. J., Lauffenburger, D. A., Murphy-Ullrich, J. E., & Wells, A. (1998). EGF receptor regulation of cell motility: EGF induces disassembly of focal adhesions independently of the motility-associated PLC γ signaling pathway. *Journal of Cell Science*, 111(5), 615–624.
- Xu, H. W., Huang, Y. J., Xie, Z. Y., Lin, L., Guo, Y. C., Zhuang, Z. R., Lin, X. P., Zhou, W., Li, M., Huang, H. H., Wei, X. L., Man, K., & Zhang, G. J. (2013). The expression of cytoglobin as a prognostic factor in gliomas: a retrospective analysis of 88 patients. *BMC Cancer*, 13, 247.
- Yassin, M., Kissow, H., Vainer, B., Joseph, P. D., Hay-Schmidt, A., Olsen, J., & Pedersen, A. E. (2018). Cytoglobin affects tumorigenesis and the expression of ulcerative colitis-associated genes under chemically induced colitis in mice. *Scientific Reports* 2018 8:1, 8(1), 1–16.
- Zhang, B., Liu, Y., Li, Y., Zhe, X., Zhang, S., & Zhang, L. (2018). Neuroglobin promotes the proliferation and suppresses the apoptosis of glioma cells by activating the PI3K/AKT pathway. *Molecular Medicine Reports*, 17(2), 2757–2763.
- Zhang, J., Lan, S. J., Liu, Q. R., Liu, J. M., & Chen, X. Q. (2013). Neuroglobin, a Novel Intracellular Hexa-Coordinated Globin, Functions as a Tumor Suppressor in Hepatocellular Carcinoma via Raf/MAPK/Erk. *Molecular Pharmacology*, 83(5), 1109–1119.
- Zhang, R., Li, J. B., Yan, X. F., Jin, K., Li, W. Y., Xu, J., Zhao, J., Bai, J. H., & Chen, Y. Z. (2018). Increased EWSAT1 expression promotes cell proliferation, invasion and epithelial-mesenchymal transition in colorectal cancer. *European Review for Medical and Pharmacological Sciences*, 22(20), 6801–6808.
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, 10(1).

2.4 Androglobin expression pattern in basal metazoans confirms its conserved functional association with cilia

Carina Osterhof, [REDACTED]: Expression analysis of Androglobin at the root of the metazoan tree (unpublished)

Supplementary files: electronic Appendix

Own contributions to this publication:

- Mining of literature and public repositories to assess the data basis and choose suitable organisms for comparative analysis
- Quality filtering, mapping, and quantification of bulk RNA-Seq data
- Quality filtering, pre-processing and cluster analysis of count matrices from single cell sequencing data; annotation of cell identity via marker genes and literature analysis
- Differential expression analysis of various sample types (bulk RNA-Seq data, single cell clusters) and correlation analysis of gene expression across cell populations
- Gene ontology enrichment analysis of gene lists
- Pseudotime calculation and analysis of developmental trajectories in embryonal datasets
- Visualisation of results (together with [REDACTED])

Project planning, analysis, and interpretation of data as well as wording of the manuscript were realized together with Prof. Dr. T. Hankeln. The project was managed by Prof. Dr. T. Hankeln.

Androglobin expression pattern in basal metazoans confirms its conserved functional association with cilia

Carina Osterhof¹, [REDACTED] and Thomas Hankeln¹

¹Institute of Organismic and Molecular Evolution, Molecular Genetics and Genome Analysis, Johannes Gutenberg University, Mainz, Germany

*correspondence: hankeln@uni-mainz.de

Abstract

Androglobin is the most distinctive member of the globin protein family: its characteristic globin domain is embedded within a large protein of 1500 amino acids, which additionally contains a calpain protease domain and several other domains of currently unknown function. In addition, the globin domain is permutated and interrupted by a calmodulin binding motif. The Adgb gene/protein is highly conserved across the metazoan phylogenetic tree. Initially, mammalian Adgb was thought to be mainly expressed in testis and – at a lower level – in lung tissue, but we could recently show that it is also found in the female reproductive tract. In all these tissues, Adgb is expressed in cell types carrying either motile cilia or flagella. Analysis of Adgb KO mice also revealed a pivotal function of Adgb during late stages of spermatogenesis.

Since it is difficult to observe motile cilia directly in mammalian tissues, we studied Adgb's expression pattern in several other model organisms for developmental studies. Specifically, we analysed publicly available bulk and single cell mRNA sequencing data from a variety of taxa at the root of the metazoan tree. We confirmed that not only the Adgb gene structure, but also its mRNA expression pattern is highly conserved – as is the case for many ciliary proteins. Androglobin was expressed in different types of ciliated or flagellated cells, for example on the surface of embryos in sea urchins, in comb cells of the sea walnut, in ciliated cells of the digestive tract of the starlet sea anemone and of course, in sperm cells of sponges. The morphological similarities between choanocytes, a major expression site of Adgb in sponges, and choanoflagellates could indicate that also these unicellular organisms are a suitable model. All these organisms, either isolated or as a combination, can now be used to further elucidate the exact function of Adgb in flagellar motility and biosynthesis, but we propose to especially focus on the cnidarian *Nematostella vectensis*, since it is genetically manipulable and embryos as well as adults are translucent and thus easy to image.

Introduction

Globins are a family of small metalloproteins, which share a characteristic α -helical protein structure and an iron atom-containing heme prosthetic group (Kendrew et al., 1958; Dickerson & Geis, 1983; Bolognesi et al., 1997). In vertebrates, the best studied globins are the oxygen-supplying respiratory proteins hemoglobin (Hb), present in erythrocytes, and myoglobin (Mb), which is mainly found in skeletal and heart muscle tissue. In the past 20 years, however, several additional representatives of the gene family were discovered in many metazoan species: neuroglobin (Ngb) in neurons of the central and peripheral nervous system of vertebrates (Burmester et al., 2000; Reuss et al., 2002) has possibly a neuroprotective function (Van Acker et al., 2018); cytoglobin (Cygb), which is present in all vertebrate classes and is expressed in extracellular matrix-producing cell types, neurons and melanocytes of mammals (Kawada et al., 2001; Burmester et al., 2004; Fujita et al., 2014); Globin X (GbX) which exists in fish, amphibians and reptiles and may possibly protect membranes from reactive oxygen species (Blank et al., 2011; Koch & Burmester, 2016); Globin E (GbE) as an respiratory protein that ensures oxygen supply in the retina of birds and turtles (Rösner et al., 2005; Schwarze et al., 2015) and possibly in lungfish oocytes (Lüdemann et al., 2019); Globin Y (GbY) as a globin type of unknown function, limited to fish, amphibians, reptiles and monotreme mammals (Fuchs et al., 2006; Schwarze & Burmester, 2013). The substantial amount of work on these novel family members in the past years has led to the conclusion that globins are expressed in many different tissues and cell types where they may fulfill additional, non-classical cellular functions besides oxygen transport and storage, such as the protection against reactive oxygen or nitrogen species, the prevention of hypoxia-triggered apoptosis, and/or the involvement in intracellular signalling processes (Burmester & Hankeln, 2014; Keppner et al., 2020).

In 2012, androglobin (Adgb) was identified as the newest and -at the same time- most peculiar member of the vertebrate globin protein family (Hoogewijs et al., 2012). Adgb is a large chimeric protein of about 1500 amino acids, which contains an embedded globin domain. This globin domain is permuted with respect to its characteristic alpha helices, and it is interrupted by a calmodulin binding motif. Adgb was initially identified bioinformatically in the echinoderm species *Strongylocentrotus purpuratus* and the cephalopod *Branchiostoma floridae*. Subsequent phylogenetic analysis revealed that Adgb homologues exist in nearly

every branch of the metazoan tree. Orthologous copies of the Adgb gene could be found in basal metazoan taxa such as the cnidarian *Nematostella vectensis* (Cnidaria), the placozoan *Trichoplax adherens* (Placozoa) and even in the choanoflagellate *Monosiga brevicollis* (Hoogewijs et al., 2012), which suggests an elementary and possibly conserved function for Adgb in animals. The gene name hints at the predominant expression in mammalian testis tissue, where lower levels of Adgb could be correlated with impaired fertility (Hoogewijs et al., 2012). In addition, we could recently show that Adgb is expressed in multi-ciliated cells of the mammalian lung, brain and the female reproductive tract (Koay et al., 2021).

Upon ablation of flagellar genes, severe phenotypes can be observed: A study in Adgb KO-mice confirmed its pivotal role in reproduction, showing that Adgb ablation leads to fertility impairments and structural changes in sperm morphology (Keppner et al., 2022). These fertility impairments greatly aggravate breeding of these animals and complicate functional studies. In addition, motile ciliary defects in lungs or other tissues are difficult to study directly in mammals, since beat frequencies need to be monitored *in vivo* (Jing et al., 2017) or in sophisticated cell culture systems (Lee et al., 2020). Due to Adgb gene loss events in *Caenorhabditis elegans* and *Drosophila melanogaster*, these two non-mammalian model systems are not suitable for the functional analysis of Adgb as well. Both spermatogenesis and ciliogenesis are extremely conserved processes (Mitchell, 2007). We therefore examined the expression pattern of Adgb in taxa at the root of the metazoan tree, which have recently emerged as new model systems to study highly conserved processes such as regeneration or development (e.g. *Nematostella vectensis*, Röttinger, 2021). Here, we will test whether the gene structure and distinct expression of Adgb is conserved. We intend to identify model organisms which could facilitate unravel Adgb's role in the development or maintenance of metazoan ciliary and flagellar structures.

Material and Methods

Data acquisition

For this study, we only used publicly available transcriptomic data from previously published studies. Raw data for RNA-Seq experiments was downloaded from the EBI² or NCBI³ web servers in fastq-format. Read count matrices of single-cell RNA-Seq data were also downloaded from these repositories or taken from the supplementary data of the corresponding publication. The full list of datasets can be found in Supplementary file 1.

Processing of RNA-Sequencing data

Quality scores of RNA-Seq data were evaluated using fastqc⁴. Adapter trimming and quality filtering was performed using bbdduk from the BBtools suite⁵. Parameters were chosen depending on the sequence length and overall quality of the dataset and selected mode for trimming was retained for all datasets of the same study. Reads were then mapped against the corresponding reference genome (Supplementary table 1) with HISAT2 (Kim et al., 2019) and expression levels quantified with StringTie as transcripts per million (TPM) (Pertea et al., 2015).

Differential gene expression and ontology analysis

Differential gene expression analysis was performed in R using Bioconductor package DESeq2 (Love et al., 2014). Gene ontology enrichment analysis was subsequently performed with g:Profiler (Raudvere et al., 2019).

Processing of single cell RNA-sequencing data

Single cell sequencing data were downloaded as count matrices from the EBI or NCBI web servers. Quality filtering and cluster analysis were performed in R with Seurat 4 (Hao et al., 2021). Cell identities were assigned via expressed marker genes, in accordance with the original publications. Differential expression analysis was performed using Seurat's build-in "find.markers" function with default parameters. Pearson correlation coefficients were calculated from expression matrices. Gene ontology enrichment was performed using

² <https://www.ebi.ac.uk/ena/browser/>

³ <https://www.ncbi.nlm.nih.gov/sra>

⁴ <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

⁵ <https://sourceforge.net/projects/bbmap/>

g:Profiler. A subset of datasets or single clusters were subjected to pseudotime and trajectory analysis using Monocle 3 (Qiu et al., 2017).

Results and Discussion

Data selection

To study the expression profile of *Adgb* in basal metazoans, we first identified organisms that were already studied extensively via transcriptomics due to their important role as model systems. We searched the SRA database for organisms, for which staged developmental profiles were available. Thus, we narrowed down the time point of highest *Adgb* expression during development. We also focused on datasets from untreated wildtype animals to exclude possible influences thereof. Whenever possible, we analysed bigger datasets which included multiple time points or replicates from one working group only to minimize technical variation. We included mRNA -Seq runs of isolated tissues from adult stages, to study spatial expression differences in these taxa. Finally, to study which cell type(s) *Adgb* mRNA expression emanates from, we compared the results from bulk analysis to single cell mRNA-Seq profiles. Meeting these requirements, we analysed RNA-Seq data from the following taxa: *Amphimedon queenslandica* (Porifera), *Mnemiopsis leidyi* (Ctenophora), *Nematostella vectensis* (Cnidaria), and *Strongylocentrotus purpuratus* (Echinoderm) (the full list of datasets can be found in Supplementary file X). A phylogenetic tree describing the evolutionary relationship between these taxa is given in Figure 1 (adapted and modified from: Sebé-Pedrós et al., 2017).

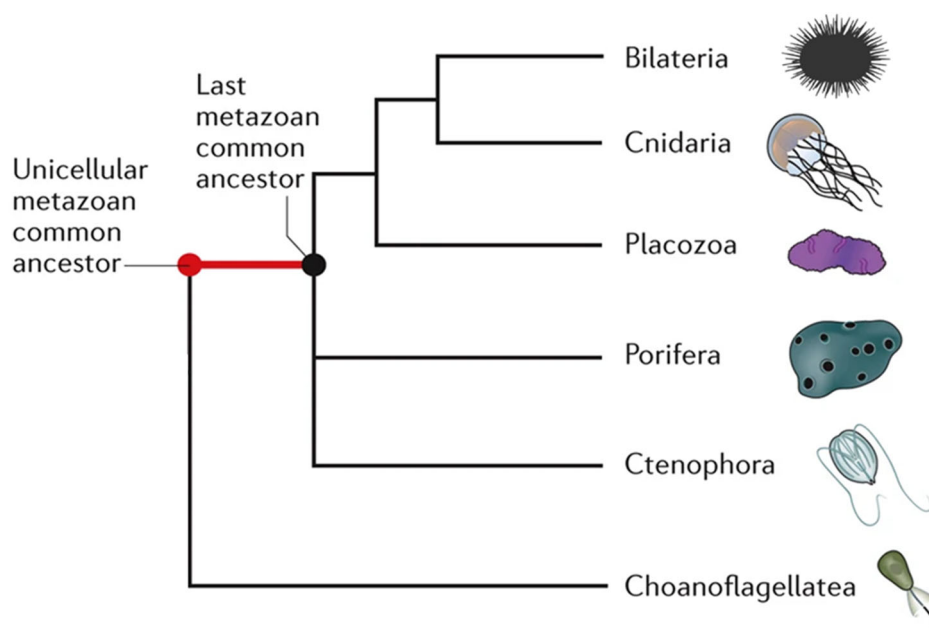


Figure 1: Phylogenetic tree of basal metazoan taxa that carry the *Adgb* gene. Note that the exact position of Porifera and Ctenophora in the metazoan clade is still a matter of debate. Adapted and modified from: Sebé-Pedrós et al., 2017

In the Porifera *Amphimedon queenslandica*, Adgb is associated with fully differentiated ciliated cells

Porifera are a basal metazoan taxon and consist of only a limited number of cell types. First, we reanalysed transcriptomic data on isolated cell populations which are thought to be homologous to cell types present in the last common ancestor of multicellular animals and unicellular holozoans (Sogabe et al., 2019): archeocytes, amoeboid and pluripotent precursor cells; choanocytes, feeding cells that propel water via the characteristic flagellated collar complex; and pinacocytes, epithelial cells which form the inner and outer lining of the sponge (Funayama, 2013). With the only exception of one archeocyte sample, Adgb expression was exclusively detected in choanocytes (Figure 2A). Intriguingly, only 5 out of 10 choanocyte samples were positive for Adgb mRNA, 4 of which were derived from the same individual. Since choanocytes may dedifferentiate into the more stem-cell like archeocytes (Nakanishi et al., 2014; Sogabe et al., 2016), these inter-individual differences could suggest deviating differentiation stages of these cells.

We then analysed single cell mRNA sequencing data derived from adult *A. queenslandica*. Identities of clusters were derived from marker gene expression as described in the original publication (Sebé-Pedrós et al., 2018). Adgb mRNA was expressed in sperm cells and in a second cell cluster termed “collagen cells” (Figure 2B). To assess patterns of co-expression, we performed Pearson correlation analysis of Adgb and all other expressed genes from the global transcriptomic data. GO-enrichment analysis of the TOP100 correlated genes returned several highly significant terms, such as cilium organization, assembly, and motility (full list: Supplementary file 2), reflecting the high expression in sperm cells. We then calculated Pearson correlation coefficients for all genes of the second Adgb expressing cluster. Surprisingly, subsequent GO-term analysis again showed a strong enrichment for cilia-related terms in these gene lists, suggesting a second ciliated cell type that does not cluster with the first (Supplementary file 3, Figure 2D). Terms such as “dynein complex” and “motile cilium” even suggested the presence of moving cilia in these cells. Due to the elongated shape of the cell in the UMAP cluster and the more prominent Adgb expression towards the edge of the cluster, we assumed an underlying developmental process with gradual changes in gene expression. Therefore, we computed trajectories and pseudotime values with Monocle 3 (Qiu et al., 2017). The cells from the collagen cluster could indeed be arranged along a

