Identification of unique α4 chain structure and conserved anti-angiogenic activity of α3NC1 type IV collagen in zebrafish

Valerie S. LeBleu[1,2,3*#], Jianli Dai[1#], Susan Tsutakawa[4], Brian A. MacDonald[2], Joseph L. Alge[1,5], Malin Sund[3], Liang Xie[3], Hikaru Sugimoto[1,3], John Tainer[1,7], Leonard I. Zon[6] and Raghu Kalluri[1,3,8,9*]

[1]*Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, TX*
[2]*Feinberg School of Medicine and Kellogg School of Management, Northwestern University, Chicago, IL*
[3]*Division of Matrix Biology, Beth Israel Deaconess Medical Center, Boston, MA*
[4]*Lawrence Berkeley National Laboratory, University of California, Berkeley, CA*
[5]*Department of Pediatrics, Baylor College of Medicine, Houston, TX*
[6]*Department of Hematology/Oncology, Children's Hospital, Boston, MA*
[7]*Department of Molecular and Cellular Oncology, University of Texas MD Anderson Cancer Center, Houston, TX*
[8]*School of Bioengineering, Rice University, Houston, TX*
[9]*Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX*

[#] co-first authors

*****To whom correspondence should be addressed: Raghu Kalluri: Department of Cancer Biology, Metastasis Research Center, University of Texas MD Anderson Cancer Center, Houston, TX 77054 rkalluri@mdanderson.org; Valerie LeBleu: Feinberg School of Medicine & Kellogg School of Management, Northwestern University, Chicago, IL 60611 valerie.lebleu@northwestern.edu

Running Title: Type IV collagen in zebrafish development

Abstract

Type IV collagen, the most abundant component of basement membranes, is essential for the formation of the extracellular scaffold that supports tissue architecture and function. Collagen IV is present in all multicellular species, with lower organisms typically possessing two type IV collagen genes, encoding α1 and α2 chains. The human genome encodes for six different type IV collagen genes, α1 to α6 chains. The α chains assemble into trimeric protomers, the building blocks of the type IV collagen network. The detailed evolutionary conservation of type IV collagen network organization remains to be studied. Here we report on the molecular evolution of type IV collagen genes. Specifically, we report on the zebrafish α4 NC1 domain, which, in contrast with its human ortholog, contains an additional cysteine residue and lacks the Met^{93} and Lys^{211} residues involved in sulfilimine bond formation between adjacent protomers. This may alter α4 chain interactions with other α chains, as supported by temporal and anatomic expression patterns of collagen IV chains during zebrafish development. Despite the divergence between zebrafish and human α3 NC1 domain (human endogenous inhibitor of angiogenesis, Tumstatin), the zebrafish α3 NC1 domain exhibits conserved anti-angiogenic activity in human endothelial cells. Our work supports that sequence identity, tissue expression, and function of type IV collagen have remained largely conserved between zebrafish and humans, with one possible difference involving the α4 NC1 domain.

Key words: collagen IV, basement membrane, Tumstatin, angiogenesis, zebrafish

Abbreviations: NC1 domain, non-collagenous 1 domain; GBM, glomerular basement membrane; hpf, hour post fertilization.

Introduction

Type IV collagen (collagen IV) network assembly is essential for basement membrane organization and crucial for multicellular life [1-6]. The formation of the type IV collagen network is a complex process that involves both intracellular and extracellular steps. In the endoplasmic reticulum, three collagen IV α chains assemble into heterotrimeric helical molecules, known as protomers, which serve as the building blocks of the type IV collagen network [7]. After being secreted from cells, type IV collagen protomers undergo oligomerization and self-assembly into insoluble sheet-like supramolecular networks [8-11]. Lower organisms, such as *C.elegans* and *Drosophila*, have two distinct chains, namely α1 and α2, whereas mammals express six distinct chains named α1 through α6, encoded by six distinct genes (*Col4a1* – *Col4a6*) [12]. The chain composition of type IV collagen network is in part dictated by unique type IV collagen protomer nucleation and formation, which are specified by α chain non-collagenous (NC1) domain interaction [13-15].

Despite many possible combinations, predominantly only three distinct protomers α1α2α1, α3α4α5, and α5α6α5 have been identified in basement membranes [16, 17]. The collagen IV α1α2α1 network, considered to be more primordial, is ubiquitously found in basement membranes of all tissues. The α1α2α1 network is thought to be critical to the evolution of multicellular life [18-20]. Indeed, genetic deletion of collagen IV α1 and/or α2 genes in mice results in a loss of basement membrane integrity and embryonic lethality [21-23], and loss of the collagen IV α1 alone in *C. elegans* and *Drosophila* leads to developmental arrest and embryonic lethality [4, 24-26]. It is also reported that collagen IV α1 and α2 mutations correlate with diseases in humans [27]. In contrast, collagen IV α3-α6 chains, only found in higher-order organisms, have restricted tissue distribution during development, and present in more specialized basement membranes [16]. For instance, the α3α4α5 (IV) network plays a prominent role in the mammalian kidney, where it is a critical component of the filtration barrier in the glomerular basement membrane (GBM). Mutations in α3-, α4-, and α5-chain encoding genes leads to the absence of α3α4α5 (IV) protomer formation in GBM, which is associated with several nephropathies, including Alport syndrome, thin basement

