Evolutionary History of *GLIS* Genes Illuminates Their Roles in Cell Reprograming and Ciliogenesis

Yuuri Yasuoka ,*,[1] Masahito Matsumoto,[1,2,3,4] Ken Yagi,[1] and Yasushi Okazaki[1,2]

[1]Laboratory for Comprehensive Genomic Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

[2]Diagnostics and Therapeutics of Intractable Diseases, Intractable Disease Research Center, Graduate School of Medicine, Juntendo University, Tokyo, Japan

[3]Advanced Diabetic Therapeutics, Department of Metabolic Endocrinology, Graduate School of Medicine, Juntendo University, Tokyo, Japan

[4]Department of Biofunction Research, Institute of Biomaterials and Bioengineering, Tokyo Medical and Dental University, Tokyo, Japan

***Corresponding author:** E-mail: yuuri.yasuoka@riken.jp.

Associate editor: Meredith Yeager

Abstract

The GLIS family transcription factors, GLIS1 and GLIS3, potentiate generation of induced pluripotent stem cells (iPSCs). In contrast, another GLIS family member, GLIS2, suppresses cell reprograming. To understand how these disparate roles arose, we examined evolutionary origins and genomic organization of *GLIS* genes. Comprehensive phylogenetic analysis shows that *GLIS1* and *GLIS3* originated during vertebrate whole genome duplication, whereas *GLIS2* is a sister group to the *GLIS1/3* and *GLI* families. This result is consistent with their opposing functions in cell reprograming. Glis1 evolved faster than Glis3, losing many protein-interacting motifs. This suggests that Glis1 acquired new functions under weakened evolutionary constraints. In fact, GLIS1 induces induced pluripotent stem cells more strongly. Transcriptomic data from various animal embryos demonstrate that *glis1* is maternally expressed in some tetrapods, whereas vertebrate *glis3* and invertebrate *glis1/3* genes are rarely expressed in oocytes, suggesting that vertebrate (or tetrapod) Glis1 acquired a new expression domain and function as a maternal factor. Furthermore, comparative genomic analysis reveals that *glis1/3* is part of a bilaterian-specific gene cluster, together with *rfx3*, *ndc1*, *hspb11*, and *lrrc42*. Because known functions of these genes are related to cilia formation and function, the last common ancestor of bilaterians may have acquired this cluster by shuffling gene order to establish more sophisticated epithelial tissues involving cilia. This evolutionary study highlights the significance of GLIS1/3 for cell reprograming, development, and diseases in ciliated organs such as lung, kidney, and pancreas.

Key words: **ortholog group, gene duplication, neofunctionalization, comparative transcriptomics, microsynteny, ciliogenic gene cluster.**

Article

Introduction

Among Krüppel-like zinc-finger transcription factors, GLI-similar transcription factors (GLIS) constitute a large family, together with GLI and ZIC (Hatayama and Aruga 2010; Kang et al. 2010; Scoville et al. 2017; Aruga and Hatayama 2018). GLI, GLIS, and ZIC share a DNA-binding domain consisting of five C2H2 zinc-finger domains, two of which, near the N-terminus, are characterized as a tandem pair of CWCH2 motifs (Hatayama and Aruga 2010; Aruga and Hatayama 2018). It has been proposed that GLI, GLIS, and ZIC originated from a common ancestral gene and that ZIC is an early branching gene group, relative to GLI and GLIS (Layden et al. 2010; Aruga and Hatayama 2018). However, phylogenetic relationships between GLI and GLIS have never been determined, possibly because zinc-finger domains from only a small number of taxa have been used for phylogenetic analysis (Kim et al. 2003; Kang et al. 2010; Layden et al. 2010; Aruga and Hatayama 2018).

In mammals, three *GLIS* genes (*GLIS1-3*) have been identified and their embryonic expression patterns in mice and adult organ expression levels in mice and/or humans have been determined (reviewed in Kang et al. [2010]). All three genes are most abundantly expressed in kidney, with moderate expression in various other organs: *GLIS1* is expressed in brain, thymus, adipose tissue, colon, testis, and placenta (Kim et al. 2002; Nakashima et al. 2002); *GLIS2* is in brain, lung, heart, esophagus, intestine, colon, thyroid, liver, and prostate (Zhang and Jetten 2001; Zhang et al. 2002); and *GLIS3* is in brain, lung, thymus, thyroid, liver, pancreas, spleen, testis, ovary, and uterus (Kim et al. 2003; Senee et al. 2006; Beak et al. 2008). *GLIS3* is also expressed in a number of human cancers, suggesting that elevated *GLIS3* expression leads to cancer progression (Kang et al. 2010). Remarkably, only *Glis1* is significantly expressed in early mouse embryos (oocyte to two-cell stage), possibly related to its proreprograming functions, discussed below (Maekawa et al. 2011).

Knock-out mouse studies have revealed that Glis2 is essential for normal renal functions (Attanasio et al. 2007; Kim et al. 2008), and that Glis3 is required for kidney development, pancreatic β-cell development, and spermatogenesis

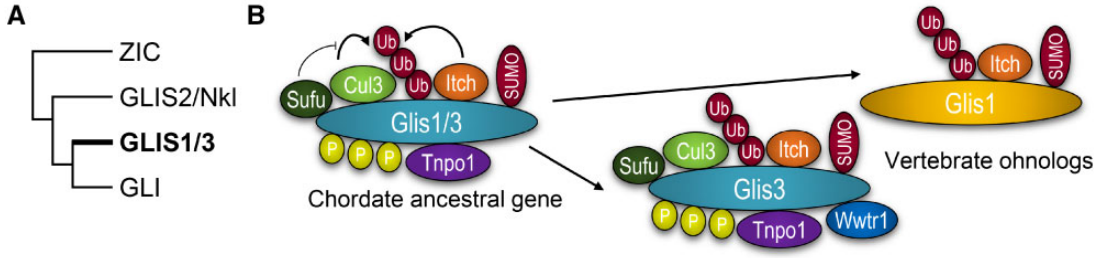

FIG. 1. *GLIS1* and *GLIS3* evolved as ohnologs in vertebrates. (*A*) A schematic phylogenetic tree of *GLI*/*GLIS*/*ZIC* genes supported by ORTHOSCOPE analysis and ML trees (see supplementary figs. S1–S4, Supplementary Material online, for more details). (*B*) Presumed evolution of protein–protein interaction motifs of GLIS1/3 after WGD in vertebrates. GLIS3 retains all conservative motifs in chordates, whereas GLIS1 lost many of them. Ubiquitination/SUMOylation motifs and an ITCH-binding motif are strongly conserved among GLIS1/3 in vertebrates. See supplementary figures S5 and S6, Supplementary Material online, for phylogenetic analysis and sequence alignment, respectively.

(Kang, Beak, et al. 2009; Kang, Kim, et al. 2009; Kang et al. 2016). In studies of kidney disease, Glis2 and Glis3 were localized to primary cilia (Attanasio et al. 2007; Kang, Beak, et al. 2009). GLIS2 dysfunction was responsible for nephronophthisis (NPHP), and loss of GLIS3-function leads to neonatal diabetes and hypothyroidism (NDH), polycystic kidneys, and other abnormalities (Jetten 2018). Notably, in generation of induced pluripotent stem cells (iPSCs) from human and mouse fibroblasts by so-called Yamanaka factors (Oct4, Sox2, Klf4, and c-Myc), replacement of c-Myc with GLIS1 decreases tumorigenicity (Maekawa et al. 2011).