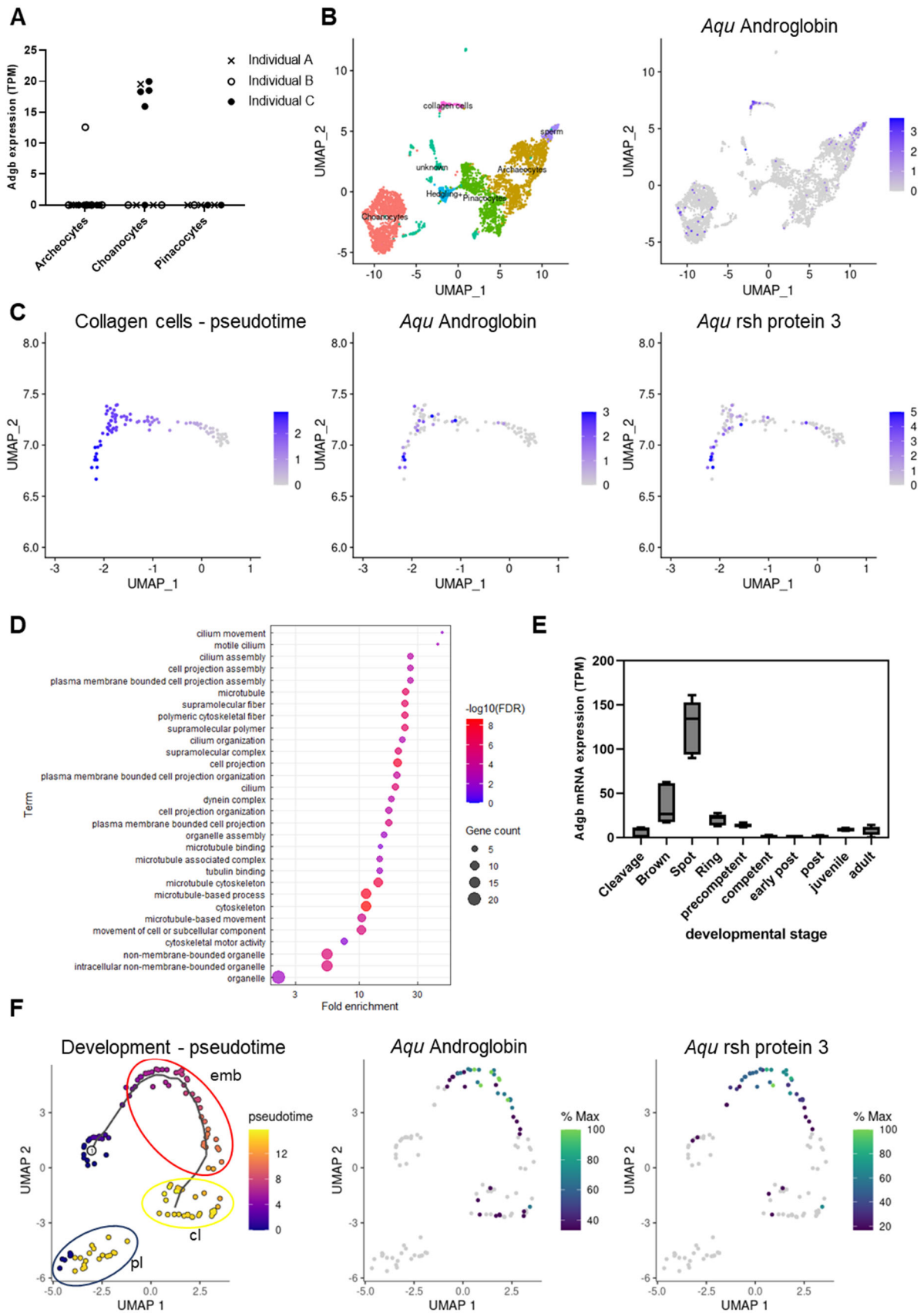
developmental trajectory (Figure 2C), where *Adgb* expression was higher in cells with higher pseudotime values. We also plotted for comparison the expression of a potential radial spoke head gene homolog, which encodes a constituent of motile cilia, and we found the same distribution as for *Adgb*. This indicated that the collagen cluster represents cells differentiating into a cell type carrying motile cilia and that *Adgb* mRNA expression was higher towards the end of this differentiation process. In other demosponges closely related to *A. queenslandica*, a rare cell type was described, which synthesizes high amounts of collagen and is also motile: the lophocyte (reviewed in: Ehrlich et al., 2018; Garrone, 1985). The described collagen cell cluster may thus represent lophocyte cells, however, due to scarce literature we were unable to identify additional potential markers for validation. In contrast to what was found in the bulk sequencing data, *Adgb* was only expressed in very few cells of the choanocyte cluster in the single cell dataset, which could reflect the inter-individual variability in gene expression seen before. However, the cellular assignments of the bulk RNA-Seq data relied on morphological characterization whereas the single cell data were classified via marker gene expression. These datasets could therefore represent differing subsets or subtypes of choanocytes.

Finally, we analysed the expression of *Adgb* mRNA during development of *A. queenslandica* (Figure 2E). *Adgb* mRNA levels rose during early embryogenesis, peaked in the spot stage and then fell rapidly. *Adgb* was weakly expressed in precompetent larvae, but not in competent and post-competent stages. Expression levels rose again in juvenile and adult animals, as also reflected by the single cell sequencing analysis. In the second study, we reanalysed transcriptomic profiles from single embryos. The bulk mRNA samples were clustered and visualised in dimensionally reduced embeddings using Monocle3 and subjected to pseudotime and developmental trajectory analysis (Figure 3F). Competent and post-settlement larval stages were assigned using differentially expressed genes from another transcriptomic study (Conaco et al., 2012), to determine the correct direction of the developmental trajectory. Also in this single embryo dataset, *Adgb* expression peaked during early development (red circle), with residual expression in (pre-)competent larvae. GO term enrichment analysis performed on the top correlated genes showed again that *Adgb* expression coincides with genes enriched for ciliary GO terms, as exemplified by the expression pattern of radial spoke head homolog 3 (Supplementary file 4). Morphological data confirmed this association: the first ciliated cells appear after cleavage in *A. queenslandica*

embryos. During the spot stage, pigment cells accumulate at the posterior pole (the pigment spot) and the ciliated cells migrate and form an outer ciliated layer (Degnan et al., 2005; Borisenko et al., 2019). Studies in two demosponges (*Reneira sp.* and *A. queenslandica*) showed that after settlement, when Adgb expression drops to zero in our analysis, these cells resorb their cilia, wander inwards and transdifferentiate into choanocytes (Leys & Degnan, 2002; Sogabe et al., 2016).

In conclusion, Adgb mRNA was expressed in a variety of ciliated cells from different developmental stages in sponges.

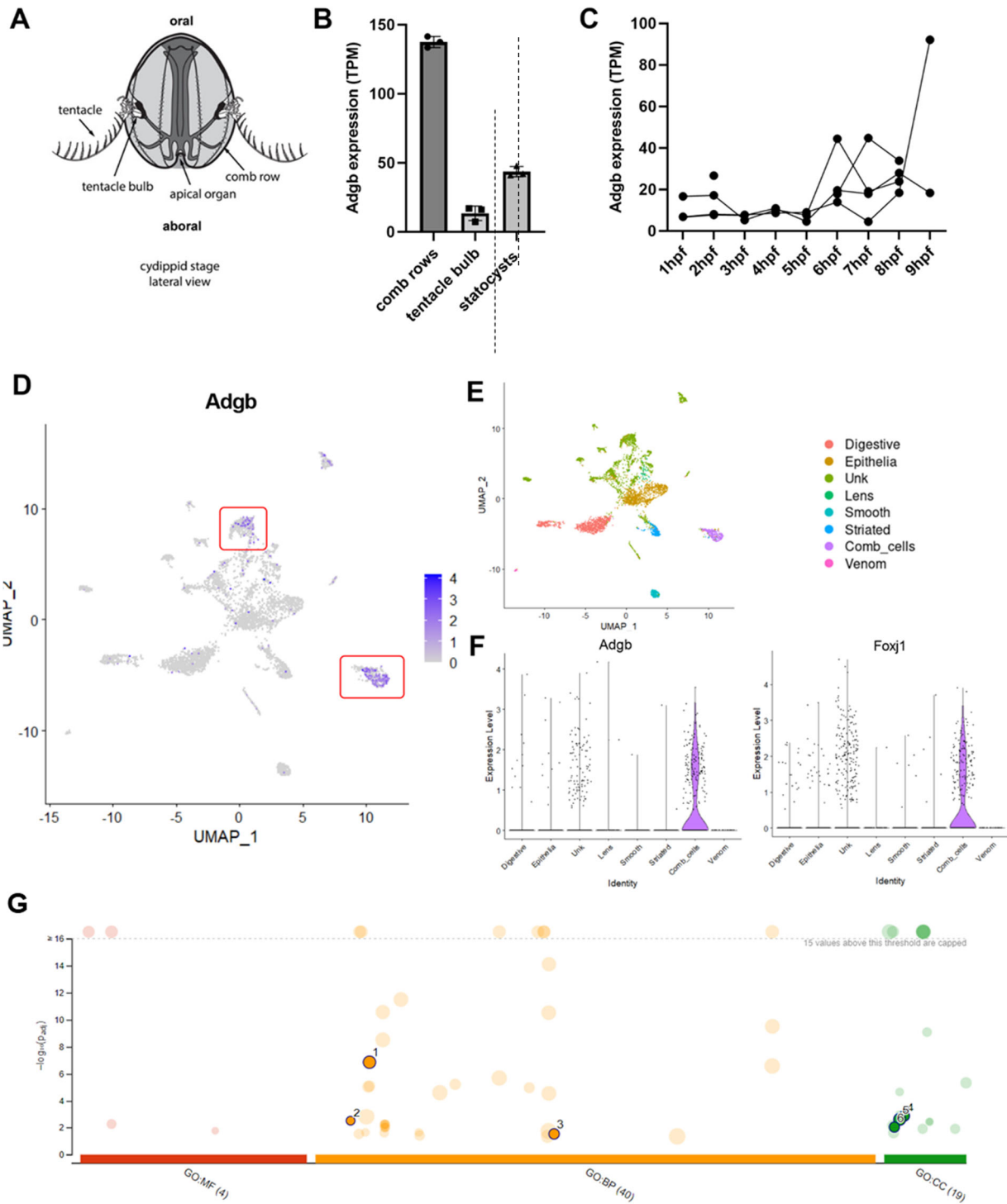
↓**Figure 2:** Expression of Adgb mRNA in different tissues and developmental stages of *A. queenslandica*. (A) Bulk mRNA sequencing data on isolated cell populations show high amounts of Adgb in Choanocytes, but not in every individual. (B) single cell sequencing analysis of an adult specimen detected Adgb mRNA in sperm cells and collagen cells. (C) Pseudotime analysis on collagen cell cluster shows an underlying developmental profile, with higher expression of Adgb and the radial spoke head protein 3 homolog as a ciliary marker in cells with higher pseudotime. (D) GO term enrichment analysis on the TOP100 genes from the collagen cell cluster with highest Pearson correlation coefficient to Adgb showed high occurrence of terms related to motile ciliogenesis, indicating that this cluster rather represented another subtype of ciliated cells. Terms with a size ≤ 5 were excluded from visualisation (E) Adgb mRNA expression derived from bulk RNA sequencing datasets during development of *A. queenslandica*. Adgb expression increased during early development until it peaks in the “spot” stage and decreased again shortly after. (F) Adgb expression in transcriptomes from single embryos during development of *A. queenslandica*. Emb – embryogenesis; cl – competent larvae; pl – post-competent larvae.



Adgb is associated with cilia in a variety of tissues and developmental stages of *Mnemiopsis leidyi*

First, we analysed Adgb mRNA expression during early development of *Mnemiopsis leidyi*. Adgb mRNA expression was lower in early stages and rose – though not in all replicates – from 6 hours post fertilization (hpf) onwards. In 1 replicate from the 9hpf timepoint, expression rose as high as 92 TPM. After 9 hours of development, the first ciliary bands appear on the *M. leidyi* embryos and thus coincide with the surge in Adgb mRNA expression.

We then reanalysed RNA-sequencing data on dissected comb rows and tentacle bulbs (Babonis et al., 2018) as well as statocysts (PRJNA787267, unpublished) from adult specimens. Adgb mRNA expression in the ciliated comb rows was ten times higher than in tentacle bulbs and three times higher than in statocysts (Figure 3B). Subsequent single cell expression analysis of adult *M. leidyi* showed a distinct cluster with a high amount of Adgb mRNA expressing cells (Figure 3D). In accordance with the original publication, we identified this cluster as comb cells. Lower amounts of Adgb could also be identified in second cluster (hereafter called: Adgb cluster 2), corresponding to clusters from the original publication which were not assigned a clear cell type identity. Gene ontology analysis on marker genes from this cluster showed again a strong association with cilia (Figure 3G, full list: Supplementary file 5). A potential homolog of FoxJ1, the master transcription factor of motile ciliogenesis that regulates Adgb in mammals (Koay et al., 2021), was also expressed in this additional clusters (Figure 3F). Differential expression analysis of comb cells and Adgb cluster 2 with subsequent GO term analysis revealed that Adgb cluster 2 showed higher transcription of genes related to ribosome biogenesis and translation, whereas there was no sign for elevated transcriptional activity in comb cells. This could mean that Adgb cluster 2 represents a precursor of comb cells, although it showed no proximity to the comb cell cluster in dimensionality reduced plots and no affinity in trajectory analysis (data not shown). Alternatively, these cells may represent a different type of ciliated cells. Adgb expression is strongly linked to flagella biosynthesis in sperm cells of mammals (Keppner et al., 2022) and could also be found in sperm cells of *A. queenslandica*. In the initial description of this *M. leidyi* dataset, the authors did not clearly identify germ cell clusters. Therefore it is tempting to speculate that Adgb cluster 2 might represent cells differentiating into sperm cells.



ID	Source	Term ID	Term Name	P _{adj} (query_1)
1	GO:BP	GO:0007017	microtubule-based process	1.354×10^{-7}
2	GO:BP	GO:0003341	cilium movement	3.245×10^{-3}
3	GO:BP	GO:0044782	cilium organization	3.032×10^{-2}
4	GO:CC	GO:0030286	dynein complex	1.437×10^{-3}
5	GO:CC	GO:0015630	microtubule cytoskeleton	2.598×10^{-3}
6	GO:CC	GO:0005929	cilium	9.483×10^{-3}

Figure 3: Adgb mRNA expression analysis in the Ctenophore *Mnemiopsis leidyi*. (A) schematic depiction of relevant tissues in adult *M. leidyi*. (B) Adgb mRNA expression in adult tissues. Note that statocysts were analysed in a different study and direct comparison of expression values should be performed carefully. (C) Adgb mRNA expression during early stages of development. (D) Adgb mRNA expression in single cell sequencing data from adult *M. leidyi*. Adgb was expressed in comb cells and a subset of unknown cells (cf. E), and both clusters coexpress the ciliary transcription factor FoxJ1 (F). (G) GO term enrichment analysis based on marker genes from the unknown Adgb expressing clusters showed, among others, strong enrichment for ciliary terms.

Expression pattern of Adgb in tissues of *Nematostella vectensis* and other cnidarians also hints at ciliated cells

During development of the cnidarian *N. vectensis*, Adgb was expressed at low levels throughout all stages. In adult *N. vectensis*, Adgb mRNA expression was 10 times higher in testis tissue than in ovaries whereas there was only little expression in muscle tissue (Figure 4D). There was also moderate expression in mesenterial filaments, an organ associated with digestion. Expression was again higher in male individuals. Since the gonads are in close proximity to these filaments, this could be due to residual testis tissue. However, in female individuals, mesenterial expression of Adgb mRNA was highest among the organs analysed. Histological analysis of mesenterial filaments in another Anthozoan species, *Lophelia pertusa*, described a ciliated cell type next to cnidocytes, mucocytes and secretory cells (Strömberg & Östman, 2017). Either of those could possibly be the source of Adgb mRNA expression in *N. vectensis* mesenteries. In any case, Adgb mRNA expression was also detected in isolated nematosomes. These are small and mobile cellular aggregates containing cnidocytes and a flagellated cell type, but no gland cells (Robson, 1957; Williams, 1979). Transcriptomic analysis on cell fractions enriched for or devoid of cnidocytes from Cnidarian tentacles showed no difference in Adgb expression, indicating that Adgb is not expressed in these cells (Supplementary Table 2, Sunagar et al., 2018). The higher expression in mesenteries and nematosomes is therefore most probably resulting from the presence of cilia in one or more different cell types.

Unfortunately, we could not validate either hypothesis on the single-cell level. No Adgb expression was detected in adult or juvenile samples of *Nematostella vectensis* (Supplementary Figure 1). These datasets were generated by the MARS Seq technique (Jaitin et al., 2014), which possibly detects a lower amount of expressed genes in total and results in a high amount of drop-outs (Ziegenhain et al., 2017). Hence, if endogenous Adgb expression was low, it would be difficult to detect. However, other single cell sequencing studies of adult tissues such as *A. queenslandica* or *M. leidy* discussed above that were also based on this protocol successfully detected Adgb mRNA expression. In a single cell sequencing study on the adult polyp of *Hydra vulgaris*, Adgb mRNA expression was more abundant (Siebert et al., 2019; Supplementary Figure 1C). The highest expression could be found in the male germ line cluster (mgl1+2, upper left corner). In the single-cell study from *N. vectensis*, no germ cells were

identified, only the somatic portion of the gonad. In summary, although single cell evidence is scarce, *Adgb* appears as predominantly expressed in male germ cells of cnidarians and hypothetically in ciliated cell types of the mesenteries.