membrane nephropathy, and focal segmental glomerulosclerosis [5, 28-31]. The α3α4α5 protomer is also noted in the basement membranes of many other tissues, including brain, alveoli, testes, cochlea, the synapses of neuromuscular junctions, and lung epithelia in mammals [32-38]. Genetic disruption of the α3α4α5 network composition in these basement membranes is associated with multiple organ impairment in Alport syndrome [39]. In addition, autoantibodies against the NC1 domain of the collagen IV α3 chain results in Goodpasture syndrome, which is characterized by pulmonary hemorrhage and glomerulonephritis [40, 41].

Beyond its likely role in structural stability of Collagen IV network in basement membranes[13], the cleavage of the NC1 domain releases the bioactive α3 NC1 fragment, Tumstatin, which exerts anti-angiogenic and pro-apoptotic functions that significantly influences tumor growth [42, 43]. Interestingly, zebrafish, like mammals, express six collagen IV α chains [44]. The development of the collagen IV α3α4α5 and α5α6α5 networks are thought to result from more recent evolutionary events, and these networks may not be as highly conserved as the primordial α1α2α1 (IV) network. Indeed, the α3-chain cryptic epitope/s targeted by the autoantibodies in Goodpasture syndrome do not appear to be conserved in zebrafish [45]. Little is otherwise known about the evolutionary divergence of collagen IV basement membrane composition from the zebrafish to mammals [46, 47]. Here we report on the degree of homology of NC1 domain between the six zebrafish collagen IV α chains and their human homologues. We describe temporospatial expression pattern of the six collagen IV chains in zebrafish development at the mRNA level. We also demonstrate that the zebrafish α3 chain NC1 domain (zTumstatin) is present with conserved anti-angiogenic activity when compared to human Tumstatin.

Material and methods

NC1 Domain Molecular Analysis and 3D rendering

Amino acid sequences were obtained from UniProt database. The color scheme used for alignments is from the algorithm in Clustal X. Each amino acid is colored when it meets the criteria specific for the residue type. This analysis is conducted by Clustal X and viewed by Jalview. The 3D modeling of the structures of the [(α1)2(α2)]2 NC1 hexamers from human placenta basement membranes (Protein Data Bank ID 1LI1) were rendered using the molecular graphics visualization program YASARA (YASARA Bio-sciences) and analyzed to generate modeled structures using the molecular graphics software InsightII (Accelrys). Percentage of homology was calculated by multiplying the query coverage and percent identity of each comparison with usage of NCBI blast tool.

Animal Care

AB and Tu strain zebrafish were bred and maintained as described [45]. C57BL/6 mice (Charles River Laboratories) were used for the *in vivo* angiogenesis assay. Animal studies were reviewed and approved by the institutional animal care and use committee at Beth Israel Deaconess Medical Center and Boston Children's Hospital.

RT-PCR Determination of Embryonic and Adult Expression Pattern

At each of the 13 developmental stages described therein, total RNA was isolated, and cDNA was generated from pools of 25 embryos. For adult AB strain, ten distinct organs were isolated, placed in TRIzol™ (Invitrogen) reagent and immediately homogenized. Generation of cDNA was performed using Superscript II™ reverse transcriptase (Invitrogen) and oligo dT primers (Invitrogen). RT-PCR primers (below) were designed to regions more unique than those used to clone the NC1 domains, and using the NCBI BLAST program to assure that none of these primers could cross-amplify any other zebrafish type IV collagen α chains.

zα1 NC1 Forward:	5'-TCCTCAATGGACCACGGCTTCCTT-3'
zα1 NC1 Reverse:	5'-CCTGCTGATGTGAGTTCTTAG-3'
zα2 NC1 Forward:	5'-CGCAGTGTGAACGTTGGCTAT-3'
zα2 NC1 Reverse:	5'-TTGGCAGCGGATGATGCGGGA-3'
zα3 NC1 Forward:	5'-AGAGACGGATTCCTTTTCACC-3'
zα3 NC1 Reverse:	5'-AGGCTGTTGCTTCATACAGAC-3'
zα4 NC1 Forward:	5'-GGAAGAAATGAGACACTTTCT-3'
zα4 NC1 Reverse:	5'-TCATTGGTTGGGGTCATTCAT-3'
zα5 NC1 Forward:	5'-ACTTTGTCTGGTGCACACGGT-3'
zα5 NC1 Reverse:	5'-TTACGTCCTCTTCATACACAC-3'
zα6 NC1 Forward:	5'-ACGCGCAGCACTGGCATTGGT-3'
zα6 NC1 Reverse:	5'-ACTGAGACGCGAGCGGAACTG-3'