In other vertebrates, Glis2 was first identified as a neuronal Krüppel-like protein (Nkl) in *Xenopus*, and misexpression of Glis2 induced extra primary neurons (Zhang and Jetten 2001). In zebrafish, Glis2 was described as NPHP7 and loss of function analysis showed that Glis2 is required for cilium motility (Kim et al. 2013). Loss of function experiments involving Glis3 led to a significant decrease in β-cell mass in zebrafish (O'Hare et al. 2016). Glis3 deficiency also resulted in a medaka mutant with shortened renal cilia and caused polycystic kidney disease (Hashimoto et al. 2009). These studies imply conserved roles of Glis2 and Glis3 for primary cilium function and kidney development, and Glis3 for pancreatic development in vertebrates. However, little is known about Glis1 functions in other vertebrates.

Results and Discussion

Phylogenetic Relationships of *GLIS* Genes Are Consistent with Differences in Gene Functions

To infer *GLIS* gene ancestry, we first performed comprehensive phylogenetic analysis using a species tree-based ortholog group identification tool, ORTHOSCOPE (Inoue and Satoh 2018). The result demonstrates that *GLIS1* and *GLIS3* are "ohnologs" derived from a single ancestral gene (*GLIS1/3*) via two-rounds of whole genome duplication (WGD) in vertebrates (supplementary figs. S1 and S2, Supplementary Material online). In addition, the orthologous gene group *GLIS1/3* is a sister group to the *GLI* subfamily. On the other hand, *GLIS2* and its invertebrate orthologs form an outgroup relative to *GLI* and *GLIS1/3* (fig. 1*A*). These results were further validated by maximum likelihood (ML) trees constructed using the same sequence set as used with ORTHOSCOPE (supplementary figs. S3 and S4, Supplementary Material online). Therefore, these data clarify relationships between *GLI*, *GLIS1*, *GLIS2*, and *GLIS3* with higher reliability than previous studies (Materna et al. 2006; Shimeld 2008; Hatayama and Aruga 2010; Layden et al. 2010).

The deep evolutionary origins of Glis1/3 and Glis2 are reconciled with their structural and functional differences. First, a nuclear localization signal is located at the fourth zinc-finger domain (ZF4) of Glis3, but at ZF3 of Glis2 (Beak et al. 2008; Vasanth et al. 2011; Hatayama and Aruga 2012). Second, Glis3 contains a ciliary localization signal in the N-terminal region that is conserved among Gli transcription factors and presumably binds to Transportin1 (TNPO1) (Han et al. 2017; Jetten 2018). However, the ciliary localization signal motif is not present in Glis2, although Glis2 is reportedly localized in the primary cilium (Attanasio et al. 2007; Jetten 2018). Third, Glis1 and Glis3 contain transactivation domains in the C-terminal region, and self-repressive domains in the N-terminal region (Kim et al. 2002, 2003). In contrast, Glis2 contains transactivation domains in the N-terminal region, and self-repressive domains in the C-terminal region (Zhang et al. 2002). Fourth, both GLIS1 and GLIS3 have comparable activity in reprograming of human adipose-derived stromal cells, together with Yamanaka factors, whereas GLIS2 suppresses reprograming (Lee et al. 2017).

The phylogenetic tree of Glis1/3 in chordates further demonstrates that Glis1 evolved faster than Glis3 after WGD in vertebrates (supplementary fig. S5, Supplementary Material online). This implies that Glis1 experienced additional evolutionary events such as acquisition of new functions, loss of ancestral functions, and adaptive evolution, whereas Glis3 retained ancestral form and functions. Actually, GLIS1 can replace Klf4 as a transcription factor for iPSC generation, but other GLI/GLIS/ZIC factors, GLI4, GLIS2, GLIS3, and ZIC4, cannot (Maekawa et al. 2011). Hereafter, we focus on similarities and differences between the ohnologs, Glis1 and Glis3.

Glis1 Lacks Several Conserved Motifs

The Cullin3 complex binds Glis3 for degradation via polyubiquitination, whereas Supressor of Fused (Sufu) inhibits its degradation by binding to Glis3 via a YGH motif (ZeRuth et al. 2011). Similarly, an E3 ubiquitin ligase, Itch, reportedly binds to Glis3 via a PPPY motif for degradation

(ZeRuth et al. 2015). For transactivation activity, a Hippo signal pathway regulator Wwtr1/TAZ binds to a P/LPXY motif in the C-terminal region of Glis3 but does not bind to Glis1 or Glis2 (Kang, Beak, et al. 2009). Glis3 also interacts with CBP/p300 through its C-terminal transactivation domain (ZeRuth et al. 2013). Interacting partner proteins of Glis1 and Glis2 have been little studied compared with those of Glis3.

Because of the deep origins of Glis1/3 and Glis2, sequence homology between Glis1 and Glis3 helps us to understand evolution of vertebrate ohnologs. As a result of its faster evolutionary rate (supplementary fig. S5, Supplementary Material online), Glis1 lost several protein motifs that are conserved in Glis3 and invertebrate Glis1/3 (supplementary fig. S6, Supplementary Material online). For example, Glis1 lacks a binding motif for Sufu, which is conserved in Glis3 and Gli (ZeRuth et al. 2011). Glis1 also lacks all putative phosphorylation sites identified in Glis3 (ZeRuth et al. 2015). On the other hand, a binding motif for Itch is widely conserved among chordate Glis1/3 genes. In addition, ubiquitination/sumoylation motifs are also conserved, suggesting that Itch is heavily involved in degradation of Glis1 and Glis3. Because invertebrate Glis1/3 does not have a Wwtr1-interacting motif, Glis3 may have acquired that motif for transactivation. Thus, Glis1 lost many ancestral features and may have changed its functions, whereas Glis3 is highly conserved and probably retains ancestral functions (fig. 1*B*). It is worth investigating whether Glis1 acquired new binding partners for cell reprograming.

Faster Evolving *glis1/3* Genes Are Prone to Elimination from Genomes after Extra WGD

We next performed syntenic analysis to examine genomic organization of *glis1* and *glis3* (fig. 2*A* and *B*). When the human genome is compared with sarcopterygian and shark genomes, both loci are well conserved, whereas in actinopterygians, both loci are highly rearranged, with a few exceptions such as *ndc1*, *lrrc42*, and *rfx3*. Those will be further discussed below.

Among teleosts, medaka and zebrafish retain two *glis1* genes (*glis1a* and *glis1b*) that originated from teleost-specific WGD, whereas fugu lost *glis1a* (fig. 2*A*). Because Glis1a sequences are more derived than Glis1b (supplementary fig. S5, Supplementary Material online), Glis1a must be less constrained and has been more easily eliminated during teleost evolution. By contrast, extant teleosts only retain a single copy of *glis3* (fig. 2*B*), possibly due to rapid gene loss in an early stage of teleost evolution (Inoue et al. 2015). Because teleost *glis3* genes have longer branch lengths in phylogenetic trees than other vertebrate *glis3* genes (supplementary fig. S5, Supplementary Material online), teleost Glis3 possibly evolved under weakened functional constraints.

Another case of additional WGD in vertebrates is the African clawed frog, *Xenopus laevis*, which has an allotetraploid genome composed of two subgenomes, denoted L and S (Session et al. 2016). Surveying *glis1/3* genes in *X. laevis*, we found that *glis1* and surrounding genes have been eliminated from the S subgenome (fig. 2*C*), whereas the *glis3* locus is conserved, except for copy number variations of crystallin gamma (*cryg*) genes (fig. 2*D*). As with Glis1a and Glis3 in teleosts, *Xenopus* Glis1 is more divergent among vertebrate Glis1 (supplementary fig. S5, Supplementary Material online). Taken together, loss of faster evolving genes in vertebrate genomes may reflect differences of evolutionary constraints on Glis1/3 functions.