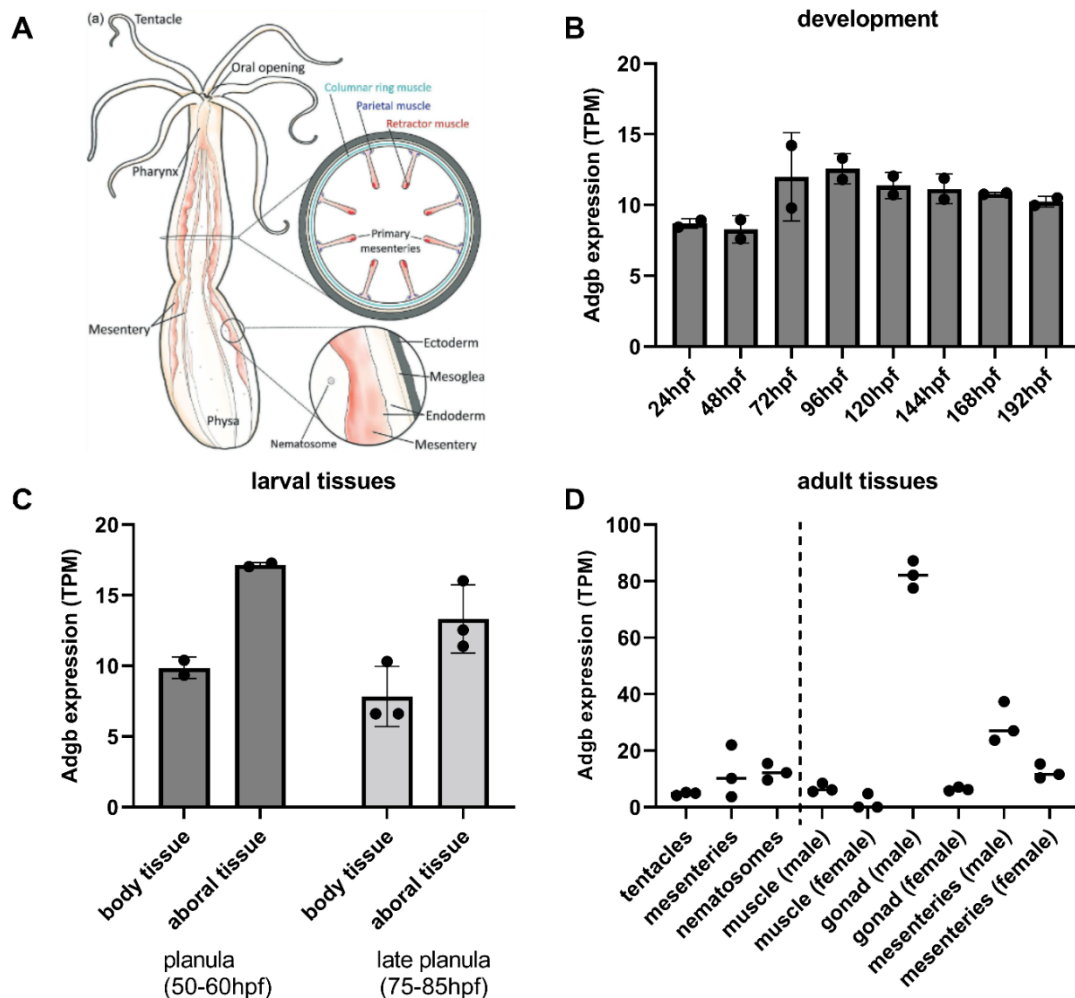


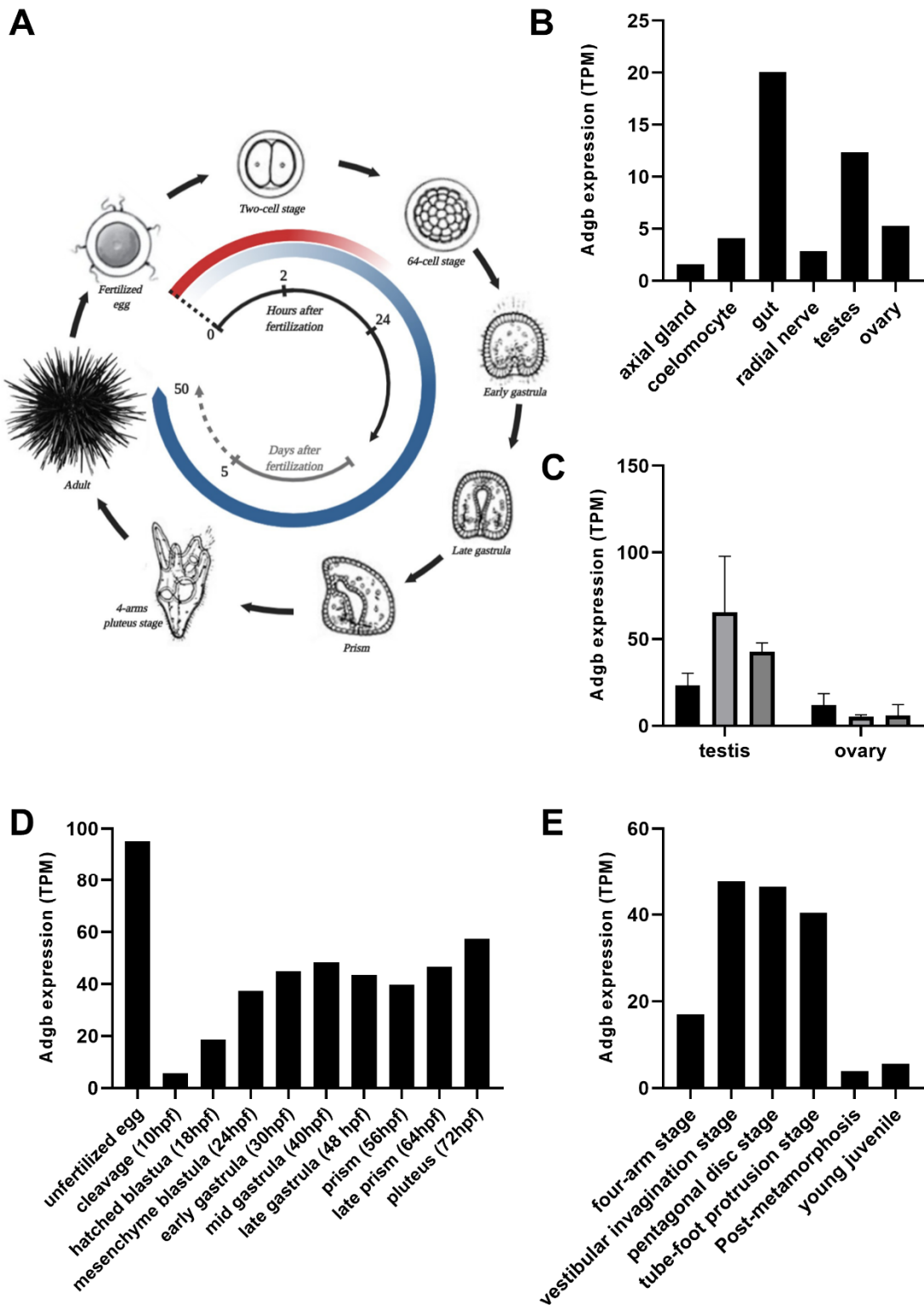
Figure 4: *Adgb* expression analysis in bulk sequencing data of the cnidarian *Nematostella vectensis*. (A) schematic overview of tissues in adult *N. vectensis*. (B) *Adgb* mRNA is constantly expressed during development at low levels. (C) In larvae, *Adgb* mRNA expression is higher in aboral tissues than body tissues in both early and later stages. (D) *Adgb* mRNA expression in adult *N. vectensis* tissues. Dotted line separates the different studies that were analysed. *Adgb* expression is highest in male gonads and mesenteries. In female specimens, *Adgb* expression is highest in mesenteries. Lower levels can be found in nematosomes, whereas there is nearly no expression detectable in tentacles, muscle tissue and female gonads.

Expression of Adgb in the echinoderm *Strongylocentrotus purpuratus*

The first study we reevaluated focused on the early stages of *S. purpuratus* embryogenesis. The amount of Adgb mRNA expression varied greatly during development of *S. purpuratus* (Figure 5D). Adgb was already maternally deposited in unfertilized eggs. The mRNA then vanished quickly within the first few hours. With the onset of embryonal transcription, Adgb mRNA levels rose again during the blastula stage until they peaked during gastrulation (~40 hpf). This surge in expression coincides with the formation of a ciliary tuft at the blastula stage and subsequent ciliation on the whole surface of the early gastrula embryo (Eldon et al., 1990; Cameron et al., 1993). Adgb mRNA levels then remained elevated until pluteus stage. A second study on the transcriptome the larval stages of *S. purpuratus* showed that Adgb mRNA levels surged in the motile larval stages (Figure 5E). In addition, Adgb levels increased from the four arm stage to vestibular invagination stage, where ciliary bands develop into so called epaulets, which assist the larvae in feeding and swimming (Smith et al., 2008). Adgb expression then dropped sharply at the end of metamorphosis when the sea urchin becomes sessile.

In the adult sea urchin, Adgb levels were higher in testis than in ovary tissue, possibly reflecting an association with flagellar structures (Figure 5C). Highest expression levels were found in the gut and low levels of expression in the axial gland or radial nerve (Figure 5B). Intriguingly, histological analysis of the gut from sea urchin *Strongylocentrotus intermedius* showed stratified epithelia, but no signs of cellular protrusions such as microvilli or cilia (Hu et al., 2021), in contrast to, for example, the composition of mesenteries in *N. vectensis*. Unfortunately, no single cell sequencing data from adult *S. purpuratus* specimen was available so that no definitive cell type allocation could be made here.

↓ **Figure 5:** Adgb mRNA expression in different stages of *Strongylocentrotus purpuratus* development. (A) schematic depiction and duration of stages in development. Adgb mRNA expression analysis in different adult tissues (B) and gonads (C) showed high expression levels in testis tissue and the gut. During development (D), Adgb mRNA was deposited in the unfertilized egg. Expression then drops sharply. In later stages, Adgb mRNA levels increased again and were highest during gastrulation and in the late pluteus. In later developmental stages, Adgb levels increased during further development of the larvae and then dropped again upon onset of metamorphosis (E).



Conclusion

In mammals, Adgb is prominently expressed in late stages of spermatogenesis and in ciliated epithelial cells of the female reproductive tract, lung, and brain (Koay et al., 2021). Our work here shows that this expression pattern is highly conserved and that association of Adgb with ciliated cell types can be traced back to the most basal representatives of multicellular organisms. Some expression modalities such as the strong association with fully differentiated ciliated cells such as collagen cells from *A. queenslandica* show striking similarities with our findings in mammals, indicating that also the function of Adgb is highly conserved across the metazoan tree. Arguing for an even deeper phylogenetic origin, Adgb seems also to be present in the ciliated protozoan *Tetrahymena thermophila* (Joachimik et al., 2021), a unicellular eukaryote related to *Paramecia* (Cavalier-Smith, 1993). Whereas effects of Adgb ablation in mammals are severe and lead to infertility and abnormal sperm morphology (Keppner et al., 2022), loss of Adgb in *T. thermophila* only caused minor modification of ciliary motility (Joachimik et al., 2021). It is therefore advisable to expand the set of model organisms for future functional studies of Adgb. Unfortunately, classical model organisms such as *Drosophila melanogaster* or *Caenorhabditis elegans* are not suitable for this, since both taxa lost their copy of Adgb (Hoogewijs et al., 2012) as well as motile cilia on somatic cell types (Choksi et al., 2014). The work presented here can help identifying suitable taxa for future functional analysis. In the simple metazoan *A. queenslandica*, archeocytes differentiate into choanocytes on the one hand and sperm cells on the other. Both differentiated cell types (and possibly the somewhat elusive lophocytes as well) expressed Adgb and would therefore be an interesting model to study Adgb's regulatory mechanisms and to study whether Adgb expression impacts cellular fate or is resulting thereof. However, there are no protocols established for genetic manipulation of *A. queenslandica* and also for no other closely related taxon of the demosponge genus. *N. vectensis*, on the other hand, has been used frequently as a model system for regeneration (Amiel & Röttinger, 2021) and can also be genetically manipulated (Ikmi et al., 2014; Karabulut et al., 2019). Since the expression data in *N. vectensis* was not fully conclusive, it would be interesting to perform in situ expression analysis on this organism to determine the cell type expressing Adgb in mesenteries and to confirm the potential expression in germ cells. *S. purpuratus*, an echinoderm and thus sister clade to the vertebrates, could also be used to study Adgb's function. It is an established model organisms in ciliary research, but in comparison to *A. queenslandica* and *N. vectensis* it is more complex

S. purpuratus also lacks some intriguing properties such as *N. vectensis*' translucency during adulthood which greatly facilitates imaging and thus the analysis of internal cilia (DuBuc et al., 2014). In summary, we propose to focus on *N. vectensis* as a model for future studies of Adgb's function.

It could also be desirable to study Adgb's function not in a multicellular, but unicellular model system. Flagellated choanocytes of sponges, possibly a major expression site of Adgb, are structurally very similar to choanoflagellates. Both cells are ciliated and carry a collar of microvilli around the flagellum. Choanoflagellates are also the closest contemporary relatives of multicellular animals and carry a putative Adgb gene (Hoogewijs et al., 2012). It has been proposed that multicellularity itself evolved from a unicellular, flagellated, and collared ancestor and that choanocytes and choanoflagellates are homologous to each other (e.g. Cavalier-Smith, 2017; Nielsen, 2008). However, recent transcriptomic (Sogabe et al., 2019) and morphological analysis (Mah et al., 2014) suggest that both cell types evolved convergently, and that multicellularity rather is derived from transdifferentiating cells resembling modern, pluripotent stem cells (Sogabe et al., 2019). Nonetheless, the structural resemblance and the possibility of genetic manipulation (Hoffmeyer & Burkhardt, 2016) make choanoflagellates an intriguing candidate to complement functional studies on Adgb in e.g. *N. vectensis*.

Here, we could generate a convincing hypothesis that Adgb's expression pattern is highly conserved over a variety of organisms relying only on publicly available bulk and single cell sequencing data, without performing any additional wet lab experiments. The additional point of view on these data even led to potential new classifications of unknown or poorly described cell clusters. These types of reanalyses can generate new knowledge and help annotating cell cluster identities from sequencing efforts, which do not have the resources to rely on expert annotation in large consortia, such as the Fly Cell Atlas (Li et al., 2022). Now these new hypotheses may be validated with appropriate wet lab experiments to confirm these cell type identities to test whether Adgb can be used as a marker for ciliated and flagellated cells of basal metazoans *in vivo*.

Acknowledgments

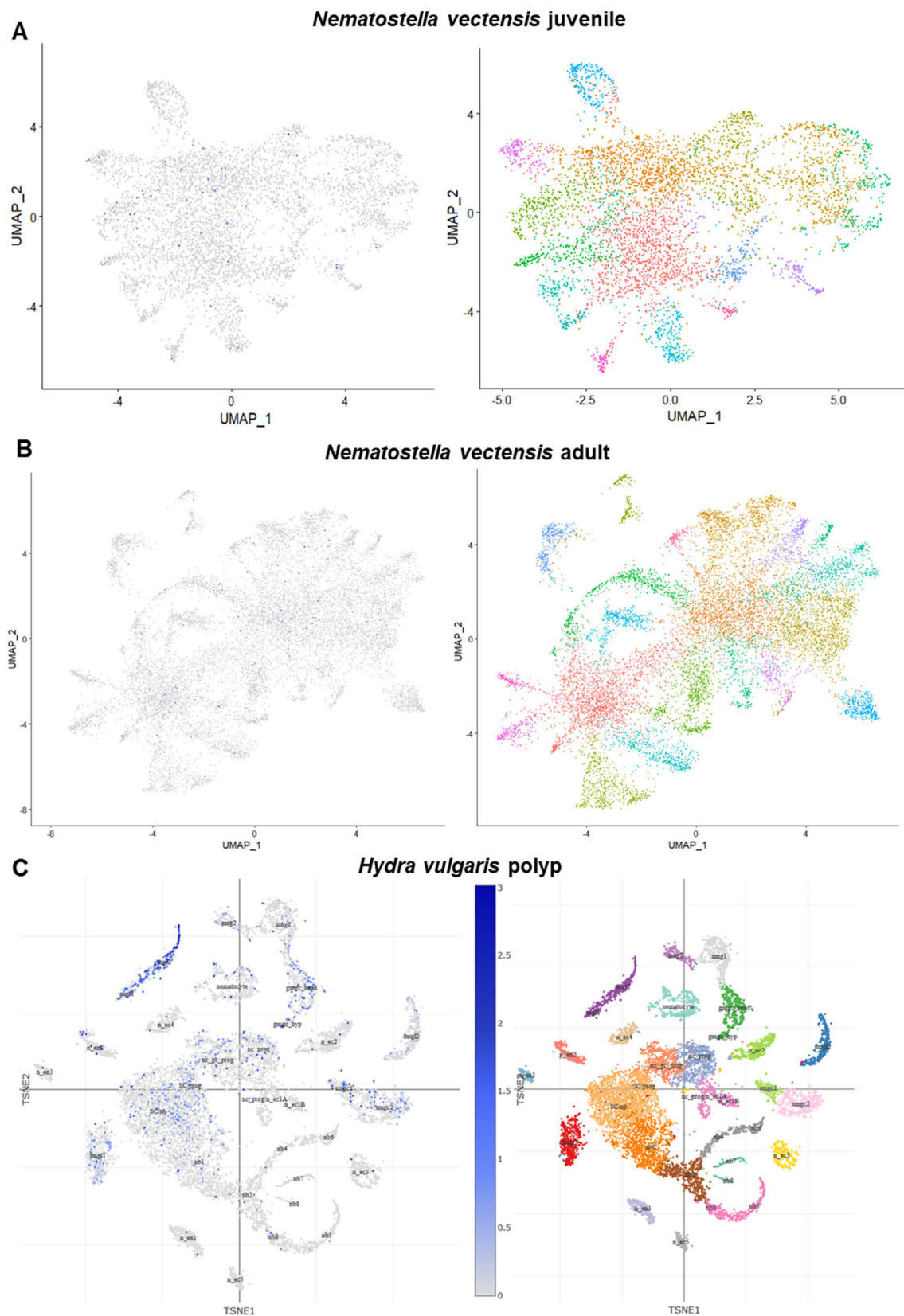
This work was supported by the German Research Foundation to T.H. (HA 2103/9-1).

Supplementary table 2: List of organisms, their corresponding reference genomes and annotation versions used in bulk RNA-Seq analysis in this study.

Organism	Genome version	annotation
Amphimedon queenslandica	Aqu1 (GCA_000090795.1)	Aqu1.52 (gene set v2.1)
Mnemiopsis leidyi	MneLei_Aug2011 (GCA_000226015.1)	MneLei_Aug2011.52
Nematostella vectensis	ASM20922v1 (GCA_000209225.1)	ASM20922v1.52
Strongylocentrotus purpuratus	Spur_5.0 (GCA_000002235.4)	Spur_5.0.49

Supplementary table 2: Adgb mRNA expression in Cnidocytes, remapped and quantified from study PRJNA391807 by Sunagar et al., 2018. There is no detectable difference between the Cnidocyte-positive, -superpositive and -negative samples.

RUN	SAMPLE ID	GROUP	ADGB EXPRESSION (TPM)
SRR5864937	Neg1	negative	4.466142
SRR6706292	Neg2	negative	7.61013
SRR6706323	Neg3	negative	7.219751
SRR6710013	Neg5	negative	5.05338
SRR6710014	Neg6	negative	7.842983
SRR6710017	Neg7	negative	5.67276
SRR6713982	Pos1	positive	2.161914
SRR6713983	Pos2	positive	4.428854
SRR6713984	Pos3	positive	2.994717
SRR6713985	Pos5	superpositive	5.155812
SRR6713986	Pos6	superpositive	2.35182
SRR6713987	Pos7	superpositive	2.752329



Supplementary figure 4: Single cell mRNA expression of Adgb in Cnidarians. Adgb mRNA was not enriched in any of the cell clusters in (A) juvenile and (B) adult *Nematostella vectensis*, but was enriched in male germ cells of the interstitial lineage from *Hydra vulgaris* (Siebert et al., 2019; Figure from Single cell atlas of Broad institute, https://singlecell.broadinstitute.org/single_cell).