Probe Generation and Whole Mount *in-situ* Hybridization of Embryos

Whole-mount *in situ* hybridization was performed as described [48]. Digoxygenin-labeled antisense RNA probes were synthesized using a DIG RNA Labeling NTP mix (Roche) and either SP6 or T7 RNA polymerase (Invitrogen). Primers used for generating probes to amplify the 3'- untranslated regions of type IV collagen transcripts (α1- α6) in zebrafish are:

zα1 NC1 Forward:	5'-TCCTGCCTCAGCGCTCTTTCTCAT-3'
zα1 NC1 Reverse:	5'-CATAAGTTTTCATAAATAACTAAG-3'
zα2 NC1 Forward:	5'-CTCCTGTCCCGCATCAGCCGCTGC-3'
zα2 NC1 Reverse:	5'-GCAAGATAAAACCCCCTGTTGCC-3'
zα3 NC1 Forward:	5'-CAACAGCCTCAAGGTGACCTTGCT-3'
zα3 NC1 Reverse:	5'-GGGCATGCCTTGCTTACAATGGAAACCGCC-3'
zα4 NC1 Forward:	5'-TGATACAAGCCAATGGTATTTTCT-3'
zα4 NC1 Reverse:	5'-TTACTATTGTAGGTGAAATATCCC-3'
zα5 NC1 Forward:	5'-CACATCAGCAGGTGTCAAGTGTGCATG-3'
zα5 NC1 Reverse:	5'-CCATACGCCGCTGCACAAAAGAAACAGCGC-3'
zα6 NC1 Forward:	5'-AGCTTGTGTAGATGCACTGCAGGA-3'
zα6 NC1 Reverse:	5'-GTGAAGCAGCCATTGACATCCATA-3'

Recombinant Protein Production

A cDNA encoding zTumstatin (zebrafish Tumstatin), was subcloned into the pcBFT vector containing a FLAG tag. The vectors were then transfected into human embryonic kidney cells (HEK293) with Fectin-system (Invitrogen). Protein from culture supernatant was isolated as previously described [45, 49].

In Vitro Cell Proliferation Assay

Human umbilical vein endothelial cells (HUVEC) were obtained from Lonza and cultured in EGM-2 cell culture medium (Lonza). Assays were carried out in 96-well plates, with 6,000 cells per well and using WST-1 reagent (Roche). Absorbances were measured at 450nm and 650nm on a Versamax turntable microplate reader (Molecular Devices). Three independent wells were measured for each condition and the data presented as absorbance (A) values at 650nm subtracted from 450nm (A_{450} - A_{650}). Positive control: 10% FBS; Negative control: 0.1% FBS. HEK293 produced zebrafish Tumstatin was supplemented to 10% FBS at the indicated concentration.

Migration Assay

HUVEC (5,000 cells/ well) were seeded onto the top chambers of a 48-well Boyden chamber (Neuro Probe Inc) in the presence of un-supplemented media or media supplemented with 10 μM zTumstatin. HUVEC were separated from bottom wells containing EBM culture media (Lonza) supplemented with 0.1% FBS and 40 ng/ml of recombinant VEGF (R&D). The negative control did not contain VEGF in the bottom well. Top and bottom wells were separated by 8 μm polycarbonate membranes (Neuro Probe Inc). Membranes were fixed and stained using the PROTOCOL HEMA 3 Stain Set (Fisher Scientific) after removing all cells from the top side of the membrane. The number of migrated cells was scored under a 20x objective by light microscopy. Each well was evaluated with 2 to 4 images captured per well. For each group, 4 to 5 wells were evaluated.

Matrigel™ Plug Assay

Matrigel™ (BD Biosciences) was mixed with 20 units/ml heparin (Pierce), 50 ng/ml bFGF and VEGF (R&D), and 10 μmol/L zTumstatin. Control plugs did not contain zTumstatin. The Matrigel™ mixture was injected subcutaneously in C57Bl/6 mice, 2 plugs per mouse. After 7 days, the mice were sacrificed, and the Matrigel™ plugs were removed and fixed in 10% buffered formalin. The plugs were then embedded in paraffin, sectioned, and stained by hematoxylin and eosin. Sections were examined by light microscopy, and the number of blood vessels counted from five or more images taken for each plug, and the number of

vessels counted was averaged for each plug. A total of 8 plugs from 4 mice were analyzed in each group, statistical analyses were performed on plugs.

Quantification and statistical analysis

GraphPad Prism software was used for graphic representation and statistical analysis. One-way ANOVA method was used in **Fig. 5b-c**, compared to positive control. Two-tailed unpaired Student's t test was used in **Fig. 5e**. The data is presented as the mean +/- the standard error of the mean. Significance of statistical tests is reported in all graphs as follows: ****, $p<0.0001$; ***, $p<0.001$; **, $p<0.01$.