Expression Profiles Suggest Neofunctionalization of Glis1 in Vertebrate Oocytes

In mice, *Glis1* is enriched in unfertilized eggs and one-cell stage embryos (Maekawa et al. 2011). Transcriptomic data of mouse early embryos (Tang et al. 2011; Xue et al. 2013) further demonstrate that *Glis1*, but not *Glis3*, is maternally expressed (fig. 3*A* and supplementary fig. S7*B*, Supplementary Material online). However, in humans, *GLIS1* and *GLIS3* are not expressed in oocytes (supplementary fig. S7*A*, Supplementary Material online), suggesting gain or loss of maternal expression of *Glis1* in mice or humans, respectively. To examine these possibilities, we next examined transcriptomic data from bovine preimplantation embryos (Jiang et al. 2014). The data showed that *Glis1*, but not *Glis3*, is expressed in bovine oocytes (supplementary fig. S7*C*, Supplementary Material online), supporting the possibility that the mammalian ancestor possessed *Glis1* expression in oocytes, but that humans lost it.

To infer the origin of maternal expression of *glis1*, we further surveyed transcriptomic data of other vertebrates. Interestingly, transcriptomic data of chicken early embryos (Hwang et al. 2018) demonstrated that both *glis1* and *glis3* are expressed in oocytes (fig. 3*B*). High temporal-resolution transcriptomic data from *Xenopus tropicalis* embryos (Owens et al. 2016) showed that *glis1*, but not *glis3*, is expressed as a maternal factor (fig. 3*C*). In teleosts, both medaka and zebrafish transcriptomic data during developmental stages (Ichikawa et al. 2017; White et al. 2017) indicate that *glis1/3* genes are not expressed in oocytes, except for medaka *glis3*, which exhibits fairly low-level expression (fig. 3*D* and supplementary fig. S7*D*, Supplementary Material online). These results indicate that *glis1* may have been expressed maternally, at least in the tetrapod ancestor, and that lineage-specific gain and loss of maternal expression occurred for *glis1/3* genes in several lineages. To determine the origin of maternal expression of *glis1* in vertebrates, we need to examine more expression profiles, especially from basal vertebrates such as chondrichthyans and cyclostomes. Because the elephant shark genome retains more conserved synteny around *glis1* and *glis3* with tetrapods than those of actinopterygians (fig. 2*A* and *B*), it is possible that chondrichthyans retain maternal expression of *glis1* as an ancestral feature of gnathostomes.

Given the likely tetrapod origin of *glis1* maternal expression, it is interesting that *dmrtb1* (also called *dmrt6*) adjoins *glis1* in tetrapod genomes (fig. 3*A*). DMRTB1 was also screened as a candidate proreprograming transcription factor, although DMRTB1 could not be replaced with c-MYC as GLIS1 could (Maekawa et al. 2011). In addition, transcriptomic data showed that *dmrtb1* is also maternally expressed

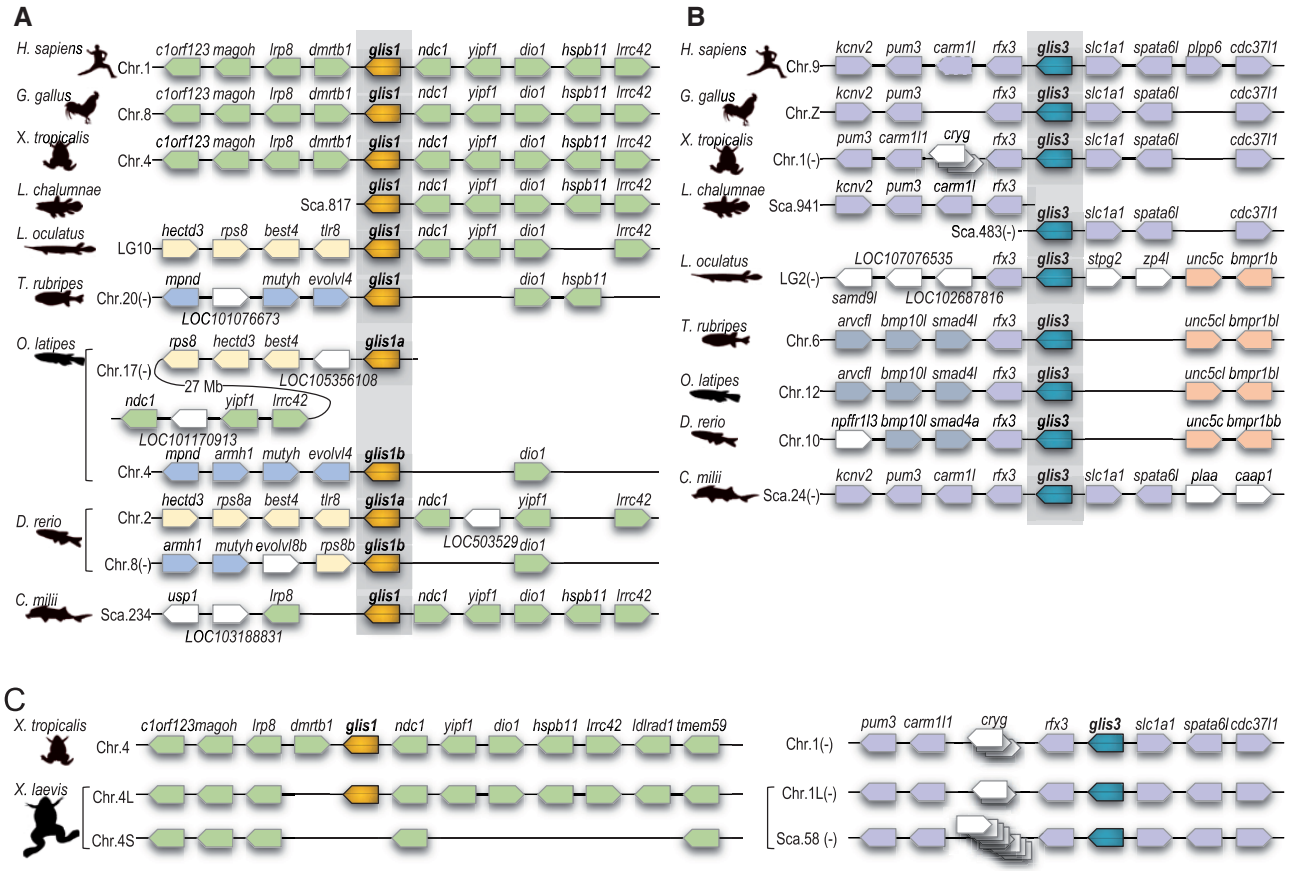

Fig. 2. Synteny around *glis1* and *glis3* is highly conserved in vertebrates. (*A*) Conserved synteny around *glis1*. Green, yellow, and blue boxes indicate conserved syntenic protein-coding genes among vertebrates, actinopterygians, and teleosts, respectively. (*B*) Conserved synteny around *glis3*. Purple, orange, and gray boxes indicate conserved syntenic protein-coding genes among vertebrates, actinopterygians, and teleosts, respectively. Human *CRAM1L1* became a pseudogene, as indicated by the dashed line. (*C*) Genomic organization around *glis1* in *Xenopus*. *glis1* and surrounding genes (*yipf1*, *dio1*, *hspb11*, *lrrc42*, and *idlrad1*) have been eliminated from the S subgenome in *X. laevis*, and *dmrtb1* was eliminated from both subgenomes. (*D*) Synteny around *glis3* is widely conserved in *X. laevis* subgenomes, but the number of tandemly duplicated *cryg* genes varies between *X. tropicalis* (3), the *X. laevis* L subgenome (2), and the *X. laevis* S subgenome (8).

in human, mouse, and frog oocytes, but not in bovine or chicken oocytes (supplementary fig. S8, Supplementary Material online). Thus, it is worth examining the possibility that GLIS1 and DMRTB1 work collaboratively as maternal factors. It should be noted that *dmrtb1* was eliminated from both subgenomes of *X. laevis* (fig. 2C), as previously described (Watanabe et al. 2017), although it is maternally expressed in *X. tropicalis* (supplementary fig. S8, Supplementary Material online). This fact implies that maternal gene regulatory networks vary considerably between *X. laevis* and *X. tropicalis*.