References

- Amiel, A. R., & Röttinger, E. (2021). Experimental Tools to Study Regeneration in the Sea Anemone *Nematostella vectensis*. *Methods in Molecular Biology*, 2219, 69–80.
- Babonis, L. S., DeBiasse, M. B., Francis, W. R., Christianson, L. M., Moss, A. G., Haddock, S. H. D., Martindale, M. Q., & Ryan, J. F. (2018). Integrating Embryonic Development and Evolutionary History to Characterize Tentacle-Specific Cell Types in a Ctenophore. *Molecular Biology and Evolution*, 35(12), 2940–2956.
- Blank, M., Wollberg, J., Gerlach, F., Reimann, K., Roesner, A., Hankeln, T., Fago, A., Weber, R. E., & Burmester, T. (2011). A membrane-bound vertebrate globin. *PLoS ONE*, 6(9).
- Bolognesi, M., Bordo, D., Rizzi, M., Tarricone, C., & Ascenzi, P. (1997). Nonvertebrate hemoglobins: Structural bases for reactivity. In *Progress in Biophysics and Molecular Biology* (Vol. 68, Issue 1, pp. 29–68). Elsevier Ltd.
- Borisenko, I., Podgornaya, O. I., & Ereskovsky, A. V. (2019). From traveler to homebody: Which signaling mechanisms sponge larvae use to become adult sponges? *Advances in Protein Chemistry and Structural Biology*, 116, 421–449.
- Burmester, T., Haberkamp, M., Mitz, S., Roesner, A., Schmidt, M., Ebner, B., Gerlach, F., Fuchs, C., & Hankeln, T. (2004). Neuroglobin and Cytoglobin: Genes, Proteins and Evolution. *IUBMB Life*, 56(11), 703–707.
- Burmester, T., & Hankeln, T. (2014). Function and evolution of vertebrate globins. *Acta Physiologica*, 211(3), 501–514.
- Burmester, T., Welch, B., Reinhardt, S., & Hankeln, T. (2000). A vertebrate globin expressed in the brain. *Nature*, 407(6803), 520–523.
- Cameron, R. A., Britten, R. J., & Davidson, E. H. (1993). The embryonic ciliated band of the sea urchin, *Strongylocentrotus purpuratus* derives from both oral and aboral ectoderm. *Developmental Biology*, 160(2), 369–376.
- Cavalier-Smith, T. (1993). Kingdom protozoa and its 18 phyla. *Microbiological Reviews*, 57(4), 953.
- Cavalier-Smith, T. (2017). Origin of animal multicellularity: precursors, causes,

- consequences—the choanoflagellate/sponge transition, neurogenesis and the Cambrian explosion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1713).
- Choksi, S. P., Lauter, G., Swoboda, P., & Roy, S. (2014). Switching on cilia: transcriptional networks regulating ciliogenesis. *Development*, 141(7), 1427–1441.
- Conaco, C., Neveu, P., Zhou, H., Arcila, M. L., Degnan, S. M., Degnan, B. M., & Kosik, K. S. (2012). Transcriptome profiling of the demosponge *Amphimedon queenslandica* reveals genome-wide events that accompany major life cycle transitions. *BMC Genomics*, 13(1), 1–19.
- Degnan, B. M., Leys, S. P., & Larroux, C. (2005). Sponge Development and Antiquity of Animal Pattern Formation. *Integrative and Comparative Biology*, 45(2), 335–341.
- Dickerson, R. E., & Geis, I. (1983). *Hemoglobin : structure, function, evolution, and pathology*. Benjamin/Cummings Pub. Co.
- DuBuc, T. Q., Dattoli, A. A., Babonis, L. S., Salinas-Saavedra, M., Röttinger, E., Martindale, M. Q., & Postma, M. (2014). In vivo imaging of *Nematostella vectensis* embryogenesis and late development using fluorescent probes. *BMC Cell Biology*, 15(1), 1–14.
- Ehrlich, H., Wysokowski, M., Zółtowska-Aksamitowska, S., Petrenko, I., & Jesionowski, T. (2018). Collagens of Poriferan Origin. *Marine Drugs*, 16(3).
- Eldon, E. D., Montpetit, I. C., Nguyen, T., Decker, G., Valdizan, M. C., Klein, W. H., & Brandhorst, B. P. (1990). Localization of the sea urchin Spec3 protein to cilia and Golgi complexes of embryonic ectoderm cells. *Genes & Development*, 4(1), 111–122.
- Fujita, Y., Koinuma, S., De Velasco, M. A., Bolz, J., Togashi, Y., Terashima, M., Hayashi, H., Matsuo, T., & Nishio, K. (2014). Melanoma transition is frequently accompanied by a loss of cytoglobin expression in melanocytes: A novel expression site of cytoglobin. *PLoS ONE*, 9(4).
- Funayama, N. (2013). The stem cell system in demosponges: Suggested involvement of two types of cells: Archeocytes (active stem cells) and choanocytes (food-entrapping flagellated cells). *Development Genes and Evolution*, 223(1–2), 23–38.

- Garrone, R. (1985). The Collagen of the Porifera. In Bairati, A. & R. Garrone (Eds.), *Biology of Invertebrate and Lower Vertebrate Collagens* (pp. 157–175). Springer US.
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., ... Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573–3587.e29.
- Hoffmeyer, T. T., & Burkhardt, P. (2016). Choanoflagellate models — *Monosiga brevicollis* and *Salpingoeca rosetta*. *Current Opinion in Genetics & Development*, *39*, 42–47.
- Hoogewijs, D., Ebner, B., Germani, F., Hoffmann, F. G., Fabrizio, A., Moens, L., Burmester, T., Dewilde, S., Storz, J. F., Vinogradov, S. N., & Hankeln, T. (2012). Androglobin: A Chimeric Globin in Metazoans That Is Preferentially Expressed in Mammalian Testes. *Molecular Biology and Evolution*, *29*(4), 1105–1114.
- Hu, F., Chi, X., Yang, M., Ding, P., Yin, D., Ding, J., Huang, X., Luo, J., Chang, Y., & Zhao, C. (2021). Effects of eliminating interactions in multi-layer culture on survival, food utilization and growth of small sea urchins *Strongylocentrotus intermedius* at high temperatures. *Scientific Reports 2021 11:1*, *11*(1), 1–11.
- Ikmi, A., McKinney, S. A., Delventhal, K. M., & Gibson, M. C. (2014). TALEN and CRISPR/Cas9-mediated genome editing in the early-branching metazoan *Nematostella vectensis*. *Nature Communications*, *5*.
- Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., & Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science (New York, N.Y.)*, *343*(6172), 776–779.
- Jing, J. C., Chen, J. J., Chou, L., Wong, B. J. F., & Chen, Z. (2017). Visualization and Detection of Ciliary Beating Pattern and Frequency in the Upper Airway using Phase Resolved Doppler Optical Coherence Tomography. *Scientific Reports 2017 7:1*, *7*(1), 1–11.
- Joachimiak, E., Osinka, A., Farahat, H., Świdarska, B., Sitkiewicz, E., Poprzeczko, M., Fabczak, H., & Wloga, D. (2021). Composition and function of the C1b/C1f region in the ciliary central apparatus. *Scientific Reports 2021 11:1*, *11*(1), 1–17.

- Karabulut, A., He, S., Chen, C. Y., McKinney, S. A., & Gibson, M. C. (2019). Electroporation of short hairpin RNAs for rapid and efficient gene knockdown in the starlet sea anemone, *Nematostella vectensis*. *Developmental Biology*, 448(1), 7–15.
- Kawada, N., Kristensen, D. B., Asahina, K., Nakatani, K., Minamiyama, Y., Seki, S., & Yoshizato, K. (2001). Characterization of a Stellate Cell Activation-associated Protein (STAP) with Peroxidase Activity Found in Rat Hepatic Stellate Cells. *Journal of Biological Chemistry*, 276(27), 25318–25323.
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610), 662–666.
- Keppner, A., Correia, M., Santambrogio, S., Koay, T. W., Maric, D., Osterhof, C., Winter, D. V., Clerc, A., Stumpe, M., Chalmel, F., Dewilde, S., Odermatt, A., Kressler, D., Hankeln, T., Wenger, R. H., & Hoogewijs, D. (2022). Androglobin, a chimeric mammalian globin, is required for male fertility. *ELife*, 11.
- Keppner, A., Maric, D., Correia, M., Koay, T. W., Orlando, I. M. C., Vinogradov, S. N., & Hoogewijs, D. (2020). Lessons from the post-genomic era: Globin diversity beyond oxygen binding and transport. *Redox Biology*, 37, 101687.
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915.
- Koay, T. W., Osterhof, C., Orlando, I. M. C., Keppner, A., Andre, D., Yousefian, S., Alonso, M. S., Correia, M., Markworth, R., Schödel, J., Hankeln, T., & Hoogewijs, D. (2021). Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis. *Journal of Biological Chemistry*, 296.
- Koch, J., & Burmester, T. (2016). Membrane-bound globin X protects the cell from reactive oxygen species. *Biochemical and Biophysical Research Communications*, 469(2), 275–280.
- Lee, D. D. H., Petris, A., Hynds, R. E., & O’Callaghan, C. (2020). Ciliated epithelial cell differentiation at air-liquid interface using commercially available culture media.

Methods in Molecular Biology (Clifton, N.J.), 2109, 275.

- Leys, S. P., & Degnan, B. M. (2002). Embryogenesis and metamorphosis in a haplosclerid demosponge: Gastrulation and transdifferentiation of larval ciliated cells to choanocytes. *Invertebrate Biology, 121*(3), 171–189.
- Li, H., Janssens, J., de Waegeneer, M., Kolluru, S. S., Davie, K., Gardeux, V., Saelens, W., David, F. P. A., Brbić, M., Spanier, K., Leskovec, J., McLaughlin, C. N., Xie, Q., Jones, R. C., Brueckner, K., Shim, J., Tattikota, S. G., Schnorrer, F., Rust, K., ... Aerts, S. (2022). Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science, 375*(6584).
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12), 550.
- Lüdemann, J., Verissimo, K. M., Dreger, K., Fago, A., Schneider, I., & Burmester, T. (2019). Globin E is a myoglobin-related, respiratory protein highly expressed in lungfish oocytes. *Scientific Reports 2019 9:1, 9*(1), 1–11.
- Mah, J. L., Christensen-Dalsgaard, K. K., & Leys, S. P. (2014). Choanoflagellate and choanocyte collar-flagellar systems and the assumption of homology. *Evolution & Development, 16*(1), 25–37.
- Mitchell, D. R. (2007). The evolution of eukaryotic cilia and flagella as motile and sensory organelles. *Advances in Experimental Medicine and Biology, 607*, 130.
- Nakanishi, N., Sogabe, S., & Degnan, B. M. (2014). Evolutionary origin of gastrulation: Insights from sponge development. *BMC Biology, 12*(1), 1–9.
- Nielsen, C. (2008). Six major steps in animal evolution: Are we derived sponge larvae? *Evolution and Development, 10*(2), 241–257.
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology, 33*(3), 290–295.
- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., & Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods 2017 14:10, 14*(10), 979–982.

- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, 1.
- Reuss, S., Saaler-Reinhardt, S., Weich, B., Wystub, S., Reuss, M. H., Burmester, T., & Hankeln, T. (2002). Expression analysis of neuroglobin mRNA in rodent tissues. *Neuroscience*, 115(3), 645–656.
- Robson, E. A. (1957). A Sea-Anemone from Brackish Water. *Nature* 1957 179:4563, 179(4563), 787–788.
- Rösner, A., Fuchs, C., Hankeln, T., & Burmester, T. (2005). A Globin Gene of Ancient Evolutionary Origin in Lower Vertebrates: Evidence for Two Distinct Globin Families in Animals. *Molecular Biology and Evolution*, 22(1), 12–20.
- Röttinger, E. (2021). *Nematostella vectensis*, an Emerging Model for Deciphering the Molecular and Cellular Mechanisms Underlying Whole-Body Regeneration. *Cells*, 10(10).
- Schwarze, K., Singh, A., & Burmester, T. (2015). The Full Globin Repertoire of Turtles Provides Insights into Vertebrate Globin Evolution and Functions. *Genome Biology and Evolution*, 7(7), 1896–1913.
- Sebé-Pedrós, A., Chomsky, E., Pang, K., Lara-Astiaso, D., Gaiti, F., Mukamel, Z., Amit, I., Hejnol, A., Degnan, B. M., & Tanay, A. (2018). Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nature Ecology and Evolution*, 2(7), 1176–1188.
- Sebé-Pedrós, A., Degnan, B. M., & Ruiz-Trillo, I. (2017). The origin of Metazoa: a unicellular perspective. *Nature Reviews Genetics* 2017 18:8, 18(8), 498–512.
- Siebert, S., Farrell, J. A., Cazet, J. F., Abeykoon, Y., Primack, A. S., Schnitzler, C. E., & Juliano, C. E. (2019). Stem cell differentiation trajectories in *Hydra* resolved at single-cell resolution. *Science*, 365(6451).
- Smith, M. M., Smith, L. C., Cameron, R. A., & Urry, L. A. (2008). The larval stages of the sea urchin, *Strongylocentrotus purpuratus*. *Journal of Morphology*, 269(6), 713–733.
- Sogabe, S., Hatleberg, W. L., Kocot, K. M., Say, T. E., Stoupin, D., Roper, K. E., Fernandez-Valverde, S. L., Degnan, S. M., & Degnan, B. M. (2019). Pluripotency and the origin of

- animal multicellularity. *Nature* 2019 570:7762, 570(7762), 519–522.
- Sogabe, S., Nakanishi, N., & Degnan, B. M. (2016). The ontogeny of choanocyte chambers during metamorphosis in the demosponge *Amphimedon queenslandica*. *EvoDevo*, 7(1), 1–13.
- Sunagar, K., Columbus-Shenkar, Y. Y., Fridrich, A., Gutkovich, N., Aharoni, R., & Moran, Y. (2018). Cell type-specific expression profiling unravels the development and evolution of stinging cells in sea anemone. *BMC Biology*, 16(1), 108.
- Van Acker, Z. P., Luyckx, E., & Dewilde, S. (2018). Neuroglobin Expression in the Brain: a Story of Tissue Homeostasis Preservation. *Molecular Neurobiology* 2018 56:3, 56(3), 2101–2122.
- Williams, R. B. (1979). Studies on the nematosomes of *Nematostella vectensis* Stephenson (Coelenterata: Actiniaria). *Journal of Natural History*, 13(1), 69–80.
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., & Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular Cell*, 65(4), 631-643.e4.

3 Discussion and Outlook

Androglobin, a phylogenetically ancestral member of the globin family, is highly conserved across the metazoan tree and predominantly expressed in mammalian testis tissue, with lower levels of expression in both lung and brain tissue (Hoogewijs et al., 2012). In this work, we provided a detailed description of a novel expression site of Adgb, the female reproductive tract. Adgb is expressed in fallopian tubes and endometrial tissue in humans, mice and cattle. Via re-analysis of publicly available bulk and single cell RNA-Seq data we observed that Adgb expression in lung, brain and the female reproductive tract always originates from fully differentiated cells carrying motile cilia. As a result, Adgb presumably exerts the same function in all these organs and this function is hypothetically also associated with the presence or development of motile cilia. In addition, we have identified potential regulatory elements for the Adgb gene locus. FoxJ1, the master transcription factor for motile ciliogenesis, is co-expressed with Adgb. Our cooperation partners could subsequently proof that FoxJ1 can bind to the Adgb promotor and induce its expression. Overexpression of Adgb also increases the proportion of ciliated cells *in vivo*. Adgb could therefore exhibit a function in ciliogenesis itself (Chapter 2.1; Koay et al., 2021).

In testis tissue, Adgb is detected in round and elongating spermatids and thus closely associated with the emergence of the flagellum, a special form of a cilium pivotal for motility of sperm cells and reproductive success. Upon Adgb ablation, male mice suffer from infertility. We have analysed the transcriptomic changes in Adgb KO testis tissue compared to wildtype littermates and found a small, but distinct set of genes that suggest defects in late stages of spermatogenesis but revealed no pathway in particular. Closer phenotypic analysis conducted by our cooperation partners of spermatids via microscopy showed impaired maturation, abnormal sperm shape and a variety of ultrastructural defects in organisation of both microtubules and mitochondria. Mature sperm displayed a variety of flagellar deformations. Mass spectrometry analysis of immunoprecipitated Adgb also conducted in the Hoogewijs lab suggests an interaction with Sept10. The specific function of this protein is unknown, however, it forms complexes with other members of the septin family which were already described as vital for the proper formation of the connecting midpiece of sperm (Shen et al., 2020). The interaction of Sept10 and Adgb was subsequently confirmed both *in vivo* and *in vitro*. Upon ablation of Adgb, Sept10 also mislocalizes in sperm cells. Mechanistically, Adgb could

contribute to the proteolysis of Sept10, which is required for its activation. All in all, these data suggest a pivotal role for Adgb in the maturation of structurally sound sperm cells (Chapter 2.2; Keppner et al., 2022).

In the initial publication from 2012, the Adgb gene was identified in nearly every branch of the metazoan tree. Given its pivotal function in reproduction, we interrogated whether not only the Adgb gene, but also its expression pattern as well as the association with ciliated or flagellated cells is conserved beyond vertebrates. We therefore analysed Adgb mRNA expression in bulk and single cell data from taxa at the root of the metazoan tree. Adgb was detected in a variety of cell types carrying motile cilia or flagella, such as sperm cells (sponge), comb cells (sea walnut), ciliated cells on the surface of embryos (sea urchin) as well as ciliated cells from the digestive tract (sea anemone). Overall, these results show a strong conservation of Adgb's expression pattern and presumably also its cilia-association function in metazoan taxa (Chapter 2.4).

Many globin family members such as Mb or Cygb have previously been studied in the context of cancer. Depending on the globin family member as well as the cancer type in question, globins may have both oncogenic and protective function in malignancies and can serve as prognostic markers (see also Chapter 1.3). To analyse Adgb's role in cancer, we have stably overexpressed Adgb in a lung cancer cell model. Subsequent transcriptomic analysis showed, among others, hallmarks of enhanced motility, which suggest an oncogenic effect of Adgb – in line with the findings of other studies in brain and prostate cancer models. However, these effects were not pronounced enough to be detectable in phenotypic motility assays. We then analysed a wide variety of transcriptomic data from in-vivo cancer samples. Despite its presumable oncogenic role, we found that Adgb is consistently downregulated in cancer entities derived from tissues with endogenous Adgb expression. Adgb mRNA levels also decreased with degree of malignancy and are not detectable in metastases. We validated these findings via qPCR in an independent cohort of lung and testis cancer samples. We concluded that while Adgb may have oncogenic properties in cell culture models, it is simply not expressed in vivo, such that Adgb undergoes sophisticated regulatory mechanisms that silence the gene upon dedifferentiation of tissues. Adgb is thus not suitable as prognostic marker in cancer (Chapter 2.3).

3.1 Androglobin as part of the ciliary proteome

Androglobin is mainly expressed in cells carrying motile cilia or flagellar structures (Chapter 2.1; Koay et al., 2021) and this expression pattern is highly conserved throughout the metazoan tree (Chapter 2.4). A proteomic study of cilia in a sea urchin (*Strongylocentrotus purpuratus*), a sea anemone (*Nematostella vectensis*), and a choanoflagellate (*Salpingoeca rosetta*) suggested that Adgb is not only expressed in these ciliated cell types, but localizes to the cilium itself (Table 3, Sigg et al., 2017): mass spectrometry analysis on isolated cilia identified peptides derived from Adgb in all three organisms in nearly all replicates analysed (exception: replicate 2 of *S. purpuratus*). The authors also subdivided cilia from *S. purpuratus* into an axonemal fraction and one containing the membrane and matrix of the cilium before analysis via mass spectrometry. Adgb counts tend to be higher in the axonemal fraction of isolated cilia. In *S. rosetta*, Adgb was even exclusively found in ciliary samples, whereas it was not detectable in the cell bodies or the microvilli.

Table 3: Adgb peptide counts in proteomic studies of isolated cilia (from: Sigg et al 2017). Protein composition of motile cilia from *S. purpuratus*, *N. vectensis*, and *S. rosetta* were analysed via mass spectrometry. Peptides derived from Adgb protein could be identified in cilia from all three organisms. No peptides were identified in the cell bodies and microvilli from *S. rosetta*.