Results

The zebrafish type IV collagen alpha 4 chain is distinct from its human homolog

Type IV collagen NC1 domains in mammals are critical for protomer assembly and present with anti-angiogenic activities [50-53], and the head-to-head organization of the Col4α1-Col4α2 pairs, Col4α3-Col4α pairs and Col4α5-Col4α6 pairs was found in human, mouse and zebrafish species (**Supplementary Fig. 1**). A high degree of sequence homology between the NC1 domains of zebrafish α1, α2, α5, and α6 chains and the murine and human orthologs was noted, with more pronounced dissimilarity for α3 and α4 chains (**Tables 1-2, Fig. 1, Supplementary Fig. 2**). Comparative analysis of the zebrafish NC1 subdomains across the six Col4 chains revealed that specific features, critical for the maintenance of the tertiary structure of the NC1 domain and protomer assembly, are highly conserved between zebrafish and human (**Fig. 2a**). All six zebrafish Col4 chains contain the HSQ tripeptides that signal the beginning of the NC1 subdomain I and II (**Fig. 2a**, cyan). Each zebrafish α chain also contains twelve conserved cysteine residues (**Fig. 2a**, green) that maintain the domain structure essential for protomer assembly. It is reported that the collagen IV network is stabilized by protomer dimerization at adjoining NC1 domains, preceded by a chloride-dependent NC1 domain conformational change through salt bridges formed between the R76 D78, E175-R179, and N187 residues [10]. These residues are mostly conserved in all six zebrafish α chains, except for R179 in the α2 chain, for which a R179K substitution in is noted (**Fig. 2a**, red, red arrow). In addition, the zebrafish α3 and α4 NC1 domains lack the N187 residue (**Fig. 2a**, red). It is speculated that once the NC1 domain alignment occurs, protomers are linked by a sulfilimine bond between the M93 and K211 residues in a reaction catalyzed by the highly conserved peroxidasin enzyme [9, 11, 54]. Five of the six zebrafish α chains (except for the α4 chain) contain the M93 and K211 residues associated with the formation of the sulfilimine bond and protomer dimerization (**Fig. 2a**, yellow) [9].

In contrast with other collagen IV α chains, the zebrafish α4 chain has two important divergent features. First, the zebrafish collagen IV α4 chain contains an additional cysteine immediately before the third cysteine residue (Cys-3) in the NC1 subdomain I, which we refer to as cysteine-13 (Cys-13; **Fig. 2a**,

magenta). The Cys-3 to Cys-4 disulfide bond forms a β-hairpin structure in the SM2 subdomain (β6–β7) in the NC1 domain that inserts into the SM3' subdomain of an adjacent NC1 domain [8, 55-57]. Three-dimensional modeling analysis reveals that the sulfur atoms of Cys-3 and Cys-13 are a mere 7.4 angstroms apart, and such extreme proximity makes Cys-13 a plausible alternative binding partner with Cys-4. The omission of one amino acid residue makes the β–hairpin in SM2 of the zebrafish α4-chain shorter than in other species and effectively reduces the 7.4 angstrom distance (**Fig 1**). This deletion, coupled with the natural flexibility in the peptide backbone of the domain (asterisk, **Fig. 2b**), makes Cys-13 a plausible alternative binding partner for Cys-4. *In silico* analysis with Polyphen-2 indicates that the Cys-13 substitution in human Col1a1 is likely damaging (score 1.000). We therefore hypothesize that Cys-13 in the zebrafish α4-chain alters the structure of the β-hairpin region of NC1 domain, and potentially allows for distinct interactions with other α chains and formation of novel protomer (**Fig. 2b**, **Supplemental Fig. 3**) [58]. Second, the zebrafish α4 chain lacks the M93 and K211 residues required for the peroxidasin-catalyzed sulfilimine bond formation that stabilizes end-to-end protomer dimerization [9]. Such deficit would prevent the covalent linking of α4 containing-protomers at the C terminus. These results suggest that zebrafish collagen IV networks containing the α4 chain may be weaker than their human homolog.

Unique temporal expression patterns of type IV collagen chains in zebrafish embryologic development

To gain further insight into the biosynthesis of collagen IV protomers in zebrafish embryogenesis, we performed RT-PCR on cDNA isolated from zebrafish embryos at 13 distinct developmental time points. The results show three distinct expression patterns: temporally unrestricted expression, post-gastrulation expression, and late-onset expression (**Fig. 3a**). Expression of α1 and α2 chains mRNA was not restricted to a particular developmental stage, and the synchronous expression patterns of the α1 and α2 chains supports that zebrafish synthesize the α1α2α1 (IV) protomer throughout embryogenesis (**Fig. 3a**). In

addition, transcription of genes encoding the α5 and α6 chains was synchronously detected in post-gastrulation embryos, suggesting that synthesis of the α5α6α5 (IV) protomer does not occur in the pre-gastrulation stages of embryogenesis (**Fig. 3a**). Finally, in late stage of embryos, when there was also a high level of α5 chain gene transcription, transcription of α3 and α4 chains was observed, supporting that the possible synthesis of the α3α4α5 (IV) protomer in later stages of zebrafish embryonic development (**Fig. 3a**). Our results support that the zebrafish embryo expresses the Col4 chains, suggesting the possible synthesis of the three collagen IV protomers known in mammals.