To reveal the ancestral expression pattern of *glis1/3* before WGD, we further examined transcriptomic data of bilaterian embryos. Recently published transcriptomic data of the cephalochordate, *Branchiostoma lanceolatum* (Marletaz et al. 2018), showed that *glis1/3* is not expressed in eggs and early stage embryos (fig. 3E) just as *glis3* is not in many vertebrates (fig. 3A, C, and D and supplementary fig. S7A–D, Supplementary Material online). Among nonchordate deuterostomes, *glis1/3* is maternally expressed in the sea urchin *Strongylocentrotus purpuratus*, as shown by QPCR (quantitative polymerase chain reaction) (Materna et al. 2006) and RNA-seq (Tu et al. 2012) (fig. 3F). However, transcriptomic data of another sea urchin, *Paracentrotus lividus* (Gildor et al. 2016), showed that *glis1/3* is not expressed in eggs (supplementary fig. S7E, Supplementary Material online), indicating that *glis1/3* is not necessarily a maternal factor in deuterostomes. In protostomes, *glis1/3* is rarely expressed in eggs and early embryos, as shown in transcriptomic data from brachiopods (Luo et al. 2015) and scallops (Wang et al. 2017) (fig. 3G and supplementary fig. S7F, Supplementary Material online). These data suggest that, after WGD, *glis3* retained ancestral expression patterns, whereas *glis1* may have acquired new expression domains and functions in vertebrate (or tetrapod) oocytes.

glis1/3 Retains Evolutionarily Conserved Microsynteny in Bilaterians

To further examine conserved microsynteny around the *glis1/3* locus in animals, we surveyed invertebrate genomic data. We found that *glis1/3*, *rfx3*, *ndc1*, *hspb11*, and *lrrc42* are clustered in 100–600-kb regions of most bilaterian genomes (fig. 4A and supplementary table S1, Supplementary Material online). *Drosophila* lost two genes and the remaining three

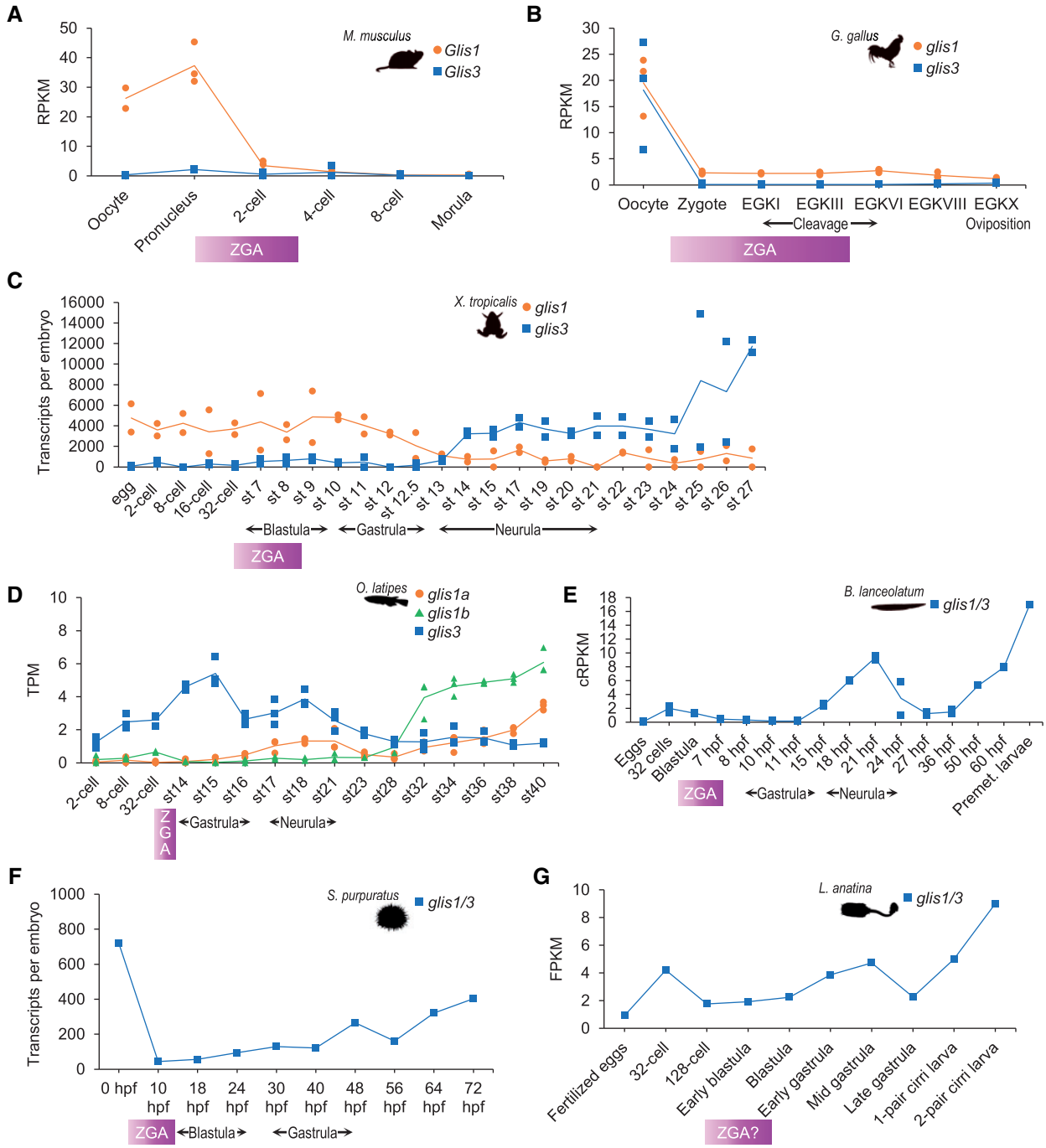

Fig. 3. *glis1* is maternally expressed in tetrapods, but vertebrate *glis3* and invertebrate *glis1/3* are rarely expressed as maternal factors. Expression levels of *glis1/3* genes in various animal embryos are shown in graphs. Individual data from biological replicates are indicated in orange circles for *glis1* (or *glis1a* in teleosts), blue squares for *glis3* (or *glis1/3* in invertebrates), and green triangles for *glis1b* in teleosts. Lines represent the average of biological replicates. ZGA represents the period of zygotic gene activation, proposed for each species or closely related species (Tadros and Lipshitz 2009; Yang et al. 2016; Jukam et al. 2017). (*A*) Early mouse embryos (Xue et al. 2013). *Glis1* but not *Glis3* is expressed before ZGA and immediately degraded after ZGA in mouse. RPKM, reads per kilobase of exon per million mapped reads. See supplementary figure S7*A*–*C*, Supplementary Material online, for more data sets from mammals (human, mouse, and bovine). (*B*) Early chicken embryos (Hwang et al. 2018). Both *glis1* and *glis3* are expressed in chicken oocytes but greatly reduced in zygotes afterward. (*C*) Frog (*X. tropicalis*) embryos (Owens et al. 2016). *glis1* is expressed maternally at levels of ~10,000 transcripts per egg (0hpf) or embryo (others), whereas *glis3* is only expressed zygotically in *Xenopus*. (*D*) Medaka embryos (Ichikawa et al. 2017). *glis3* is weakly expressed before ZGA but *glis1a* and *glis1b* are not in medaka. TPM, transcripts per million. See supplementary figure S7*D*, Supplementary Material online, for zebrafish data. (*E*) Amphioxus embryos (Marletaz et al. 2018). *glis1/3* is hardly expressed before 18 hpf (neurula) in amphioxus. cRPKM, corrected (per mappability) reads per kb of mappable positions and million reads. (*F*) Sea urchin (*S. purpuratus*) embryos (Tu et al. 2012). *glis1/3* is maternally expressed but is greatly reduced at ZGA in *S. purpuratus*. See supplementary figure S7*E*, Supplementary Material online, for data from another sea urchin (*P. lividus*). (*G*) Brachiopod embryos (Luo et al. 2015). *glis1/3* is rarely