Organism (Adgb acc.) →	<i>S. purpuratus</i>	<i>N. vectensis</i>	<i>S. rosetta</i>
Analysed cell fraction ↓	(W4XW92)	(A7RWT3)	(F2TY84)
Whole cilia	4.66 (n=3)	6 (n=2)	2 (n=2)
Axoneme	6 (n=3)	-	-
Membrane + matrix	3.66 (n=3)	-	-
Cell bodies	-	-	0 (n=1)
microvilli	-	-	0 (n=1)

Protein import from the cytosol into the cilium is a tightly controlled process (Kee et al., 2012). It is orchestrated by the ciliary pore complexes (CPC), which carry a high resemblance to the nuclear pore complexes (NPC) that regulate protein traffic into the nucleus. Both structures allow for passive diffusion of smaller proteins; however, large proteins (>~70 kDa) require a signalling peptide for active transport (Kee & Verhey, 2013). Since the deciliation protocol used in the proteomic study by Sigg et al. (2017) excluded the basal body region (Stephens,

1995), a colocalisation to ciliary structures from the cytosolic side is unlikely for Adgb. Due to its large size, the full length Adgb protein likely needs to be transported actively into the cilium.

Upon first discovery, a nuclear localisation signal (NLS) has been predicted bioinformatically at the C-terminus of Adgb (Hoogewijs et al., 2012). However, we could find no evidence of Adgb in the nucleus in a variety of immunofluorescent stainings. In support of this, stainings of *B. taurus* testis sections remained devoid of nuclear Adgb signal (Chapter 2.1). The asynchronous nature of spermatogenesis provides all meiotic stages in one tissue section, which shows that in *B. taurus* there is also no cell cycle-dependent nuclear localisation. So, either the NLS is inactive, or it has a different function. Studies performed on the NLS of the interflagellar transport component KIF17 have indeed shown that it can serve both as nuclear and ciliary localisation signal (CLS). The full-length protein localizes to the cilium, whereas a fragmented version carrying the C-terminus and the signalling peptide only is transported into the nucleus (Dishinger et al., 2010). The NPC and the CPC share common constituents. It is thus tempting to speculate that the NLS of Adgb could in fact be a CLS. Preliminary results of transfection assays with NLS reporter constructs in starved A549 cells could not verify a translocation to the cilium. However, A549 cells carry only non-motile cilia, whereas the proteome studies were conducted in organisms with motile cilia. In addition, we analysed the NLS in isolation of the rest of the protein. The studies performed on KIF17 have already shown that the protein context can be relevant for the mode of transportation (as discussed in: Kee & Verhey, 2013). Studies on the ciliary protein PC1 have for example demonstrated that disruption of its coiled coil motif leads to decreased ciliary localisation efficiency (Su et al., 2015). Adgb does also carry a coiled-coil motif upstream of the putative NLS/CLS (Hoogewijs et al., 2012), which could be relevant for it passing through the CPC.

In future studies, localisation experiments could be performed in a different model, for example in primary human tracheal epithelial cells differentiated under air liquid interphase conditions. These cells develop functioning, beating cilia (e.g.: Coles et al., 2020). First, immunofluorescent stainings should be performed to confirm Adgb's localisation to motile cilia in this system, if possible even by electron microscopy. Via subsequent transfection of different reporter constructs, which carry the NLS/CLS, the C-terminal region upstream and signalling peptide, or the full-length version of Adgb, it can be determined which protein parts are relevant for the correct localisation of Adgb in ciliated cells.

3.2 Loss of Androglobin is an indicator of ciliary or flagellar specialization in metazoans

The Androglobin gene is highly conserved across the metazoan tree. Our work on the expression patterns in basal multicellular animals (Chapter 2.4) shows that not only the gene itself, but also the mode of expression associated with flagellar or ciliary structures is conserved. Since the proper assembly of these structures is pivotal for embryonic development as well as sexual reproduction, and Adgb has already been shown to be important in murine male gamete generation, it is reasonable to assume that Adgb also serves a notable function in these organisms.

Nonetheless, the Adgb gene was lost (or heavily mutated) independently several times during evolution, for example in the model species *Caenorhabditis elegans* (Hoogewijs et al., 2012). In *C. elegans*, loss of Adgb coincides with the absence of motile cilia. Only a subset of sensory neurons is ciliated in this organism and several important substructures, such as the basal body, show specific abnormalities (Nechipurenko & Sengupta, 2017). Sperm in *C. elegans* are motile, but they do not propel themselves via flagellar beating. The cells form pseudopods and migrate in an amoeboid-like fashion. A similar situation is described for *Daphnia pulex*, a crustacean which carries no Adgb ortholog as well. Sperm of *Daphnia* are either tailless or form short pseudopodia (Ebert, 2005). A study on the evolution of the dynein motor protein gene family shows that *C. elegans* lost all but 2 copies of the axonemal dyneins, in line with its lack of motile ciliary structures. A reduction in number of axonemal dynein genes can also be found in Branchiopoda such as *Daphnia*: the two-headed inner-arm dyneins DHC5 and DHC6 are absent here, which suggests structural changes or even impairment of ciliary motility (Kollmar, 2016). The same study identifies additional organisms whose axonemal dynein content is greatly reduced and which could therefore be ideal candidates to test the hypothesis of ciliary reduction with concomitant loss of Adgb: the cnidarian *Kodoa iwatai*, the nematode *Strongyloides ratti*, and the arachnid *Tetranychus urticae* all show the same reduction in axonemal gene diversity as *C. elegans*. These genomes should be searched for Adgb orthologs in future studies.

In *D. melanogaster*, another taxon without an Adgb gene, ciliation is again different than in mammals: *Drosophila* is a model system for basal body and primary cilia research, but all cilia are immotile. Only sperm cells carry flagella, which confer motility during insemination.

Intriguingly, *Drosophila* harbours two species-specific globin genes that are exclusively expressed in the testis: glob2 and glob3 (Gleixner et al., 2012; Prothmann et al., 2020). Preliminary analysis conducted on a RNAi mediated glob3-knockdown strain suggest a function of glob3 in spermatogenesis. Male glob3 knockdown flies show structural irregularities at the post-meiotic stage and display reduced fertility (Prothmann, 2019). The ablation of Adgb in mice leads also to defects in spermatogenesis. As described in Chapter 2.2 (Keppner et al., 2022), spermatids from Adgb KO-mice develop structural deficiencies from the post-meiotic stage onwards. Maturation is impaired and sperm cells exhibit acrosomal defects and malformed flagella. The effect is more pronounced than the phenotypical manifestation of a *Drosophila* glob3, however, residual expression levels and compensation through the paralogous glob2 gene could attenuate severe defects in spermatogenesis. If Adgb is substituted for by glob2 and 3 in *Drosophila*, functional analysis of these two genes could generate additional hypotheses of Adgb's exact mode of action.

Curiously, Adgb is not the only globin whose expression could be linked to ciliated cells: GbY, a globin confined to fish, amphibians, reptiles excluding birds, and platypus, was found in multi-ciliated cells of the epidermis in adult *Xenopus laevis*. RNA-Seq data suggests that Adgb is not co-expressed here. In contrast to this *X. laevis* data, single cell RNA sequencing analysis shows Adgb expression in ciliated cells from a closely related sister species, *Xenopus tropicalis* – while GbY is expressed to a much lesser extent in this species (Michel Seiwert, unpublished data). This might be a second instant were another globin potentially substitutes for Adgb and thus another model system which could be considered for further functional studies.

3.3 Adgb beyond metazoans

As described in Chapter 2.4, Adgb is expressed during the motile blastula or gastrula stages of a variety of marine invertebrate embryos as well as in a diverse set of flagellated or ciliated cell types in adult organisms. In addition, Adgb has been described in Choanoflagellates (Hoogewijs et al., 2012) as well as in the alveolate *Tetrahymena thermophila* – a rather distant relative of the metazoan clade (Burki et al., 2020). In *Tetrahymena*, Adgb ablation leads to mild perturbation of ciliary motility, indicating also some functional conservation (Joachimciak et al., 2021). Despite the large topological distance in the metazoan tree Adgb's function seems to be conserved, which indicates that Adgb could also be present in other protists. Indeed, a search of the NCBI protein database for “Androglobin” in taxa beyond the metazoans resulted in several hits: in the ciliate *Stylonychia lemnae*, the dinoflagellate *Symbiodinium microadriaticum* (both Alveolata), and the haptophyte phytoplankton *Chrysochromulina tobinii*. All these automatically annotated sequences are sufficiently long (~1000-1500 AA) and carry a putative 5' protease domain, as indicated by NCBI's conserved domains feature. A globin domain however was not predicted. Preliminary BLASTp analysis suggests that additional sequences exist in taxa closely related to these examples. In the ciliate clade, putative Adgb sequences were identified in *Halteria grandinella*, *Tetrahymena thermophila*, *Blepharisma stoltei*, *Ichthyophthirius spec* and several taxa from the Paramecium clade. The sequence from *Symbiodinium microadriaticum* resulted in hits in additional *Symbiodinium* taxa, as well as in 2 other dinoflagellates and the Chromideran *Vitrella brassicaformis*. In the haptophyte lineage, 1 additional sequence was identified (results in Appendix).

For orientation, figure 3-1 shows one of the most recent eucaryotic phylogenetic trees (Burki et al., 2020). Putative Adgb sequences could be identified in taxa from at least three “supergroups”: the TSAR, Haptista, and Amorphea clades. Under the assumption that the topology is correct, this distribution implies that the Adgb gene could additionally be present in other phyla, since these three groups do not share a common ancestor. In analogy to metazoans, taxa with the Adgb gene could be ciliated or flagellated and indeed, with the ciliates and the dinoflagellates, two taxa famously known for their appendages were already identified as Adgb-positive above (e.g. Kahl, 1930; Sarjeant, 1974). But flagellar structures are far more widespread and diverse in protists. They occur on nearly every branch, can vary in

size and number and may have associated structures such as loricas or thecas (as reviewed in: Jeuck & Arndt, 2013).

Initially, Adgb was thought to be absent in fungi (Hoogewijs et al., 2012), since no putative Adgb sequences could be identified in the genomes of *Saccharomyces cerevisiae*, *Candida albicans* and *Aspergillus niger*. However, all these taxa belong to the phylum of ascomycetes, which propagate themselves through non-motile ascospores propelled via the wind (e.g. Neiman, 2005). It is possible that the absence of Adgb is again due to a secondary loss here. There are fungal taxa which reproduce via motile zoospores, e.g. Chytridiomycota, Oomycota or Myxomycetes (Webster & Weber, 2007). These groups should be considered first in the search for fungal Adgb sequences in future studies.

In contrast to this, Adgb will probably not be identified in the Amoebozoa sister clade to Opisthokonta from the Amorphea group (Figure 3-1). As the name indicates, Amoebozoa are not flagellated and crawl via protrusion of pseudopodia (amoeboid movement). They share this characteristic with a variety of other taxa. For example, the former clade of Heliozoa consist of ameboid protists with axopodia, stiff pseudopodia stabilized by microtubules. The clade was dissolved after molecular data revealed that the different subgroups spread over the phylogenetic tree (Nikolaev et al., 2004), however, the morphological similarities hint at organisms that could have all lost the Adgb gene. Many amoeboid taxa belong to the superphylum of Rhizaria (Nikolaev et al., 2004), a sister clade to the flagellated alveolates (Burki et al., 2020). This position renders both branches compelling for comparative analyses.

Despite the shared name, prokaryotic and eukaryotic flagellar have nothing in common. Eukaryotic flagellar are intracellular protrusions that carry the canonical microtubular 9+2 conformation and are composed of a variety of proteins (e.g. Smith & Yang, 2004). Motility is conferred though sliding of microtubular filaments via dynein motor proteins, which results in bending of the flagellum (Satir, 1968). Also bacterial flagella have been studied extensively – and have even served as the key witness in a court case on “evolution versus intelligent design” (Pallen & Matzke, 2006) – but they are structurally more simple, extracellular and consist only of a single group of proteins, the flagellins (Namba & Vonderviszt, 1997). Their motion comes from a special motor protein complex at the basal membrane of the flagellum, which translates an ion flux into rotation (as reviewed in: Nakamura & Minamino, 2019). Archaeal flagellar are structurally similar to those of bacteria (Metlina, 2004), but are

assembled from archaea specific genes and thus a product of convergent evolution (Thomas et al., 2001). It has been suggested that these kingdoms developed different forms of flagellar due to energy constraints: the energetically expensive eukaryotic structures can only be sustained by sufficiently big cells, whereas archaea and bacteria had to develop cheaper systems for motility (Schavemaker & Lynch, 2022). Ciliates, for example, could only evolve their high number of cellular appendages due to their larger size in comparison with other unicellular organisms (Schavemaker & Lynch, 2022). Because of these substantial differences in composition, structure, and energy expenditure, we hypothesize that Adgb will not be found in genomes from either Bacteria or Archaea.

It will be very interesting to elucidate whether Adgb indeed occupies a far more basal position in the phylogenetic tree than previously anticipated and to test the hypothesis that it could have already been present, together with the flagellum, in the last common ancestor of all contemporary eucaryotes (Mitchell, 2007).

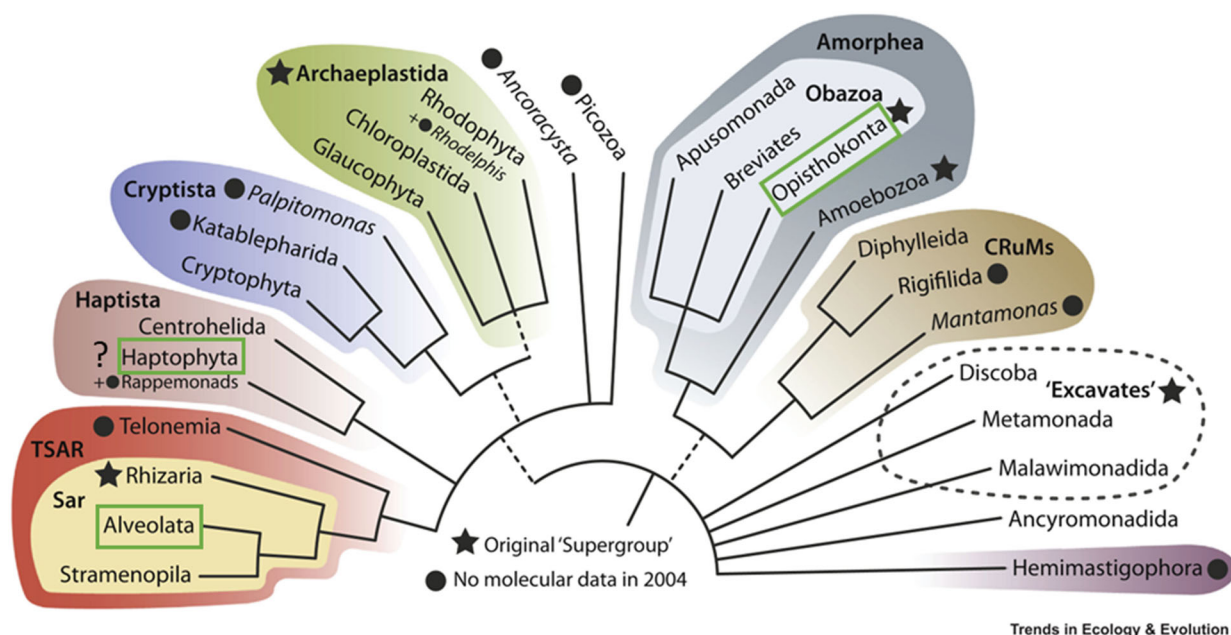


Figure 3-1: Occurrence of putative Adgb sequences across the eucaryotic tree. Adgb was identified in Choanoflagellates and Metazoans (Opisthokonta) and Ciliates (Alveolata). Database search and subsequent BLAST analysis identified putative sequences in Dinoflagellates (Alveolata) and the Haptophyta clade.

3.4 Androglobin: the maverick of the protein family?

In many regards, Adgb can be considered as a distant cousin to the globin protein family. Its multi-domain structure is unique among metazoan globins, and its globin domain is both rearranged and interrupted by a calmodulin binding IQ motif. In addition, the globin domain lacks one of the two crucial histidines needed for proper binding of the prosthetic heme group (although the substituting GlnE7 has been described in other proper globins before). It also diverged very early from the other globin family members, which arose from each other through duplication events (Storz 2018, see also Chapter 1.2). It is therefore unlikely that the full length protein exerts solely a classical globin function, such as NO homeostasis or oxygen storage – even if its absence sometimes coincides with the atypical occurrence of other globin family members.

Androglobin's putative role in the maintenance of motile cilia could hypothetically emerge mainly from its calpain domain. Calpains are calcium-dependent cysteine endopeptidases from the C2 family and activated by autodigestion (Hosfield et al., 1999; Barrett & Rawlings, 2001). Adgb shows highest sequence similarity with human calpain 7, but lacks the cysteine residues which are important for the proteolytic function (Hoogewijs et al., 2012; Rawlings, 2015). The same lack of critical residues has been described for human calpain 6: Intriguingly, calpain 6 is also known to influence ciliogenesis. It is involved in regulation and stabilization of microtubule dynamics in embryonic tissue (Tonami et al., 2007) and calpain 6 deficiency reduces the proportion of ciliated cells *in vitro* (Kim et al., 2019). Overexpression of Adgb can increase the proportion of ciliated cells in a cell culture model (Chapter 2.1), so it would be possible that this effect stems from its calpain protease domain. There is also evidence that Adgb is proteolytically cleaved: in several culture system and with different antibody systems, Western Blot analysis revealed additional bands upon overexpression of Adgb (Chapter 2.2, Bracke et al., 2018).

As also described in Chapter 2.2, a mechanism has been proposed in which Adgb proteolytically cleaves (directly or indirectly) Sept10, and that proper colocalization of both proteins is critical for the formation of flagellar structures. This cleavage is depending on calmodulin (CaM) binding to Adgb's IQ domain, but not influenced by the O₂ level. Overexpression of mutated globin constructs in combination with heme-depleting conditions showed that the interaction of Adgb with CaM is independent of the incorporation of the

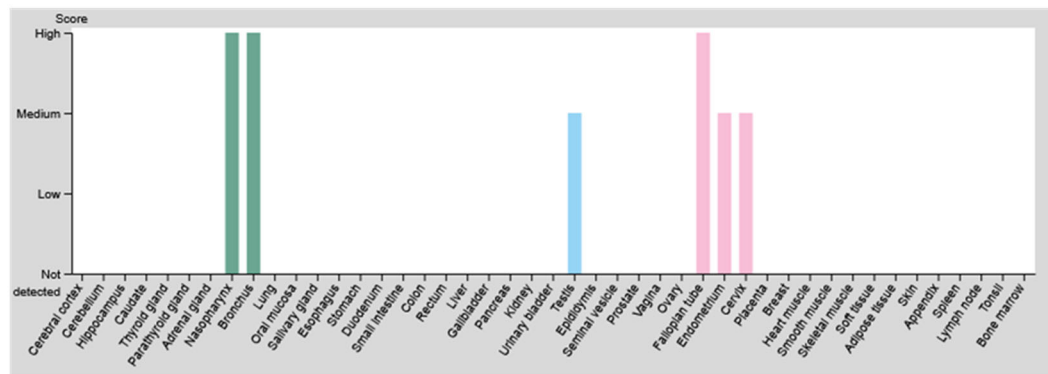
heme (Keppner et al., 2022) – adding further evidence that Adgb should not only be considered as a member of the globin protein family.