We also observed that the α1 and α4 chains are synchronously expressed in the absence of other chains at the shield time point (**Fig. 3a**, lane 3). In addition, there is no detectable α3 chain expression despite persistent expression of α4 at the Prim-11 stage (**Fig. 3a**, lane 12). This raises the possible existence of novel protomer involving the α4 chain other than the α3α4α5 (IV) protomer at this stage, such as α1α4α1. In zebrafish, expression of proteins before 6 hours post hybridization (hpf) stage is maternal. Only the α5 chain is not maternally contributed, which also supports that the α3- and α4-chains could assemble protomers with other composition than the α3α4α5 (IV) protomer. Though based on transcriptional evaluation only, our findings suggest distinct type IV collagen network composition may exist in the zebrafish development. Given that an α1-α4 heterohexamer is not formed in human, we postulated that mutation in zebrafish α4 allow for this formation. Targeting regions identified to be critical for oligomerization [59], we identified any changes in zebrafish α4 NC1 that made it similar to α1 and α2 NC1. Although there were several mutations that may render zebrafish α4 to be similar to α2, only one change made it similar to both α1 and α2 in one region, SM1' (structural motif SM1': hairpin β3'– β4' [59]) (**Fig. 1**). Examination of the α1/α2 heterohexamer structure from human placenta [56] revealed that the mutation in zebrafish α4 to glutamate (E1589) would make it analogous to E1592 and E1633 in α1 and α2 respectively (**Supplemental Fig. 4**). These analyses raise the possibility for α4 chain containing (IV) protomer that include either the α1 and α2 chain.

Organ-specific expression patterns of type IV collagen chains in zebrafish development

To define the anatomic position of collagen IV networks in zebrafish development, we performed *in situ* hybridization for each of the six collagen IV α-chain transcripts in the zebrafish embryo at 26 hpf and 36 hpf. Consistent with a recent study [60], our results show that the α1 and α2 chains were transcribed ubiquitously (**Fig. 3b**). Although the caudal expression level was low, no specific organ displayed a differential distribution of α1 or α2 chain. The α3 and α4 chains, however, displayed organ specific expression, with transcript accumulation in the brain (b) and otic vesicle (o) at 26 hpf (time point of maximal intensity; **Fig. 3c**). A similar expression pattern for the α5 chain in the brain and otic vesicles at 26 hpf supports the formation of the α3α4α5 protomer in these organs (**Fig. 4a**). At 36 hpf, the α5 and α6 chain expression in the otic vesicle, gills (g), and pronephric ducts (pn) offers support for the synthesis of the α5α6α5 protomer in these tissues (**Fig. 4b**). Prominent detection of the α6 chain in cardiac tissue (c) in the absence of α5 chain suggests a distinct type IV collagen network composition with possible novel protomer in the zebrafish cardiac development (**Fig. 4b**).

In humans and mice, organs such as intestine and skin are strictly composed of the collagen IV α1α2α1 and α5α6α5 protomers, yet other organs display a developmental restricted expression of specific protomers [38, 61, 62]. For example, during human nephrogenesis, the GBM is initially comprised of the collagen IV α1α2α1 protomer. As the glomerulus matures, the GBM collagen IV undergoes an isoform switch from the α1α2α1 to the α3α4α5 protomer [63, 64]. Our results show that the expression of the more specialized α3α4α5 protomer was not detected by *in situ* hybridization in many of the zebrafish organs during embryonic development. We posit that the absence of the α3α4α5 protomer in the developing organs of the zebrafish in its embryonic development could reflect a similar post-embryonic stage isoform switching mechanism. To test this hypothesis, we performed RT-PCR for the six collagen IV alpha chain transcripts in distinct organs in the adult zebrafish. The α1, α2, α5 and α6 chains were ubiquitously expressed in the adult zebrafish (although α5 chain expression was low in the cardiac tissue), but the α3 and α4 chains exhibited a more restricted expression pattern (**Fig. 4c**). The α3 chain was expressed in the

brain/olfactory bulb, gills, kidney, muscle, swim bladder, and testis, but was absent in the gut, heart, liver and skin (**Fig. 4c**). The α4 chain had a similar expression pattern, but it was also expressed in cardiac and hepatic tissue (**Fig. 4c**). These results support a protomer isoform switch in the post-embryonic zebrafish development, similarly to what is observed in mammals. The α4 chain expression in the adult cardiac tissue, concurrent with absent α3 and low α5 chain expression, indicates a more promiscuous α4 chain expression in zebrafish compared to human, and supports potential formation of novel protomers composition.