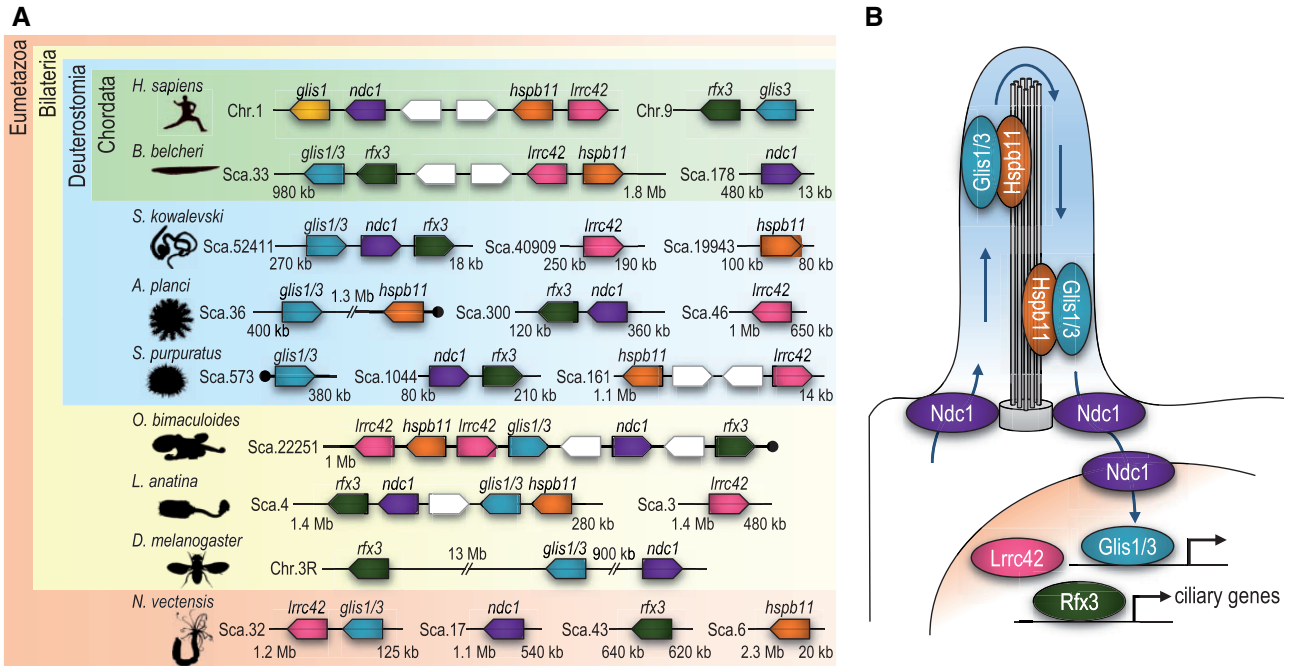

Fig. 4. Bilaterian-specific gene cluster for ciliogenesis. (*A*) Conserved synteny of ciliogenic genes (*glis1/3*, *rfx3*, *ndc1*, *hspb11*, and *lrrc42*) in bilaterians. Octopus *glis1/3* is separated into two gene models (see supplementary table S1, Supplementary Material online), but a single gene is shown in this figure. In the brachiopod genome (*Lingula anatina*), *ndc1* is separated into two gene models (see supplementary table S1, Supplementary Material online) and a gene model is identified in the opposite strand of *ndc1*. For simplification, only a single *ndc1* gene is indicated in this figure. In nonchromosomal level genome assemblies (nonhuman and nonfly), sizes of remaining regions in the scaffold are indicated. Black circles mean that no gene models are identified in the remaining region, or in other words, that the gene model is located close to the end of the scaffold. These data demonstrate that the ciliogenic gene cluster is highly conserved in humans, amphioxus, octopuses, and brachiopods. See supplementary figure S9, Supplementary Material online, for more detailed comparison of gene orders around the cluster. (*B*) Presumed functions of ciliogenic cluster genes. Glis1/3 is localized in both cilia and nuclei and may be trafficked via Ndc1 and Hspb11. Rfx3 regulates ciliogenic gene expression. Lrrc42 may function as a transcriptional regulator, together with Glis1/3 and Rfx3. See supplementary figure S10, Supplementary Material online, for *glis1/3* expression profiles in ciliated adult tissues of invertebrates.

genes are distantly located on the same chromosome, whereas interestingly, octopus retains the intact gene cluster. We also examined synteny conservation of genes neighboring the cluster in humans, amphioxus, octopuses, and brachiopods, genomes of which contain clusters that are almost intact (supplementary fig. S9, Supplementary Material online). Results showed that synteny of genes other than these five genes is not conserved among the four genomes, emphasizing remarkable conservation of these five genes as a cluster.

Among nonbilaterians, only the sea anemone (*Nematostella vectensis*) possesses *glis1/3* and *lrrc42* in the same vicinity, but others do not have a putative gene cluster. In the choanoflagellate, *Monosiga brevicollis*, a unicellular organism closely related to animals, genes other than *rfx3* are missing in its genome. These facts suggest that the common ancestor of bilaterians acquired this cluster by shuffling gene order. A characteristic feature of this highly conserved gene cluster is that all five genes belong to different gene families, in contrast to clusters of duplicated copies of the same gene family, such as Hox, ParaHox, and Wnt gene clusters (Takeuchi et al. 2016).

What then is the role of this cluster? Strikingly, the known functions of these genes are related to cilia (fig. 4*B*). Glis3 is localized in primary cilium and is associated with cystic renal diseases (Kang, Beak, et al. 2009). In amphioxus, *glis1/3* is highly expressed in gill bars, which are densely populated with ciliated cells, compared with other tissues (supplementary fig. S10*A*, Supplementary Material online). In brachiopods, *glis1/3* is enriched in lophophores, which contain ciliated tentacles (supplementary fig. S10*B*, Supplementary Material online). Among scallop adult organs, *glis1/3* is strongly expressed in the ciliated gill (supplementary fig. S10*C*, Supplementary Material online). Even in ctenophores, a basal metazoan lineage, *glis1/3* is expressed in ciliated cells (Layden et al. 2010), implying that Glis1/3 had an ancient role in ciliogenesis. Remarkably, *glis1/3* is also expressed at

Fig. 3. Continued
expressed during early embryogenesis in brachiopods. The period of ZGA has not been analyzed deeply in brachiopods, but we suppose that ZGA occurs around early blastula because expression of some developmental regulatory genes such as *bmp2/4*, *chordin*, and *brachyury* initiates at the early blastula stage. FPKM, fragments per kilobase of exon per million mapped fragments. See supplementary figure S7*F*, Supplementary Material online, for scallop embryos.

moderate levels in guts of amphioxus and brachiopods, and in digestive glands of brachiopods and scallops. These expression data imply that functions of Glis3 in pancreatic β-cell differentiation (Kang, Beak, et al. 2009; Kang, Kim, et al. 2009; O'Hare et al. 2016) originated the digestive system of the bilaterian ancestor.