As a cautionary remark, all findings except for the direct interaction of Adgb and Sept10 were analysed in vitro through overexpression analysis in a cell culture model without endogenous expression of Adgb. As described in Chapter 2.3, these systems can produce artificial results. Expression data from the human protein atlas (Uhlen et al., 2015; Karlsson et al., 2021) suggests that expression of Adgb and Sept10 is not strictly correlated in human samples (Figure 3-2). Sept10 is expressed ubiquitously in tissues (Figure 3-2 A), but there is no apparent co-expression with ADGB at the cell type level (Figure 3-2 B). However, it would be possible that the mode of normalization used by the HPA could exaggerate cell type specificity and cover low expression levels. This would be in line with the expression data reported in Chapter 2.2 (reanalysed from M. Wang et al., 2018), which shows ubiquitous, but low and constant expression levels for Sept10 during human spermatogenesis. Interestingly, Sept10 also shows a different expression pattern in murine single cell data compared to human samples, where there is a surge in Sept10 expression in round and elongating spermatids, in line with Adgb's increased expression (Chapter 2.2, reanalysed from: Green et al., 2018; Jung et al., 2019; Lukassen et al., 2018). Therefore, it would be interesting to further analyse the expression pattern of Sept10 in murine tissues other than the testis and, if feasible, also in human tissues and sperm cells to see whether this interaction can be translated to the human system.

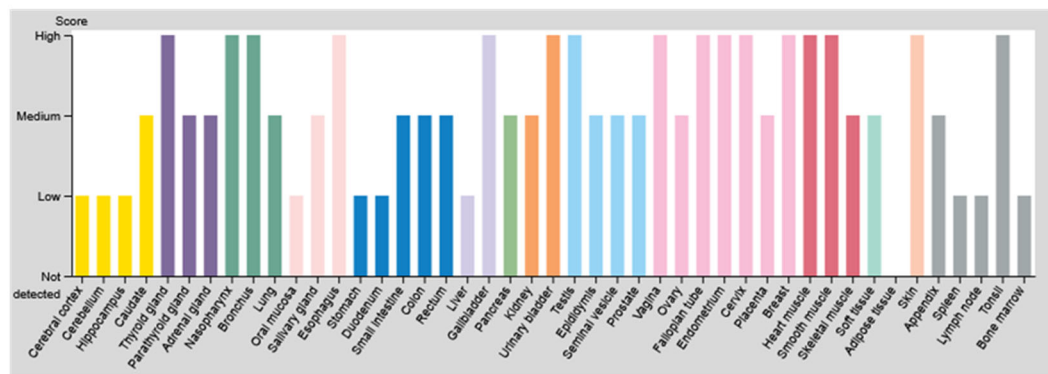
Finally, the binding of the heme group by the globin domain may not be necessary for the direct interaction of Adgb and CaM, however, it could still be possible that it could act as a signalling/activating component. The domain-specific perturbations of Adgb were introduced into these systems via overexpression, and not in the context of the whole Adgb protein. These experiments could be repeated in a more native setting, for example by creating a domain specific knockouts in a system with endogenous Adgb expression. Such a system could be either a more sophisticated mammalian cell model, such as the air liquid interface model introduced above, or a completely different, taxonomically basal organism which would ideally facilitate genomic manipulation.

A Tissue expression data (bulk RNA-Seq and antibody staining)

Adgb (RNA expression cluster 49: ciliated cells -cilium organization)



Sept10 (RNA expression cluster 11: Squamous epithelial cells – Epithelial cell function)

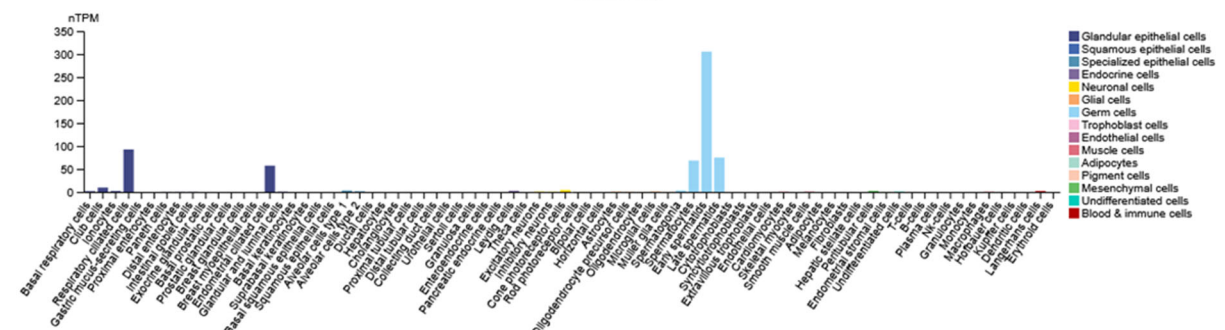


B Cell type expression data (single cell RNA-sequencing)

Single cell types Adgb

RNA single cell type specificity: Cell type enhanced (Early spermatids, Respiratory ciliated cells, Late spermatids, Spermatocytes, Endometrial ciliated cells)

Group Expression Alphabetical



Single cell types Sept10

RNA single cell type specificity: Cell type enhanced (Extravillous trophoblasts, Adipocytes, Cardiomyocytes, Hofbauer cells, Fibroblasts, Cytotrophoblasts, Smooth muscle cells)

Group Expression Alphabetical

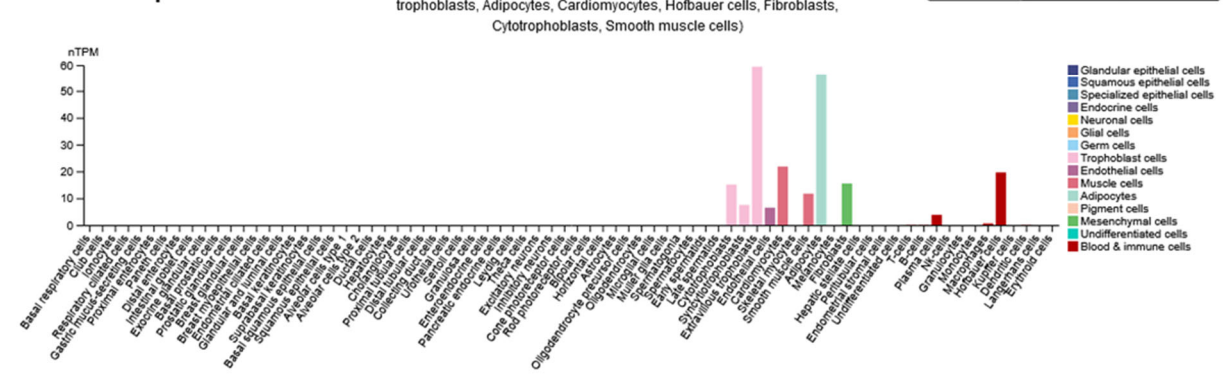


Figure 3-2: Expression data on Sept10 and Adgb from the Human protein atlas. (A) expression values derived from bulk sequencing data and antibody stainings. Adgb is ascribed to the “ciliated cells/cilium organization” mRNA expression cluster; antibody stainings show protein expression in the upper respiratory tract and male and female reproductive organs (in line with our own results, Chapter 2.1). Sept10 is expressed broadly on protein level; it is assigned to the “squamous epithelial cell cluster” deducted from bulk mRNA sequencing data. (B) Expression data on Adgb and Sept10 based on single cell mRNA sequencing data. Adgb expression was found in ciliated cells and during spermatogenesis, whereas Sept10 is expressed in trophoblasts, muscle cells, adipocytes, fibroblasts, and immune cells.

4 Summary

Androglobin is a phylogenetically ancient member of the globin protein family. Its multi-domain structure with rearranged globin domain is unique among metazoan globins. As already indicated by its name, Adgb is mainly expressed in testis tissue and, to a lower extent, in the lung and the brain. Here, we describe the female reproductive tract as a new expression site for Adgb. Adgb is expressed in the fallopian tubes and endometria of different mammals. We analysed public single-cell mRNA-Seq data and found that Adgb expression always originates from cells with motile cilia. We therefore hypothesize that Adgb exerts the same cilia-associated function in all of these organs. In addition, we found that Adgb is strongly co-expressed with FoxJ1, the master transcription factor of ciliogenesis, which is thus an intriguing candidate for the regulation of Adgb. In testis tissue, Adgb is detected in later stages of spermatogenesis and thus closely associated with the production of the flagellum, a structure very similar to motile cilia. The knockout of Adgb in mice leads to male sterility. In this context, we analysed RNA-Seq data from Adgb KO testis compared to wildtype tissue and found a small, but distinct set of deregulated genes. GO enrichment analysis suggested defects in late spermatogenesis but unfortunately did not identify specific deregulated pathways. In the initial publication, the Adgb gene was found on nearly every branch of the metazoan tree. Given its strong cilia association in mammals, we investigated whether its expression pattern is conserved beyond vertebrates. We therefore analysed RNA-Seq data from a variety of basal metazoan taxa and found that Adgb is indeed expressed in many cell types carrying flagella or motile cilia, such as sperm cells from sponges or epidermal ciliated cells in the sea walnut and embryos from sea urchins. These results indicate a strong conservation of Adgb's expression pattern, and we hypothesize that also its cilia-associated function is conserved in metazoans. Many other members of the globin family have been studied in the context of human cancer. Depending on the globin type analysed and also the cancer tissue type, globins can have either oncogenic or suppressive properties in malignancies. We therefore analysed Adgb's role in cancer. Transcriptomic analysis of an Adgb overexpressing A549 lung cancer model initially suggested a possible oncogenic effect of Adgb, in line with previous studies in brain and prostate cancer models. However, a thorough analysis of transcriptomes from cell lines and in vivo cancer entities showed that Adgb is simply not expressed in those tumours, which hints at sophisticated regulatory mechanisms that silence the Adgb gene in cancer. Adgb is therefore not suitable as a prognostic marker in malignancies.

5 Zusammenfassung

Androglobin ist ein phylogenetisch altes Mitglied der Globinproteinfamilie. Seine Multi-Domänen-Struktur mit einer rearrangierten Globin-Domäne ist einzigartig unter den Globinen der Metazoen. Wie sein Name bereits andeutet, wird Adgb vorwiegend im Hodengewebe und in geringerem Maße in der Lunge und im Gehirn exprimiert. In dieser Arbeit wurde der weibliche Fortpflanzungstrakt als ein neuer Expressionsort für Adgb identifiziert. Adgb wird in den Eileitern und Endometrien verschiedener Säugetiere exprimiert. Durch die Analyse öffentlich zugänglicher mRNA-Seq-Daten und single-cell RNA-Seq wurde festgestellt, dass die Adgb-Expression immer in differenzierten Zellen mit motilen Zilien erfolgt. Adgb könnte daher in all diesen Organen die gleiche, zilienassoziierte Funktion ausüben. Zudem wurde beobachtet, dass Adgb stark mit FoxJ1, dem Haupttranskriptionsfaktor der Zilienbildung, co-exprimiert, der somit einen interessanten Kandidaten für die Untersuchung der Regulationsmechanismen von Adgb darstellt. Im Hodengewebe wird Adgb vor allem in den späteren Stadien der Spermatogenese nachgewiesen und steht damit in engem zeitlichem Zusammenhang mit der Bildung des Flagellums, einer Struktur, die den beweglichen Zilien sehr ähnlich ist. Der Knockout von Adgb in Mäusen führt zu männlicher Sterilität. Eine differentielle Expressionsanalyse von RNA-Seq-Daten von Adgb-KO-Hoden im Vergleich zu Wildtypen ergab nur eine kleine, aber sehr distinkte Gruppe von deregulierten Genen. Die GO-Anreicherungsanalyse dieser Liste deutete auf Defekte in den späten Stadien der Spermatogenese hin, aber leider wurde kein spezieller differenziell regulierter Signalweg identifiziert. In der Erstveröffentlichung wurde das Adgb-Gen auf fast jedem Ast des Stammbaums der Metazoen gefunden. Angesichts seiner starken Assoziation mit Zilien in Säugetieren wollten wir untersuchen, ob auch sein Expressionsmuster über die Wirbeltiere hinaus konserviert ist. Bulk- und single-cell RNA-Seq-Daten einer Auswahl basaler Metazoenarten wurden analysiert und zeigten, dass Adgb tatsächlich in vielen verschiedenen Zelltypen exprimiert wird, die Geißeln oder bewegliche Zilien tragen (z. B. in Spermazellen von Schwämmen oder epidermalen Zilienzellen in der Meeresnuss und in Embryonen von Seeigeln). Diese Ergebnisse zeigen, dass das Expressionsmuster von Adgb stark konserviert ist. Dies kann daher auch für seine möglicherweise mit Zilien assoziierte Funktion in Metazoen gelten. Viele andere Mitglieder der Globinfamilie wurden im Zusammenhang mit menschlichem Krebs untersucht. Je nach untersuchtem Globin und auch je nach Art des Krebsgewebes können Globine sowohl onkogene als auch schützende Eigenschaften bei

bösartigen Erkrankungen haben. Daher wurde die mögliche Rolle von Adgb in Krebs untersucht. Die Transkriptomanalyse eines A549-Lungenkrebsmodells, welches Adgb überexprimiert, deutete auf eine onkogene Wirkung von Adgb hin, was im Einklang mit früheren Studien an Gehirn- und Prostatakrebsmodellen steht. Eine umfassende Analyse der Transkriptome von Zelllinien und Krebsentitäten zeigte jedoch, dass Adgb in diesen Entitäten einfach nicht exprimiert wird, was auf komplexe Mechanismen hinweist, die das Adgb-Gen in vivo in Tumoren abschalten. Adgb ist daher nicht als prognostischer Marker für Krebs geeignet.

6 References

- Ascenzi, P., di Masi, A., Leboffe, L., Fiocchetti, M., Nuzzo, M. T., Brunori, M., & Marino, M. (2016). Neuroglobin: From structure to function in health and disease. In *Molecular Aspects of Medicine* (Vol. 52, pp. 1–48). Elsevier Ltd.
- Barrett, A. J., & Rawlings, N. D. (2001). Evolutionary lines of cysteine peptidases. *Biological Chemistry*, 382(5), 727–733.
- Blank, M., & Burmester, T. (2012). Widespread Occurrence of N-Terminal Acylation in Animal Globins and Possible Origin of Respiratory Globins from a Membrane-Bound Ancestor. *Molecular Biology and Evolution*, 29(11), 3553–3561.
- Blank, M., Kiger, L., Thielebein, A., Gerlach, F., Hankeln, T., Marden, M. C., & Burmester, T. (2011). Oxygen supply from the bird's eye perspective: globin E is a respiratory protein in the chicken retina. *The Journal of Biological Chemistry*, 286(30), 26507–26515.
- Blank, M., Wollberg, J., Gerlach, F., Reimann, K., Roesner, A., Hankeln, T., Fago, A., Weber, R. E., & Burmester, T. (2011). A membrane-bound vertebrate globin. *PLoS ONE*, 6(9).
- Blum, A., Wang, P., & Zenklusen, J. C. (2018). SnapShot: TCGA-Analyzed Tumors. *Cell*, 173(2), 530.
- Bracke, A., Hoogewijs, D., & Dewilde, S. (2018). Exploring three different expression systems for recombinant expression of globins: *Escherichia coli*, *Pichia pastoris* and *Spodoptera frugiperda*. *Analytical Biochemistry*, 543(September 2017), 62–70.
- Brunori, M., Giuffrè, A., Nienhaus, K., Nienhaus, G. U., Scandurra, F. M., & Vallone, B. (2005). Neuroglobin, nitric oxide, and oxygen: functional pathways and conformational changes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(24), 8483–8488.
- Burki, F., Roger, A. J., Brown, M. W., & Simpson, A. G. B. (2020). The New Tree of Eukaryotes. *Trends in Ecology & Evolution*, 35(1), 43–55.
- Burmester, T., Ebner, B., Weich, B., & Hankeln, T. (2002). Cytooglobin: A Novel Globin Type Ubiquitously Expressed in Vertebrate Tissues. *Molecular Biology and Evolution*, 19(4), 416–421.

- Burmester, T., & Hankeln, T. (2009). What is the function of neuroglobin? *The Journal of Experimental Biology*, 212(Pt 10), 1423–1428.
- Burmester, T., & Hankeln, T. (2014). Function and evolution of vertebrate globins. *Acta Physiologica*, 211(3), 501–514.
- Burmester, T., Welch, B., Reinhardt, S., & Hankeln, T. (2000). A vertebrate globin expressed in the brain. *Nature*, 407(6803), 520–523.
- Cen, D., Huang, H., Yang, L., Guo, K., & Zhang, J. (2019). Long noncoding RNA STXBP5-AS1 inhibits cell proliferation, migration, and invasion through inhibiting the PI3K/AKT signaling pathway in gastric cancer cells. *OncoTargets and Therapy*, 12, 1929–1936.
- Chen, S., Huang, L., Li, G., Qiu, F., Wang, Y., Yang, C., Pan, J., Wu, Z., Chen, J., & Tian, Y. (2020). LncRNA STXBP5-AS1 suppresses stem cell-like properties of pancreatic cancer by epigenetically inhibiting neighboring androglobin gene expression. *Clinical Epigenetics*, 12(1), 1–12.
- Coles, J. L., Thompson, J., Horton, K. L., Hirst, R. A., Griffin, P., Williams, G. M., Goggin, P., Doherty, R., Lackie, P. M., Harris, A., Walker, W. T., O’callaghan, C., Hogg, C., Lucas, J. S., Blume, C., & Jackson, C. L. (2020). A revised protocol for culture of airway epithelial cells as a diagnostic tool for primary ciliary dyskinesia. *Journal of Clinical Medicine*, 9(11), 1–16.
- Corti, P., Xue, J., Tejero, J., Wajih, N., Sun, M., Stolz, D. B., Tsang, M., Kim-Shapiro, D. B., & Gladwin, M. T. (2016). Globin X is a six-coordinate globin that reduces nitrite to nitric oxide in fish red blood cells. *Proceedings of the National Academy of Sciences of the United States of America*, 113(30), 8538–8543.
- Dishinger, J. F., Kee, H. L., Jenkins, P. M., Fan, S., Hurd, T. W., Hammond, J. W., Truong, Y. N. T., Margolis, B., Martens, J. R., & Verhey, K. J. (2010). Ciliary entry of the kinesin-2 motor KIF17 is regulated by importin-beta2 and RanGTP. *Nature Cell Biology*, 12(7), 703–710.
- Doğan, A., Demirci, S., Kıratlı, B., & Şahin, F. (2017). Cytochrome: a potential marker for adipogenic differentiation in preadipocytes in vitro. *Cytotechnology*, 69(1), 157.
- Dröge, J., Pande, A., Englander, E. W., & Makalowski, W. (2012). Comparative Genomics of Neuroglobin Reveals Its Early Origins. *PLOS ONE*, 7(10), e47972.