Conserved anti-angiogenic function of the zebrafish collagen IV α3 NC1 domain, zTumstatin

Given that Tumstatin, the NC1 domain of collagen IV α3 chain, is a known anti-angiogenic cleavage product of Col4α3, we wondered whether zebrafish α3-chain NC1 domain, zTumstatin, would also exert anti-angiogenic functions, similarly to its human homolog [42, 43, 65]. To test this hypothesis, we produced recombinant zTumstatin using a eukaryotic expression system, as previously described [45]. Human umbilical vein endothelial cells (HUVEC) proliferation was inhibited in a dose-dependent manner by eukaryotic zTumstatin (**Fig. 5a**). zTumstatin also inhibited HUVEC migration in a Boyden chamber assay (**Fig. 5b**). The anti-angiogenic property of zTumstatin was also tested in vivo in a Matrigel™ plug assay. Subcutaneous Matrigel™ plugs, containing bovine FGF and VEGF, and supplemented with or without zTumstatin, were implanted in C57BL/6 mice. Control plugs (no zTumstatin) demonstrate robust angiogenesis, whereas in zTumstatin-containing plugs showed minimal levels of angiogenesis (**Fig. 5c**). Microscopic angiogenesis, evaluated by H&E staining of Matrigel™ plugs, also revealed suppressed angiogenesis (54% reduction) in zTumstatin-containing plugs compared to control (**Fig. 5d**). These results support antiangiogenic function of the zTumstatin.

Discussion

Collagen IV exists in all metazoans, including sponges [18]. Type IV collagen has pivotal function in shaping tissue and organs, transducing signaling pathways and maintaining normal development in multicellular organisms [2, 18]. Zebrafish, as one of the lowest vertebrates known to express the six collagen IV homologous to those in humans, is a powerful model system to study conserved function of collagen IV and therefore collagen IV-related disease. The evolution of the type IV collagen genes includes the initial tandem duplication of *Col4a1* or *Col4a2* and head-to-head organization of the *Col4a1-Col4a2* gene pair and a shared bi-directional promoter. The pair possibly duplicated onto distinct chromosomes, resulting in a similar head-to-head organization of the *Col4a3-Col4a4* and *Col4a5-Col4a6* pairs; and this gene arrangement is conserved between zebrafish, mouse, and human. This may explain the increased similarities in the NC1 domains, conserved to preserve specific network assembly, of the α3 chain with α1 and α5 chain, and the increased similarities in the α4 chain with the α2 and α6 chains [18, 66]. The comparison of the sequence of β-hairpin SM2 subdomain and SM3' subdomain therein also supports this interpretation.

We also show the conserved twelve cysteine residues required for maintenance of NC1 domain structure in all six zebrafish α chains when compared to human NC1 domains. The residues involved in the formation of the chloride -salt-bridge-mediated sulfilimine bond, through a conformational change of NC1 domain, are conserved in all six collagen IV α chains in zebrafish. In addition, the enzymes that catalyze collagen IV protomer cross-linking in the NC1 and 7-S domains, peroxidasin and lysyl oxidase-like 2 [11, 57, 67], are also conserved. We note that the zebrafish α4-chain NC1 domain lacks M93 and K211 residues and contains an additional cysteine residue. The M93 and K211 residues are joined by a sulfilimine bond required for linking adjacent protomers [9, 56], suggesting distinct adjacent protomer interaction for the zebrafish α4-containing protomer. 3D remodeling in our study also supports that the additional cysteine residue alters the zebrafish α4-containing protomer interaction with other protomers.

Our study also offers a comprehensive analysis of the temporospatial expression pattern of each α chain throughout zebrafish development at the RNA level. Our results support that zebrafish generates the collagen IV protomers known to be synthesized in humans: α1α2α1, α3α4α5 and α5α6α5. Similarly, in mice, spatial and temporal distribution of collagen IV isoforms in the developing eye also reveals these protomers [68]. In our study, we note that the transcriptional pattern of α1 and α2 chain is widely distributed, whereas which of the α3-α6 chain exhibits a restricted, organ specificity pattern, consistent with recent findings [60]. Synchronous expression of α3-, α4- and α5- chains in the brain suggests that α3α4α5 (IV) protomer is synthesized there at an early time point, which agrees with the results of a recent study highlighting the importance of this protomer in zebrafish neural development [66]. Furthermore, synchronous expression of these α chains in the otic vesicle and pronephros support an evolutionary conserved anatomic specificity for α3α4α5 synthesis, given that this protomer is the major component of the cochlear and glomerular basement membranes in humans. The chains expression enabling α3α4α5 network synthesis is not detectable in these organs until late time points, which supports a similar, evolutionary conserved, collagen IV isoform switching event in these tissues [63, 64].