Rfx3 positively regulates ciliary genes in most animals, and perhaps in choanoflagellates (Piasecki et al. 2010). Importantly, Rfx3 also regulates pancreatic β-cell differentiation (Ait-Lounis et al. 2010), suggesting cooperative functions with Glis3 in islets. Ndc1 is a component of the nuclear pore complex and also of the ciliary pore complex, which mediate protein transport to nuclei and cilia, respectively (Mansfeld et al. 2006; Ounjai et al. 2013). Hspb11 is an ortholog of intraflagellar transport 25 (Ift25) that participates in transport of Hedgehog signaling molecules, including Gli in primary cilium (Keady et al. 2012). Lrrc42 has been reported as a nuclear protein expressed in lung cancer (Fujitomo et al. 2014), suggesting that Lrrc42 interacts with transcription factors that may include Glis1/3 and Rfx3, to regulate ciliogenesis. Taken together, this "ciliogenic gene cluster" may serve to establish ciliated tissues in organs such as gill, gastrointestinal epithelium, lung, kidney, and pancreas, since the origin of bilaterians. In other words, formation of this cluster with other ciliogenic genes further suggests that ciliogenesis was the original function of Glis1/3, and that Glis1 has been coopted for cell reprograming in vertebrates.

Conclusions and Perspectives

In this study, we first clarified relationships between *GLI* and *GLIS* genes by comprehensive phylogenetic analysis. The similar gene names are confusing, but the first emergence of *GLIS2* by duplication of the ancestral *GLI/GLIS/ZIC* gene in metazoans greatly predates the appearance of *GLIS1* and *GLIS3* by WGD in vertebrates. Amino acid sequences of Glis1 and Glis3 were compared to identify conserved and diversified protein–protein interaction motifs. Surveys of transcriptomic data emphasized that maternal expression of *glis1* is characteristic of tetrapods. Glis1 appears to have been released from evolutionary constraints for conventional roles and has acquired new functions in oocytes. Together with proreprograming activity of *GLIS1*, we hypothesize that Glis1 was neofunctionalized for cell reprograming in vertebrates (or tetrapods). The cell reprograming activity of Glis1/3 from various animals should be examined using iPS or other reprograming assays. Then, Glis1-specific transcriptional machinery for cell reprograming should be determined.

Comparative genomic analysis revealed a highly conserved gene cluster containing *glis1/3* and other ciliogenic genes. Transcriptomic data also support ancestral roles of Glis1/3 in ciliated tissues. The next question is how these clustered genes are regulated for ciliogenesis. To answer this question, we surveyed previously identified, conserved noncoding sequences for the human genome using UCNEbase (Dimitrieva and Bucher 2013). We found two candidate cis-regulatory modules around the cluster (UCNE34150 and UCNE3883), but these are not conserved among bilaterians. To identify "the cluster controlling region," more comprehensive analysis for noncoding sequences should be performed.

This study highlights the importance of carefully considering orthologous relationships between homologs without preconceptions stemming from classical gene names, in order to better understand and predict gene functions. Expression profiles and comparative genomics provide us with many clues to unravel how genes evolved. The evolutionary history of *GLIS* genes illuminates potential functions of *GLIS1/3* genes for cell reprograming and ciliogenesis. Taken together with previous studies on GLIS1 for iPSC technologies and those on GLIS3 for development and disease in kidney and pancreas, our study will facilitate applications of GLIS1/3 to stem cell biology and medical sciences.

Materials and Methods

Phylogenetic Analysis

To identify ortholog groups of *GLIS* genes, protein-coding DNA sequences of human GLIS1, GLIS2, GLIS3, GLI2, and ZIC1 were submitted as queries to ORTHOSCOPE, a species tree-based ortholog identification tool (Inoue and Satoh 2018), with the following settings: analysis group, vertebrata; *E*-value threshold for reported sequences, $1e^{-5}$; number of hits to report per genome, 3; aligned site rate threshold within unambiguously aligned sites, 0; data set, DNA (Exclude 3rd); rearrangement BS (bootstrap) value threshold, 60%. To produce NJ and ML trees of Glis1/3, amino acid sequences of Glis1/3 were aligned with MAFFT (v7.221) (Katoh et al. 2002) using the –auto strategy. Unaligned regions were trimmed with TrimAl (v1.2rev59) (Capella-Gutierrez et al. 2009) using the –gappyout option. To generate nucleotide alignments, corresponding cDNA sequences were forced onto the amino acid alignment using PAL2NAL (Suyama et al. 2006). The maximum likelihood method with PROTGAMMAAUTO (amino acid sequences) or GTRGAMMA (nucleotide sequences) was used to construct phylogenetic trees with RAxML (v8.2.0) (Stamatakis 2014). For the nucleotide tree, we used codon partitions.

Synteny Analysis

Genomic synteny of *GLIS* genes in vertebrates was analyzed using genome assemblies of *Homo sapiens*, GRCh38.p12 (human), *Gallus gallus*, GRCg6a (chicken), *X. laevis*, xenLae2 (African clawed frog), *X. tropicalis*, xenTro9 (tropical clawed frog), *Latimeria chalumnae*, LatCha1 (coelacanth), *Lepisosteus oculatus*, LepOcu1 (spotted gar), *Danio rerio*, GRCz10 (zebrafish), *Oryzias latipes*, ASM223467v1 (Japanese medaka), *Takifugu rubripes*, FUGU5 (pufferfish), and *Callorhinchus milii*, ESHARK1 (elephant shark). For the synteny search for invertebrates, genome versions are listed in supplementary table S1, Supplementary Material online. The ciliogenic gene cluster (fig. 4*A*) fulfills a pipeline used to identify conserved microsynteny blocks in previous studies (Simakov et al. 2013, 2015; Albertin et al. 2015); Nmax 10 (maximum of 10 intervening genes) and Nmin 3 (minimum of 3 genes in a syntenic block). Unfortunately, this cluster was not detected in those studies, possibly because they used a limited number of gene families to simplify gene family assignments. The false-positive rate for

this cluster falls to <0.1%, because random genome reshuffling produces ~10% false positives in pairwise genome comparisons (Simakov et al. 2015), but the cluster was observed in more than three species across bilaterian phyla.

Transcriptomic Data from Embryos and Adult Tissues of Various Animals

Publicly available transcriptomic data of human early embryos (Xue et al. 2013), mouse early embryos (Tang et al. 2011; Xue et al. 2013), bovine early embryos (Jiang et al. 2014), chicken early embryos (Hwang et al. 2018), *X. tropicalis* embryos (Owens et al. 2016), medaka embryos (Ichikawa et al. 2017), zebrafish embryos (White et al. 2017), amphioxus embryos, and adult tissues (Marletaz et al. 2018), sea urchin embryos (*S. purpuratus* [Tu et al. 2012] and *P. lividus* [Gildor et al. 2016]), brachiopod embryos and adult tissues (Luo et al. 2015), and scallop embryos and adult tissues (Wang et al. 2017) were used to examine expression levels of *glis* genes and *dmrtb1*. Data for *X. tropicalis*, zebrafish, amphioxus, and sea urchins (*S. purpuratus*) were collected from Xenbase (http://www.xenbase.org/entry/; last accessed September 11, 2019), Expression Atlas (https://www.ebi.ac.uk/gxa/experiments/E-ERAD-475/Results; last accessed September 11, 2019), Amphiencode (http://amphiencode.github.io/; last accessed September 11, 2019), and EchinoBase (http://www.echinobase.org/Echinobase/; last accessed September 11, 2019), respectively.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Jun Inoue for conducting ORTHOSCOPE analysis and critiquing the manuscript, Naoki Irie for sharing medaka transcriptomic data, Steven D. Aird for technical editing of the manuscript, and Atsushi Hijikata and Tsuyoshi Shirai for their initial analyses of protein-interacting motifs based on the Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS) from the AMED, Japan (Grant No. JP17am0101069). We thank all members of Laboratory for Comprehensive Genomic Analysis, especially Shohei Noma, Tomoko Hirata, and Yuki Yasuoka, for technical support. This work was supported in part by a Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JSPS) (Grant Nos. 16K21559, 17KT0114, 18K14745, and 18H04825 to Y.Y.; and No. 18K19591 to M.M.), the Translational Research Network Program from the AMED, Japan (Grant No. JP18lm0203004 to M.M.), Kawano masanori memorial public interest incorporated foundation for promotion of pediatrics (M.M.), and research grants for the RIKEN Center for Integrative Medical Sciences from MEXT, Japan.