- Ebert, D. (2005). Introduction to Daphnia Biology - Ecology, Epidemiology, and Evolution of Parasitism in Daphnia [Internet]. In *Ecology, Epidemiology, and Evolution of Parasitism in Daphnia*.
- Elkholi, I. E., Elsherbiny, M. E., & Emara, M. (2022). Myoglobin: From physiological roles to potential implications in cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1877(3), 188706.
- Elsherbiny, M. E., Shaaban, M., El-Tohamy, R., Elkholi, I. E., Hammam, O. A., Magdy, M., Allalunis-Turner, J., & Emara, M. (2021). Expression of Myoglobin in Normal and Cancer Brain Tissues: Correlation With Hypoxia Markers. *Frontiers in Oncology*, 11, 1561.
- Emara, M., Salloum, N., & Allalunis-Turner, J. (2009). Expression and hypoxic up-regulation of neuroglobin in human glioblastoma cells. *Molecular Oncology*, 3(1), 45–53.
- Emara, M., Turner, A. R., & Allalunis-Turner, J. (2010). Hypoxic regulation of cytoglobin and neuroglobin expression in human normal and tumor tissues. *Cancer Cell International*, 10(1), 1–16.
- Fabrizius, A., Andre, D., Laufs, T., Bicker, A., Reuss, S., Porto, E., Burmester, T., & Hankeln, T. (2016). Critical re-evaluation of neuroglobin expression reveals conserved patterns among mammals. *Neuroscience*, 337, 339–354.
- Fago, A., Giangiacomo, L., D'Avino, R., Carratore, V., Romano, M., Boffi, A., & Chiancone, E. (2001). Hagfish Hemoglobins: Structure, function and oxygen-linked association. *Journal of Biological Chemistry*, 276(29), 27415–27423.
- Fago, A., & Weber, R. E. (1995). The hemoglobin system of the hagfish *Myxine glutinosa*: aggregation state and functional properties. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1249(1), 109–115.
- Fiocchetti, M., Cipolletti, M., Ascenzi, P., & Marino, M. (2018). Dissecting the 17 β -estradiol pathways necessary for neuroglobin anti-apoptotic activity in breast cancer. *Journal of Cellular Physiology*, 233(7), 5087–5103.
- Flonta, S. E., Arena, S., Pisacane, A., Michieli, P., & Bardelli, A. (2009). Expression and functional regulation of myoglobin in epithelial cancers. *The American Journal of Pathology*, 175(1), 201–206.

- Fordel, E., Thijs, L., Martinet, W., Lenjou, M., Laufs, T., Van Bockstaele, D., Moens, L., & Dewilde, S. (2006). Neuroglobin and cytoglobin overexpression protects human SH-SY5Y neuroblastoma cells against oxidative stress-induced cell death. *Neuroscience Letters*, 410(2), 146–151.
- Fuchs, C., Burmester, T., & Hankeln, T. (2006). The amphibian globin gene repertoire as revealed by the *Xenopus* genome. *Cytogenetic and Genome Research*, 112(3–4), 296–306.
- Fujita, Y., Koinuma, S., De Velasco, M. A., Bolz, J., Togashi, Y., Terashima, M., Hayashi, H., Matsuo, T., & Nishio, K. (2014). Melanoma transition is frequently accompanied by a loss of cytoglobin expression in melanocytes: A novel expression site of cytoglobin. *PLoS ONE*, 9(4).
- Gallagher, M. D., & Macqueen, D. J. (2017). Evolution and Expression of Tissue Globins in Ray-Finned Fishes. *Genome Biology and Evolution*, 9(1), 32–47.
- Gleixner, E., Herlyn, H., Zimmerling, S., Burmester, T., & Hankeln, T. (2012). Testes-specific hemoglobins in *Drosophila* evolved by a combination of sub- and neofunctionalization after gene duplication. *BMC Evolutionary Biology*, 12(1), 34.
- Gorr, T. A., Wichmann, D., Pilarsky, C., Theurillat, J. P., Fabrizius, A., Laufs, T., Bauer, T., Koslowski, M., Horn, S., Burmester, T., Hankeln, T., & Kristiansen, G. (2011). Old proteins – new locations: myoglobin, haemoglobin, neuroglobin and cytoglobin in solid tumours and cancer cells. *Acta Physiologica*, 202(3), 563–581.
- Green, C. D., Ma, Q., Manske, G. L., Shami, A. N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S. J., Moore, B. B., Tallquist, M. D., Li, J. Z., & Hammoud, S. S. (2018). A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Developmental Cell*, 46(5), 651-667.e10.
- Hardison, R. C. (1996). A brief history of hemoglobins: plant, animal, protist, and bacteria. *Proceedings of the National Academy of Sciences*, 93(12), 5675–5679.
- Hardison, R. C. (1998). Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *The Journal of Experimental Biology*, 201(Pt 8), 1099–1117.
- Helbo, S., Dewilde, S., Williams, D. R., Berghmans, H., Berenbrink, M., Cossins, A. R., & Fago,

- A. (2012). Functional differentiation of myoglobin isoforms in hypoxia-tolerant carp indicates tissue-specific protective roles. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, 302(6), 693–701.
- Herman, J. L., Challis, C. J., Novák, Á., Hein, J., & Schmidler, S. C. (2014). Simultaneous Bayesian Estimation of Alignment and Phylogeny under a Joint Model of Protein Sequence and Structure. *Molecular Biology and Evolution*, 31(9), 2251–2266.
- Hoffmann, F. G., Opazo, J. C., Hoogewijs, D., Hankeln, T., Ebner, B., Vinogradov, S. N., Bailly, X., & Storz, J. F. (2012). Evolution of the globin gene family in deuterostomes: Lineage-specific patterns of diversification and attrition. *Molecular Biology and Evolution*, 29(7), 1735–1745.
- Hoffmann, F. G., Opazo, J. C., & Storz, J. F. (2010). Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 107(32), 14274–14279.
- Hoffmann, F. G., Storz, J. F., Kuraku, S., Vandewege, M. W., & Opazo, J. C. (2021). Whole-Genome Duplications and the Diversification of the Globin-X Genes of Vertebrates. *Genome Biology and Evolution*, 13(10).
- Hoogewijs, D., Ebner, B., Germani, F., Hoffmann, F. G., Fabrizius, A., Moens, L., Burmester, T., Dewilde, S., Storz, J. F., Vinogradov, S. N., & Hankeln, T. (2012). Androglobin: A Chimeric Globin in Metazoans That Is Preferentially Expressed in Mammalian Testes. *Molecular Biology and Evolution*, 29(4), 1105–1114.
- Hosfield, C. M., Elce, J. S., Davies, P. L., & Jia, Z. (1999). Crystal structure of calpain reveals the structural basis for Ca(2+)-dependent protease activity and a novel mode of enzyme activation. *The EMBO Journal*, 18(24), 6880.
- Huang, B., Lu, Y., Li, X., Zhu, Z., Li, K., Liu, J., Zheng, J., & Hu, Z. (2014). Androglobin knockdown inhibits growth of glioma cell lines. *Int J Clin Exp Pathol*, 7(5), 2179–2184.
- Huang, J., Xie, N., Huang, H., Yao, J., & Hu, W. (2018). Long noncoding RNA STXBP5-AS1 inhibits cell proliferation, migration, and invasion via preventing the PI3K/AKT against STXBP5 expression in non-small-cell lung carcinoma. *Journal of Cellular Biochemistry*, 120(5), 7489–7498.

- Hutter, C., & Zenklusen, J. C. (2018). The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell*, 173(2), 283–285.
- Jeuck, A., & Arndt, H. (2013). A Short Guide to Common Heterotrophic Flagellates of Freshwater Habitats Based on the Morphology of Living Organisms. *Protist*, 164(6), 842–860.
- Joachimiak, E., Osinka, A., Farahat, H., Świdarska, B., Sitkiewicz, E., Poprzeczko, M., Fabczak, H., & Wloga, D. (2021). Composition and function of the C1b/C1f region in the ciliary central apparatus. *Scientific Reports* 2021 11:1, 11(1), 1–17.
- Jung, M., Wells, D., Rusch, J., Ahmad, S., Marchini, J., Myers, S. R., & Conrad, D. F. (2019). Unified single-cell analysis of testis gene regulation and pathology in five mouse strains. *ELife*, 8.
- Kahl, A. (1930). Allgemeiner Teil und Prostomata. In *Urtiere oder Protozoa I, Wimpertiere oder Ciliata (Infusoria)*. Fischer.
- Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., Edfors, F., Oksvold, P., von Feilitzen, K., Zwahlen, M., Arif, M., Altay, O., Li, X., Ozcan, M., Mardonoglu, A., ... Lindskog, C. (2021). A single-cell type transcriptomics map of human tissues. *Science Advances*, 7(31).
- Kawada, N., Kristensen, D. B., Asahina, K., Nakatani, K., Minamiyama, Y., Seki, S., & Yoshizato, K. (2001). Characterization of a Stellate Cell Activation-associated Protein (STAP) with Peroxidase Activity Found in Rat Hepatic Stellate Cells. *Journal of Biological Chemistry*, 276(27), 25318–25323.
- Kee, H. L., Dishinger, J. F., Lynne Blasius, T., Liu, C. J., Margolis, B., & Verhey, K. J. (2012). A size-exclusion permeability barrier and nucleoporins characterize a ciliary pore complex that regulates transport into cilia. *Nature Cell Biology* 2012 14:4, 14(4), 431–437.
- Kee, H. L., & Verhey, K. J. (2013). Molecular connections between nuclear and ciliary import processes. *Cilia*, 2(1), 1–10.
- Keppner, A., Correia, M., Santambrogio, S., Koay, T. W., Maric, D., Osterhof, C., Winter, D. V., Clerc, A., Stumpe, M., Chalmel, F., Dewilde, S., Odermatt, A., Kressler, D., Hankeln, T., Wenger, R. H., & Hoogewijs, D. (2022). Androglobin, a chimeric mammalian globin, is

- required for male fertility. *ELife*, 11.
- Keppner, A., Maric, D., Correia, M., Koay, T. W., Orlando, I. M. C., Vinogradov, S. N., & Hoogewijs, D. (2020). Lessons from the post-genomic era: Globin diversity beyond oxygen binding and transport. *Redox Biology*, 37, 101687.
- Kim, B. H., Kim, D. Y., Oh, S., Ko, J. Y., Rah, G., Yoo, K. H., & Park, J. H. (2019). Deficiency of calpain-6 inhibits primary ciliogenesis. *BMB Reports*, 52(10), 619.
- Koay, T. W., Osterhof, C., Orlando, I. M. C., Keppner, A., Andre, D., Yousefian, S., Alonso, M. S., Correia, M., Markworth, R., Schödel, J., Hankeln, T., & Hoogewijs, D. (2021). Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis. *Journal of Biological Chemistry*, 296.
- Koch, J., & Burmester, T. (2016). Membrane-bound globin X protects the cell from reactive oxygen species. *Biochemical and Biophysical Research Communications*, 469(2), 275–280.
- Kollmar, M. (2016). Fine-Tuning Motile Cilia and Flagella: Evolution of the Dynein Motor Proteins from Plants to Humans at High Resolution. *Molecular Biology and Evolution*, 33(12), 3249–3267.
- Kono, H., Hosomura, N., Amemiya, H., Kawaida, H., Furuya, S., Akaike, H., Kawaguchi, Y., Sudo, M., & Ichikawa, D. (2021). Cytoglobin as a Prognostic Factor for Pancreatic Ductal Adenocarcinoma. *Pancreas*, 50(7), 994–999.
- Kristiansen, G., Hu, J., Wichmann, D., Stiehl, D. P., Rose, M., Gerhardt, J., Bohnert, A., Ten Haaf, A., Moch, H., Raleigh, J., Varia, M. A., Subarsky, P., Scandurra, F. M., Gnaiger, E., Gleixner, E., Bicker, A., Gassmann, M., Hankeln, T., Dahl, E., & Gorr, T. A. (2011). Endogenous myoglobin in breast cancer is hypoxia-inducible by alternative transcription and functions to impair mitochondrial activity: a role in tumor suppression? *The Journal of Biological Chemistry*, 286(50), 43417–43428.
- Kristiansen, G., Rose, M., Geisler, C., Fritzsche, F. R., Gerhardt, J., Lüke, C., Ladhoff, A. M., Knüchel, R., Dietel, M., Moch, H., Varga, Z., Theurillat, J. P., Gorr, T. A., & Dahl, E. (2010). Endogenous myoglobin in human breast cancer is a hallmark of luminal cancer phenotype. *British Journal of Cancer*, 102(12), 1736–1745.

- Kugelstadt, D., Haberkamp, M., Hankeln, T., & Burmester, T. (2004). Neuroglobin, cytoglobin, and a novel, eye-specific globin from chicken. *Biochemical and Biophysical Research Communications*, 325(3), 719–725.
- Li, R. C., Morris, M. W., Lee, S. K., Pouranfar, F., Wang, Y., & Gozal, D. (2007). Neuroglobin protects PC12 cells against oxidative stress. *Brain Research*, 1190(1), 159–166.
- Liu, X., El-Mahdy, M. A., Boslett, J., Varadharaj, S., Hemann, C., Abdelghany, T. M., Ismail, R. S., Little, S. C., Zhou, D., Thuy, L. T. T., Kawada, N., & Zweier, J. L. (2017). Cytoglobin regulates blood pressure and vascular tone through nitric oxide metabolism in the vascular wall. *Nature Communications* 2017 8:1, 8(1), 1–14.
- Lüdemann, J., Verissimo, K. M., Dreger, K., Fago, A., Schneider, I., & Burmester, T. (2019). Globin E is a myoglobin-related, respiratory protein highly expressed in lungfish oocytes. *Scientific Reports* 2019 9:1, 9(1), 1–11.
- Lukassen, S., Bosch, E., Ekici, A. B., & Winterpacht, A. (2018). Characterization of germ cell differentiation in the male mouse through single-cell RNA sequencing. *Scientific Reports*, 8(1), 8–14.
- Mannino, M. H., Patel, R. S., Eccardt, A. M., Perez Magnelli, R. A., Robinson, C. L. C., Janowiak, B. E., Warren, D. E., & Fisher, J. S. (2019). Myoglobin as a versatile peroxidase: Implications for a more important role for vertebrate striated muscle in antioxidant defense. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 234, 9–17.
- Mathai, C., Jourd'heuil, F. L., Lopez-Soler, R. I., & Jourd'heuil, D. (2020). Emerging perspectives on cytoglobin, beyond NO dioxygenase and peroxidase. *Redox Biology*, 32, 101468.
- Meller, S., van Ellen, A., Gevensleben, H., Bicker, A., Hankeln, T., Gorr, T. A., Sailer, V., Dröge, F., Schröck, F., Bootz, F., Schröck, A., Perner, S., Dietrich, D., & Kristiansen, G. (2016). Ectopic Myoglobin Expression Is Associated with a Favourable Outcome in Head and Neck Squamous Cell Carcinoma Patients. *Anticancer Research*, 36(12), 6235–6242.
- Metlina, A. L. (2004). Bacterial and archaeal flagella as prokaryotic motility organelles. *Biochemistry (Moscow)* 2004 69:11, 69(11), 1203–1212.
- Mitchell, D. R. (2007). The evolution of eukaryotic cilia and flagella as motile and sensory

- organelles. *Advances in Experimental Medicine and Biology*, 607, 130.
- Nakamura, S., & Minamino, T. (2019). Flagella-Driven Motility of Bacteria. *Biomolecules*, 9(7).
- Nakatani, K., Okuyama, H., Shimahara, Y., Saeki, S., Kim, D. H., Nakajima, Y., Seki, S., Kawada, N., & Yoshizato, K. (2004). Cytooglobin/STAP, its unique localization in splanchnic fibroblast-like cells and function in organ fibrogenesis. *Laboratory Investigation; a Journal of Technical Methods and Pathology*, 84(1), 91–101.
- Namba, K., & Vonderviszt, F. (1997). Molecular architecture of bacterial flagellum. *Quarterly Reviews of Biophysics*, 30(1), 1–65.
- Nechipurenko, I. V., & Sengupta, P. (2017). The rise and fall of basal bodies in the nematode *Caenorhabditis elegans*. *Cilia*, 6(1), 1–6.
- Neiman, A. M. (2005). Ascospore Formation in the Yeast *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews*, 69(4), 565.
- Nikolaev, S. I., Berney, C., Fahrni, J. F., Bolivar, I., Polet, S., Mylnikov, A. P., Aleshin, V. V., Petrov, N. B., & Pawlowski, J. (2004). From the Cover: The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 101(21), 8066.
- Nishi, H., Inagi, R., Kawada, N., Yoshizato, K., Mimura, I., Fujita, T., & Nangaku, M. (2011). Cytooglobin, a novel member of the globin family, protects kidney fibroblasts against oxidative stress under ischemic conditions. *The American Journal of Pathology*, 178(1), 128–139.
- Oleksiewicz, U., Liloglou, T., Field, J. K., & Xinarianos, G. (2011). Cytooglobin: biochemical, functional and clinical perspective of the newest member of the globin family. *Cellular and Molecular Life Sciences* 2011 68:23, 68(23), 3869–3883.
- Opazo, J. C., Lee, A. P., Hoffmann, F. G., Toloza-Villalobos, J., Burmester, T., Venkatesh, B., & Storz, J. F. (2015). Ancient duplications and expression divergence in the globin gene superfamily of vertebrates: Insights from the elephant shark genome and transcriptome. *Molecular Biology and Evolution*, 32(7), 1684–1694.
- Pallen, M. J., & Matzke, N. J. (2006). From The Origin of Species to the origin of bacterial

- flagella. *Nature Reviews Microbiology* 2006 4:10, 4(10), 784–790.
- Petersen, M. G., Dewilde, S., & Fago, A. (2008). Reactions of ferrous neuroglobin and cytoglobin with nitrite under anaerobic conditions. *Journal of Inorganic Biochemistry*, 102(9), 1777–1782.
- Prothmann, A. (2019). *Analyse der Evolution und Funktion von Globinen in Arthropoden* [JGU Mainz].
- Prothmann, A., Hoffmann, F. G., Opazo, J. C., Herbener, P., Storz, J. F., Burmester, T., & Hankeln, T. (2020). The Globin Gene Family in Arthropods: Evolution and Functional Diversity. *Frontiers in Genetics*, 11, 858.
- Qiu, Y., Maillett, D. H., Knapp, J., Olson, J. S., & Riggs, A. F. (2000). Lamprey hemoglobin. Structural basis of the bohr effect. *The Journal of Biological Chemistry*, 275(18), 13517–13528.
- Rawlings, N. D. (2015). Bacterial calpains and the evolution of the calpain (C2) family of peptidases. *Biology Direct*, 10(1).
- Reeder, B. J., Svistunenko, D. A., & Wilson, M. T. (2011). Lipid binding to cytoglobin leads to a change in haem co-ordination: a role for cytoglobin in lipid signalling of oxidative stress. *The Biochemical Journal*, 434(3), 483–492.
- Reuss, S., Saaler-Reinhardt, S., Weich, B., Wystub, S., Reuss, M. H., Burmester, T., & Hankeln, T. (2002). Expression analysis of neuroglobin mRNA in rodent tissues. *Neuroscience*, 115(3), 645–656.
- Ribeiro, E. A., & Ramos, C. H. I. (2005). Circular permutation and deletion studies of myoglobin indicate that the correct position of its N-terminus is required for native stability and solubility but not for native-like heme binding and folding. *Biochemistry*, 44(12), 4699–4709.
- Rösner, A., Fuchs, C., Hankeln, T., & Burmester, T. (2005). A Globin Gene of Ancient Evolutionary Origin in Lower Vertebrates: Evidence for Two Distinct Globin Families in Animals. *Molecular Biology and Evolution*, 22(1), 12–20.
- Sarjeant, W. A. S. (1974). Living Dinoflagellates: Morphology. In *Fossil and Living*