Our findings unexpectedly yielded complementary evidence (homology and temporospatial expression) that the zebrafish collagen IV protomer repertoire may include novel protomers. For instance, zebrafish α4 chain was synchronously expressed with the α1 chain in the absence of α2, α3 and α5 chains at the shield time point and in cardiac tissue of adult zebrafish, wherein α3 and α5 chain expression were lacking or were minimal. These findings suggest the potential existence of a putative α1α4α1 protomer. The zebrafish α4 chain NC1 domain containing an additional cysteine residue, which is predicted to lead to structural changes of the β-hairpin loop, may enable additional diversity of protomers [55-57]. Lack of M93 and K211 residues and an additional cysteine residue may trigger distinct adjacent protomer interaction for the zebrafish α4-containing protomer. Collectively, our results suggest that apart from α3α4α5, other potential collagen IV protomers (α1α4α1) may present in zebrafish. Although, our analyses rely on gene transcription findings, other reports have shown protein-level distribution of collagen IV α3

and α4 that supports our findings [49, 69]. Collagen IV α3 protein expression in the adult zebrafish is in the gills and kidney [45], which is in agreement with our RNA analyses. Col4a4 protein was detected in the heart of the developing zebrafish [69]. Future studies are needed to evaluate zebrafish α chain interactions at the protein level.

Finally, we report on the conserved function of the zebrafish collagen IV α3 chain NC1 domain (zTumstatin). Similar to its human ortholog (Tumstatin), zTumstatin presents with anti-angiogenic activity in vitro and in vivo. The epitope that is targeted by autoantibodies in Goodpasture syndrome is however not conserved [45]. This suggests a distinct immune system maturation and regulation of the zebrafish compared to human, in particular with respect to the autoimmune response to collagen IV α3 chain in Goodpasture syndrome [70]. Our studies reveal evolutionary conservation of collagen IV NC1 domain between zebrafish and humans at the molecular and functional level and, together with other reports, made advances in understanding collagens by using model organisms [71].

Acknowledgements: The authors would like to express our gratitude to Noelle Paffett-Lugassy and Caroline Burns for guidance in isolating the numerous adult zebrafish tissues. This study was partially funded by a research grant from Emerald Foundation, NIH research grants DK 55001 (RK), Cancer Prevention and Research Institute of Texas, Metastasis Research Center at MDACC and research funds from the MDACC and Beth Israel Deaconess Medical Center for the Division of Matrix Biology. The 2008 summer sabbatical of MS was funded by SSMF and the Swedish Society for Medicine.

Conflict of Interests: None

Author Contributions: VSL, ST, JD and MS reviewed the data, figures, and revised the manuscript. BAM, HS, LX, LIZ, and JT performed or supervised experiments. JLA performed in silico homology analysis. VSL, ST, JD, BAM prepared the figures. RK oversaw the scientific design of the project and edited the manuscript.

Table 1. Homology Between Zebrafish and Human Collagen IV NC1 domain. Bold: homologs; underlined: highest percent homology.

		Homo sapiens					
	Accession number	*Col4α1*	*Col4α2*	*Col4α3*	*Col4α4*	*Col4α5*	*Col4α6*
Danio rerio	AAY43817.1	**87%**	64%	69%	58%	84%	62%
	AAX55917.1	60%	**82%**	56%	66%	62%	78%
	AAX55916.1	61%	56%	**61%**	52%	63%	55%
	AAY43816.1	52%	58%	48%	**56%**	53%	58%
	AAY43815.1	82%	61%	69%	56%	**84%**	63%
	AAY43818.2	58%	75%	53%	66%	60%	**77%**

Table 2. Homology Between Zebrafish and Human Collagen IV Protein. Bold: homologs; underlined: highest percent homology.

		Homo sapiens					
		Col4α1	*Col4α2*	*Col4α3*	*Col4α4*	*Col4α5*	*Col4α6*
Danio rerio	*Col4α1*	**64%**	44%	49%	43%	60%	45%
	Col4α2	43%	**46%**	40%	44%	42%	47%
	Col4α3	42%	37%	**41%**	38%	46%	38%
	Col4α4	41%	42%	40%	**45%**	40%	45%
	Col4α5	28%	31%	35%	55%	**39%**	28%
	Col4α6	36%	36%	31%	42%	38%	**51%**

Figure Legends

Figure 1: Sequence comparison of the human, mouse, and zebrafish NC1 domain. Comparison of amino acid sequence shows conservation among different species, and the α3 and α4 NC1 domains show higher variation. Subdomains of interests are boxed, structural motifs SM1 (hairpin β3–β4), SM2 (hairpin β6–β7), SM1' (hairpin β3'–β4'), SM3 (β9 and preceding loop), CI (chloride ion). Arrow points to amino acid residue similar in zCol4a4 and hCol4a1 and hCol4a2 NC1 domain. Hs, Homo sapiens (human); Mm Mus musculus (mouse), Dr, Dano rerio (zebrafish).