References

Ait-Lounis A, Bonal C, Seguin-Estevez Q, Schmid CD, Bucher P, Herrera PL, Durand B, Meda P, Reith W. 2010. The transcription factor Rfx3 regulates beta-cell differentiation, function, and glucokinase expression. *Diabetes* 59(7):1674–1685.

Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, Brenner S, Ragsdale CW, Rokhsar DS. 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524(7564):220–224.

Aruga J, Hatayama M. 2018. Comparative genomics of the Zic family genes. *Adv Exp Med Biol*. 1046:3–26.

Attanasio M, Uhlenhaut NH, Sousa VH, O'Toole JF, Otto E, Anlag K, Klugmann C, Treier AC, Helou J, Sayer JA, et al. 2007. Loss of GLIS2 causes nephronophthisis in humans and mice by increased apoptosis and fibrosis. *Nat Genet*. 39(8):1018–1024.

Beak JY, Kang HS, Kim YS, Jetten AM. 2008. Functional analysis of the zinc finger and activation domains of Glis3 and mutant Glis3(NDH1). *Nucleic Acids Res*. 36(5):1690–1702.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.

Dimitrieva S, Bucher P. 2013. UCNEbase—a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic Acids Res*. 41(Database issue):D101–D109.

Fujitomo T, Daigo Y, Matsuda K, Ueda K, Nakamura Y. 2014. Identification of a nuclear protein, LRRC42, involved in lung carcinogenesis. *Int J Oncol*. 45(1):147–156.

Gildor T, Malik A, Sher N, Avraham L, Ben-Tabou de-Leon S. 2016. Quantitative developmental transcriptomes of the Mediterranean sea urchin *Paracentrotus lividus*. *Mar Genomics*. 25:89–94.

Han Y, Xiong Y, Shi X, Wu J, Zhao Y, Jiang J. 2017. Regulation of Gli ciliary localization and Hedgehog signaling by the PY-NLS/karyopherin-beta2 nuclear import system. *PLoS Biol*. 15(8):e2002063.

Hashimoto H, Miyamoto R, Watanabe N, Shiba D, Ozato K, Inoue C, Kubo Y, Koga A, Jindo T, Narita T, et al. 2009. Polycystic kidney disease in the medaka (*Oryzias latipes*) pc mutant caused by a mutation in the Gli-Similar3 (glis3) gene. *PLoS One* 4(7):e6299.

Hatayama M, Aruga J. 2010. Characterization of the tandem CWCH2 sequence motif: a hallmark of inter-zinc finger interactions. *BMC Evol Biol*. 10:53.

Hatayama M, Aruga J. 2012. Gli protein nuclear localization signal. *Vitam Horm*. 88:73–89.

Hwang YS, Seo M, Lee BR, Lee HJ, Park YH, Kim SK, Lee HC, Choi HJ, Yoon J, Kim H, et al. 2018. The transcriptome of early chicken embryos reveals signaling pathways governing rapid asymmetric cellularization and lineage segregation. *Development* 145(6):dev157453.

Ichikawa K, Tomioka S, Suzuki Y, Nakamura R, Doi K, Yoshimura J, Kumagai M, Inoue Y, Uchida Y, Irie N, et al. 2017. Centromere evolution and CpG methylation during vertebrate speciation. *Nat Commun*. 8(1):1833.

Inoue J, Sato Y, Sinclair R, Tsukamoto K, Nishida M. 2015. Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. *Proc Natl Acad Sci U S A*. 112(48):14918–14923.

Inoue J, Satoh N. 2018. ORTHOSCOPE: an automatic web tool for phylogenetically inferring bilaterian orthogroups with user-selected taxa. *Mol Biol Evol*. 36(3):621–631.

Jetten AM. 2018. GLIS1-3 transcription factors: critical roles in the regulation of multiple physiological processes and diseases. *Cell Mol Life Sci*. 75(19):3473–3494.

Jiang Z, Sun J, Dong H, Luo O, Zheng X, Obergfell C, Tang Y, Bi J, O'Neill R, Ruan Y, et al. 2014. Transcriptional profiles of bovine in vivo pre-implantation development. *BMC Genomics*. 15(1):756.

Jukam D, Shariati SAM, Skotheim JM. 2017. Zygotic genome activation in vertebrates. *Dev Cell* 42(4):316–332.

Kang HS, Beak JY, Kim YS, Herbert R, Jetten AM. 2009. Glis3 is associated with primary cilia and Wwtr1/TAZ and implicated in polycystic kidney disease. *Mol Cell Biol*. 29(10):2556–2569.

Kang HS, Chen LY, Lichti-Kaiser K, Liao G, Gerrish K, Bortner CD, Yao HH, Eddy EM, Jetten AM. 2016. Transcription factor GLIS3: a new and critical regulator of postnatal stages of mouse spermatogenesis. *Stem Cells* 34(11):2772–2783.

Kang HS, Kim YS, ZeRuth G, Beak JY, Gerrish K, Kilic G, Sosa-Pineda B, Jensen J, Pierreux CE, Lemaigre FP, et al. 2009. Transcription factor