- Dinoflagellates* (pp. 13–32). Academic Press.
- Satir, P. (1968). STUDIES ON CILIA : III. Further Studies on the Cilium Tip and a “Sliding Filament” Model of Ciliary Motility. *The Journal of Cell Biology*, 39(1), 77.
- Schavemaker, P. E., & Lynch, M. (2022). Flagellar energy costs across the tree of life. *ELife*, 11.
- Schwarze, K., & Burmester, T. (2013). Conservation of globin genes in the “living fossil” *Latimeria chalumnae* and reconstruction of the evolution of the vertebrate globin family. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1834(9), 1801–1812.
- Schwarze, K., Campbell, K. L., Hankeln, T., Storz, J. F., Hoffmann, F. G., & Burmester, T. (2014). The globin gene repertoire of lampreys: Convergent evolution of hemoglobin and myoglobin in jawed and jawless vertebrates. *Molecular Biology and Evolution*, 31(10), 2708–2721.
- Schwarze, K., Singh, A., & Burmester, T. (2015). The Full Globin Repertoire of Turtles Provides Insights into Vertebrate Globin Evolution and Functions. *Genome Biology and Evolution*, 7(7), 1896–1913.
- Shen, Y. R., Wang, H. Y., Tsai, Y. C., Kuo, Y. C., Wu, S. R., Wang, C. Y., & Kuo, P. L. (2020). The SEPT12 complex is required for the establishment of a functional sperm head-tail junction. *Molecular Human Reproduction*, 26(6), 402–412.
- Sigg, M. A., Menchen, T., Lee, C., Johnson, J., Jungnickel, M. K., Choksi, S. P., Garcia, G., Busengdal, H., Dougherty, G. W., Pennekamp, P., Werner, C., Rentzsch, F., Florman, H. M., Krogan, N., Wallingford, J. B., Omran, H., & Reiter, J. F. (2017). Evolutionary Proteomics Uncovers Ancient Associations of Cilia with Signaling Pathways. *Developmental Cell*, 43(6), 744–762.e11.
- Smagghe, B. J., Trent, J. T., & Hargrove, M. S. (2008). NO dioxygenase activity in hemoglobins is ubiquitous in vitro, but limited by reduction in vivo. *PLoS One*, 3(4).
- Smith, E. F., & Yang, P. (2004). The Radial Spokes and Central Apparatus: Mechano-Chemical Transducers That Regulate Flagellar Motility. *Cell Motility and the Cytoskeleton*, 57(1), 8.
- Sosa, V., Moliné, T., Somoza, R., Paciucci, R., Kondoh, H., & Lleonart, M. E. (2013). Oxidative stress and cancer: an overview. *Ageing Research Reviews*, 12(1), 376–390.

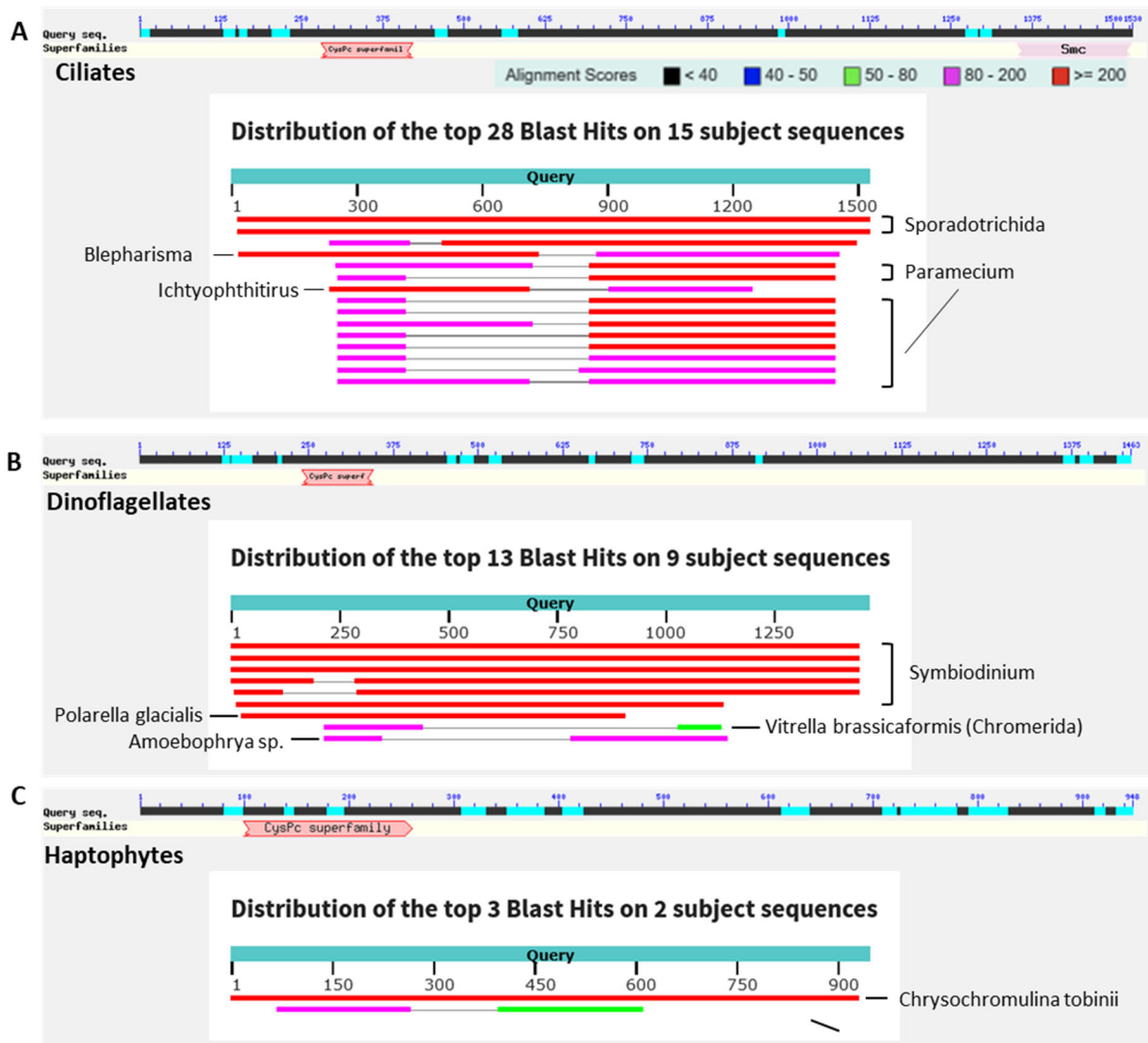
- Stephens, R. E. (1995). Ciliogenesis in sea urchin embryos--a subroutine in the program of development. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, 17(4), 331–340.
- Storz, J. F. (2018). *Hemoglobin*. Oxford University Press.
- Storz, J. F., Opazo, J. C., & Hoffmann, F. G. (2013). Gene duplication, genome duplication, and the functional diversification of vertebrate globins. *Molecular Phylogenetics and Evolution*, 66(2), 469–478.
- Su, X., Wu, M., Yao, G., El-Jouni, W., Luo, C., Tabari, A., & Zhou, J. (2015). Regulation of polycystin-1 ciliary trafficking by motifs at its C-terminus and polycystin-2 but not by cleavage at the GPS site. *Journal of Cell Science*, 128(22), 4063–4073.
- Thi Thanh Hai, N., Thuy, L. T. T., Shiota, A., Kadono, C., Daikoku, A., Hoang, D. V., Dat, N. Q., Sato-Matsubara, M., Yoshizato, K., & Kawada, N. (2018). Selective overexpression of cytoglobin in stellate cells attenuates thioacetamide-induced liver fibrosis in mice. *Scientific Reports 2018 8:1*, 8(1), 1–15.
- Thi Thanh Thuy, L., Morita, T., Yoshida, K., Wakasa, K., Iizuka, M., Ogawa, T., Mori, M., Sekiya, Y., Momen, S., Motoyama, H., Ikeda, K., Yoshizato, K., & Kawada, N. (2011). Promotion of liver and lung tumorigenesis in DEN-treated cytoglobin-deficient mice. *The American Journal of Pathology*, 179(2), 1050–1060.
- Thomas, N. A., Bardy, S. L., & Jarrell, K. F. (2001). The archaeal flagellum: a different kind of prokaryotic motility structure. *FEMS Microbiology Reviews*, 25(2), 147–174.
- Thuy, L. T. T., Hai, H., & Kawada, N. (2020). Role of cytoglobin, a novel radical scavenger, in stellate cell activation and hepatic fibrosis. *Clinical and Molecular Hepatology*, 26(3), 280.
- Thuy, L. T. T., Van Thuy, T. T., Matsumoto, Y., Hai, H., Ikura, Y., Yoshizato, K., & Kawada, N. (2016). Absence of cytoglobin promotes multiple organ abnormalities in aged mice. *Scientific Reports*, 6.
- Tiso, M., Tejero, J., Basu, S., Azarov, I., Wang, X., Simplaceanu, V., Frizzell, S., Jayaraman, T., Geary, L., Shapiro, C., Ho, C., Shiva, S., Kim-Shapiro, D. B., & Gladwin, M. T. (2011). Human Neuroglobin Functions as a Redox-regulated Nitrite Reductase. *The Journal of Biological Chemistry*, 286(20), 18277.

- Tonami, K., Kurihara, Y., Aburatani, H., Uchijima, Y., Asano, T., & Kurihara, H. (2007). Calpain 6 is involved in microtubule stabilization and cytoskeletal organization. *Molecular and Cellular Biology*, 27(7), 2548–2561.
- Trachootham, D., Alexandre, J., & Huang, P. (2009). Targeting cancer cells by ROS-mediated mechanisms: a radical therapeutic approach? *Nature Reviews Drug Discovery* 2009 8:7, 8(7), 579–591.
- Trent, J. T., & Hargrove, M. S. (2002). A ubiquitously expressed human hexacoordinate hemoglobin. *Journal of Biological Chemistry*, 277(22), 19538–19545.
- Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigartyo, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., ... Ponten, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220), 1260419–1260419.
- Van Acker, Z. P., Luyckx, E., & Dewilde, S. (2018). Neuroglobin Expression in the Brain: a Story of Tissue Homeostasis Preservation. *Molecular Neurobiology* 2018 56:3, 56(3), 2101–2122.
- Vinogradov, S. N., Hoogewijs, D., Bailly, X., Mizuguchi, K., Dewilde, S., Moens, L., & Vanfleteren, J. R. (2007). A model of globin evolution. *Gene*, 398(1–2), 132–142.
- Vinogradov, S. N., Tinajero-Trejo, M., Poole, R. K., & Hoogewijs, D. (2013). Bacterial and archaeal globins - A revised perspective. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1834(9), 1789–1800.
- Wang, M., Liu, X., Chang, G., Chen, Y., An, G., Yan, L., Gao, S., Xu, Y., Cui, Y., Dong, J., Chen, Y., Fan, X., Hu, Y., Song, K., Zhu, X., Gao, Y., Yao, Z., Bian, S., Hou, Y., ... Qiao, J. (2018). Single-Cell RNA Sequencing Analysis Reveals Sequential Cell Fate Transition during Human Spermatogenesis. *Cell Stem Cell*, 23(4), 599-614.e4.
- Webster, J., & Weber, R. (2007). Introduction to Fungi. In *Introduction to Fungi*. Cambridge University Press.
- Wright, T. J., & Davis, R. W. (2015). Myoglobin oxygen affinity in aquatic and terrestrial birds and mammals. *Journal of Experimental Biology*, 218(14), 2180–2189.

- Xu, H. W., Huang, Y. J., Xie, Z. Y., Lin, L., Guo, Y. C., Zhuang, Z. R., Lin, X. P., Zhou, W., Li, M., Huang, H. H., Wei, X. L., Man, K., & Zhang, G. J. (2013). The expression of cytoglobin as a prognostic factor in gliomas: a retrospective analysis of 88 patients. *BMC Cancer*, 13, 247.
- Yassin, M., Kissow, H., Vainer, B., Joseph, P. D., Hay-Schmidt, A., Olsen, J., & Pedersen, A. E. (2018). Cytoglobin affects tumorigenesis and the expression of ulcerative colitis-associated genes under chemically induced colitis in mice. *Scientific Reports* 2018 8:1, 8(1), 1–16.
- Zhang, B., Liu, Y., Li, Y., Zhe, X., Zhang, S., & Zhang, L. (2018). Neuroglobin promotes the proliferation and suppresses the apoptosis of glioma cells by activating the PI3K/AKT pathway. *Molecular Medicine Reports*, 17(2), 2757–2763.
- Zhang, J., Lan, S. J., Liu, Q. R., Liu, J. M., & Chen, X. Q. (2013). Neuroglobin, a Novel Intracellular Hexa-Coordinated Globin, Functions as a Tumor Suppressor in Hepatocellular Carcinoma via Raf/MAPK/Erk. *Molecular Pharmacology*, 83(5), 1109–1119.

7 Appendix

7.1 Supplementary data



7.2 Index of electronic supplement

Electronic version of this thesis

Appendix of Chapter 2.1

- Digital version of the publication
- Supplemental figures and tables
- Supplementals files 1-2

Appendix of Chapter 2.2

- Digital version of the publication
- Digital version of the figures
- Supplementary files 1-2

Appendix of Chapter 2.3

- Supplementary files 1-3

Appendix of Chapter 2.4

- Supplementary files 1-5

Appendix of Discussion

- Blast results for basal metazoan Adgb sequences

7.3 Acknowledgements

7.4 Eidesstattliche Erklärung

Hiermit versichere ich, dass ich diese Dissertation selbständig angefertigt und alle verwendeten Hilfsmittel in der Arbeit angegeben habe. Ich habe die Dissertation, oder Teile davon, an keiner anderen Fakultät, bzw. einem anderen Fachbereich eingereicht. Ich habe keinen anderen Promotionsversuch unternommen.

Mainz, September 2022

Carina Osterhof

CARINA OSTERHOF

Göttelmannstraße 12, 55130 Mainz-Weisenau · 0163/7954976

caosterh@uni-mainz.de

EDUCATIONAL PROFILE

01/2016 - PRESENT

PhD Student

Supervisor: Prof. Dr. Thomas Hankeln
AG Molecular Genetics and Genome Analysis
IOME JGU Mainz

04/2014 – 11/2015

Master of Science Biology (Final Grade: 1,2)

Master thesis: *Emission sources of airborne Bacteria*
Supervisor: Dr. Viviane Després (IMP, JGU Mainz)

04/2011 – 05/2014

Bachelor of Science Biology (Final Grade: 1,5)

Bachelor thesis: *Expressionsanalyse von Cytoglobin und Neuroglobin in verschiedenen Vertebraten*
Supervisor: Prof. Dr. Thomas Hankeln (IOME, JGU Mainz)

03/2011

Abitur (Final Grade: 1,3)

Konrad-Adenauer-Gymnasium Westerburg
VBio Award for best Biology exam

SKILLS AND EXPERTISE

Bioinformatics

- (Single cell) RNA-Sequencing
- Differential gene expression analysis, GO-Term and Pathway analysis
- Microbiome analysis

Molecular Biology

- | | |
|---------------------------|------------------------|
| - RNA/DNA Isolation | - Western Blot |
| - Molecular cloning | - Immunohistochemistry |
| - Cell culture techniques | - Histology |
| - PCR/qRT-PCR | - Microscopy |

ADDITIONAL TRAINING

- Advanced training 'Genetic engineering safety regulations' (JGU Mainz 2016)
- Practical Course: Histology Course (IMB Mainz 2016)
- FELASA-A (former B) certificate for animal handling (TARC Mainz 2017)

SCIENTIFIC CONTRIBUTIONS

PUBLICATIONS

dos Santos Guilherme, M., Todorov, H., **Osterhof, C.**, Möllerke, A., Cub, K., Hankeln, T., Gerber, S., & Endres, K. (2020). Impact of Acute and Chronic Amyloid- β Peptide Exposure on Gut Microbial Commensals in the Mouse. *Frontiers in Microbiology*, 11. <https://doi.org/10.3389/fmicb.2020.01008>

Koay, T. W.*, **Osterhof, C.***, Orlando, I. M. C., Keppner, A., Andre, D., Yousefian, S., Alonso, M. S., Correia, M., Markworth, R., Schödel, J., Hankeln, T., & Hoogewijs, D. (2021). Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis. *Journal of Biological Chemistry*, 296. <https://doi.org/10.1016/j.jbc.2021.100291>

Keppner, A., Correia, M., Santambrogio, S., Koay, T. W., Maric, D., **Osterhof, C.**, Winter, D. V., Clerc, A., Stumpe, M., Chalmel, F., Dewilde, S., Odermatt, A., Kressler, D., Hankeln, T., Wenger, R. H., & Hoogewijs, D. (2021). Androglobin, a chimeric mammalian globin, is required for male fertility. *eLife* 11:e72374 <https://doi.org/10.7554/eLife.72374>

CONTRIBUTION TO CONFERENCES

Expression of Androglobin in the female reproductive tract (Poster, O2-BiP conference 2018, Barcelona)

Androglobin in the female reproductive tract: How to identify cell-type specific expression without wet lab experiments (Talk, IV. Life Science Meeting of Junior GBM 2019, Mainz)

Expression of Androglobin in the female reproductive tract (Poster, FEBS+LS² annual meeting 2020, Zurich)

TEACHING EXPERIENCE

2016 – 2021: Teaching assistant in:

Biostatistics and Bioinformatics (B.Sc. module)

Tasks: Conceptualization, organisation, and guidance during practical work

Genetics (B.Sc. module)

Tasks: Organisation, preparation, and guidance during practical work

Molecular and bioinformatical analysis of eukaryotic genes (B.Sc. advanced module)

Tasks: Organisation, preparation, and guidance during practical work

Genomic research and sequence analysis (M.Sc. advanced module)

Tasks: Lectures, organisation, revision of reports, conceptualisation, and guidance during practical work

For all courses: conceptualisation and realisation of examinations (digital and written)

2012 – 2016: Student teaching assistant in plant physiology

Tasks: oral examination of students; revision of reports; guidance in experimenting