Glis3, a novel critical player in the regulation of pancreatic beta-cell development and insulin gene expression. *Mol Cell Biol*. 29(24):6366–6379.
Kang HS, ZeRuth G, Lichti-Kaiser K, Vasanth S, Yin Z, Kim YS, Jetten AM. 2010. Gli-similar (Glis) Kruppel-like zinc finger proteins: insights into their physiological functions and critical roles in neonatal diabetes and cystic renal disease. *Histol Histopathol*. 25(11):1481–1496.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30(14):3059–3066.
Keady BT, Samtani R, Tobita K, Tsuchya M, San Agustin JT, Follit JA, Jonassen JA, Subramanian R, Lo CW, Pazour GJ. 2012. IFT25 links the signal-dependent movement of Hedgehog components to intraflagellar transport. *Dev Cell* 22(5):940–951.
Kim YH, Epting D, Slanchev K, Engel C, Walz G, Kramer-Zucker A. 2013. A complex of BBS1 and NPHP7 is required for cilia motility in zebrafish. *PLoS One* 8(9):e72549.
Kim YS, Kang HS, Herbert R, Beak JY, Collins JB, Grissom SF, Jetten AM. 2008. Kruppel-like zinc finger protein Glis2 is essential for the maintenance of normal renal functions. *Mol Cell Biol*. 28(7):2358–2367.
Kim YS, Lewandoski M, Perantoni AO, Kurebayashi S, Nakanishi G, Jetten AM. 2002. Identification of Glis1, a novel Gli-related, Kruppel-like zinc finger protein containing transactivation and repressor functions. *J Biol Chem*. 277(34):30901–30913.
Kim YS, Nakanishi G, Lewandoski M, Jetten AM. 2003. GLIS3, a novel member of the GLIS subfamily of Kruppel-like zinc finger proteins with repressor and activation functions. *Nucleic Acids Res*. 31(19):5513–5525.
Layden MJ, Meyer NP, Pang K, Seaver EC, Martindale MQ. 2010. Expression and phylogenetic analysis of the zic gene family in the evolution and development of metazoans. *Evodevo* 1(1):12.
Lee SY, Noh HB, Kim HT, Lee KI, Hwang DY. 2017. Glis family proteins are differentially implicated in the cellular reprogramming of human somatic cells. *Oncotarget* 8(44):77041–77049.
Luo YJ, Takeuchi T, Koyanagi R, Yamada L, Kanda M, Khalturina M, Fujie M, Yamasaki SI, Endo K, Satoh N. 2015. The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat Commun*. 6:8301.
Maekawa M, Yamaguchi K, Nakamura T, Shibukawa R, Kodanaka I, Ichisaka T, Kawamura Y, Mochizuki H, Goshima N, Yamanaka S. 2011. Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. *Nature* 474(7350):225–229.
Mansfeld J, Guttinger S, Hawryluk-Gara LA, Pante N, Mall M, Galy V, Haselmann U, Muhlhausser P, Wozniak RW, Mattaj IW, et al. 2006. The conserved transmembrane nucleoporin NDC1 is required for nuclear pore complex assembly in vertebrate cells. *Mol Cell* 22(1):93–103.
Marletaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, de la Calle-Mustienes E, Bertrand S, Burguera D, et al. 2018. Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* 564(7734):64–70.
Materna SC, Howard-Ashby M, Gray RF, Davidson EH. 2006. The C2H2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development. *Dev Biol*. 300(1):108–120.
Nakashima M, Tanese N, Ito M, Auerbach W, Bai C, Furukawa T, Toyono T, Akamine A, Joyner AL. 2002. A novel gene, GliH1, with homology to the Gli zinc finger domain not required for mouse development. *Mech Dev*. 119(1):21–34.
O'Hare EA, Yerges-Armstrong LM, Perry JA, Shuldiner AR, Zaghloul NA. 2016. Assignment of functional relevance to genes at type 2 diabetes-associated loci through investigation of beta-cell mass deficits. *Mol Endocrinol*. 30:429–445.
Ounjai P, Kim KD, Liu H, Dong M, Tauscher AN, Witkowska HE, Downing KH. 2013. Architectural insights into a ciliary partition. *Curr Biol*. 23(4):339–344.
Owens NDL, Blitz IL, Lane MA, Patrushev I, Overton JD, Gilchrist MJ, Cho KWY, Khokha MK. 2016. Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. *Cell Rep*. 14(3):632–647.
Piasecki BP, Burghoorn J, Swoboda P. 2010. Regulatory Factor X (RFX)-mediated transcriptional rewiring of ciliary genes in animals. *Proc Natl Acad Sci U S A*. 107(29):12969–12974.
Scoville DW, Kang HS, Jetten AM. 2017. GLIS1-3: emerging roles in reprogramming, stem and progenitor cell differentiation and maintenance. *Stem Cell Investig*. 4:80.
Senee V, Chelala C, Duchatelet S, Feng D, Blanc H, Cossec JC, Charon C, Nicolino M, Boileau P, Cavener DR, et al. 2006. Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat Genet*. 38(6):682–687.
Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, Fukui A, Hikosaka A, Suzuki A, Kondo M, et al. 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* 538(7625):336–343.
Shimeld SM. 2008. C2H2 zinc finger genes of the Gli, Zic, KLF, SP, Wilms' tumour, Huckebein, Snail, Ovo, Spalt, Odd, Blimp-1, Fez and related gene families from *Branchiostoma floridae*. *Dev Genes Evol*. 218(11-12):639–649.
Simakov O, Kawashima T, Marletaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J, Shoguchi E, Gyoja F, et al. 2015. Hemichordate genomes and deuterostome origins. *Nature* 527(7579):459–465.
Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo DH, Larsson T, Lv J, Arendt D, et al. 2013. Insights into bilaterian evolution from three spiralian genomes. *Nature* 493(7433):526–531.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 34(Web Server issue):W609–W612.
Tadros W, Lipshitz HD. 2009. The maternal-to-zygotic transition: a play in two acts. *Development* 136(18):3033–3042.
Takeuchi T, Koyanagi R, Gyoja F, Kanda M, Hisata K, Fujie M, Goto H, Yamasaki S, Nagai K, Morino Y, et al. 2016. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological Lett*. 2:3.
Tang F, Barbacioru C, Nordman E, Bao S, Lee C, Wang X, Tuch BB, Heard E, Lao K, Surani MA. 2011. Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS One* 6(6):e21208.
Tu Q, Cameron RA, Worley KC, Gibbs RA, Davidson EH. 2012. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res*. 22(10):2079–2087.
Vasanth S, ZeRuth G, Kang HS, Jetten AM. 2011. Identification of nuclear localization, DNA binding, and transactivating mechanisms of Kruppel-like zinc finger protein Gli-similar 2 (Glis2). *J Biol Chem*. 286(6):4749–4759.
Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, Guo X, Huan P, Dong B, Zhang L, et al. 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol*. 1(5):120.
Watanabe M, Yasuoka Y, Mawaribuchi S, Kuretani A, Ito M, Kondo M, Ochi H, Ogino H, Fukui A, Taira M, et al. 2017. Conservatism and variability of gene expression profiles among homeologous transcription factors in *Xenopus laevis*. *Dev Biol*. 426(2):301–324.
White RJ, Collins JE, Sealy IM, Wali N, Dooley CM, Digby Z, Stemple DL, Murphy DN, Billis K, Hourlier T, et al. 2017. A high-resolution mRNA expression time course of embryonic development in zebrafish. *Elife* 6:e30860.
Xue Z, Huang K, Cai C, Cai L, Jiang CY, Feng Y, Liu Z, Zeng Q, Cheng L, Sun YE, et al. 2013. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500(7464):593–597.
Yang KY, Chen Y, Zhang Z, Ng PK, Zhou WJ, Zhang Y, Liu M, Chen J, Mao B, Tsui SK. 2016. Transcriptome analysis of different developmental stages of amphioxus reveals dynamic changes of distinct classes of genes during development. *Sci Rep*. 6:23195.
ZeRuth GT, Takeda Y, Jetten AM. 2013. The Kruppel-like protein Gli-similar 3 (Glis3) functions as a key regulator of insulin transcription. *Mol Endocrinol*. 27(10):1692–1705.

ZeRuth GT, Williams JG, Cole YC, Jetten AM. 2015. HECT E3 ubiquitin ligase itch functions as a novel negative regulator of Gli-Similar 3 (Glis3) transcriptional activity. *PLoS One* 10(7):e0131303.

ZeRuth GT, Yang XP, Jetten AM. 2011. Modulation of the transactivation function and stability of Kruppel-like zinc finger protein Gli-similar 3 (Glis3) by suppressor of fused. *J Biol Chem*. 286(25):22077–22089.

Zhang F, Jetten AM. 2001. Genomic structure of the gene encoding the human GLI-related, Kruppel-like zinc finger protein GLIS2. *Gene* 280(1-2):49–57.

Zhang F, Nakanishi G, Kurebayashi S, Yoshino K, Perantoni A, Kim YS, Jetten AM. 2002. Characterization of Glis2, a novel gene encoding a Gli-related, Kruppel-like transcription factor with transactivation and repressor functions. Roles in kidney development and neurogenesis. *J Biol Chem*. 277(12):10139–10149.