Figure 2. Structure of the NC1 domain of zebrafish collagen IV α-chain proteins.

a. Sequence comparison of the zebrafish α1-α6 NC1 domains. Cyan: HSQ tripeptide that begin each NC1. Green: twelve conserved cysteine residues. Magenta: additional cysteine (Cysteine 13, Cys-13) in zCol4α4. Red: residues involved in salt bridges formed between the R76 D78, E175-R179, and N187. The red arrow points to the R179K substitution in zCol4α2. Yellow: M93 and K211 residues involved in the peroxidasin-catalyzed sulfilimine bond formation. **b.** Representation of the 3D molecular remodeling of human, mouse and zebrafish NC1 domain β-hairpin SM2 subdomain and SM3' subdomain showing a high level of conservation between the α1, α2 and α5 chains, and less conservation between the α3, α4 and α6 chains. The additional cysteine (C13) in NC1 β-hairpin of zCol4α4 is indicated by the arrow. The asterisk points to the area where the zebrafish protein backbone is shorter than the human and mouse homologues.

Figure 3. Embryonic expression of the six collagen IV α chains in zebrafish.

a. RT-PCR products for the zebrafish type IV collagen α chains at the indicated developmental stages (top band) range in size from 666 bp to 704 bp. The β-actin control band is shown below. Three expression patterns are noted: developmentally unrestricted (α1 and α2 chains); post-gastrulation (α5 and α6 chains); and late-onset expression (α3 and α4 chains). Expression before 6 hours post fertilization (hpf) stage is

maternal. **b**. In situ hybridization for zebrafish α1 and α2 chains at 26 hpf and 36 hpf. **c**. In situ hybridization for zebrafish α3 and α4 chains at 26 hpf, dorsal and lateral views. The α3 and α4 chain are localized in the developing brain (b), with α4 chain also detected in the otic vesicle (o). hpf, hours post fertilization, MBT, mid-blastula transition.

Figure 4. Tissue distribution of the six collagen IV α chains in zebrafish development

a-b. In situ hybridization for zebrafish α5 and α6 chains at 26 hpf and 36 hpf. Zebrafish α5 chain is expressed in the developing brain (b) at 26 hpf and otic vesicle (o) and gills (g) at 36 hpf. The distribution of α6 chain overlaps with that of the α5 chain, with additional strong expression in the developing cardiac tissue (c) at 36 hpf. At the 36 hpf stage, α5 chain was detected in the pronephric duct (pn) of the future kidney, the gills, and the otic vesicles. Dorsal and two lateral views depicting α6 chain in the gills, the pronephric duct, and cardiac tissue (c). **c**. RT-PCR products for the six α chains are shown above the positive control β-actin. All six α chains are detectable in brain, gills, kidney, skeletal muscle, swim bladder and testes. Gut/intestine and skin express the α1, α2, α5, and α6 chains only. Liver and heart both display expression of all α chains except the α3 chain.

Figure 5. Recombinant zTumstain, the NC1 domain of the α3 chain of type IV collagen, shows conserved antiangiogenic activity on human endothelial cells.

a. Schematic depiction of Col4 α3/α4/α5 protomer and cleavage site releasing Tumstatin (α3 NC1 domain). **b**. Relative HUVECs proliferation when treated with the indicated concentration of eukaryotically-produced zTumstatin, n = 3 wells per condition. Positive control: 10% FBS, Negative control: 0.1% FBS. **c**. HUVECs migration in a Boyden-chamber assay, n = 4-6 wells per condition. **d**. Representative macroscopic image of control and zTumstatin Matrigel™ plug (encircled) and blood vessels (arrows). **e**. Representative H&E section of control and zTumstatin Matrigel™ plug (arrows point to blood vessels) and

quantification of averaged blood vessels per 200x field per plug, n = 8 plugs per group (2 plug per mouse, 4 mice per group). b-c: One-way ANOVA, e: unpaired two-tailed t test, *ns*: not significant, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$.

Supplementary Figure Legends

Supplementary Figure 1: Chromosome mapping for α chains between human, mouse, and zebrafish.

The locus of each human (a), mouse (b) and zebrafish (c) collagen IV α chain is mapped to chromosome in the graphic scheme.

Supplementary Figure 2: Sequence comparison of the human, mouse, and zebrafish NC1 domain.

Comparison of amino acid sequence shows conservation among different species, and the α3 and α4 NC1 domains show higher variation. The amino acid sequence is obtained from UniProt database. The color scheme used for alignments is from the algorithm in Clustal X. Each amino acid is colored when it meets the criteria specific for the residue type. This analysis is conducted by Clustal X and viewed by Jalview.

Supplementary Figure 3: Alignment of human, mouse, and zebrafish SM2 (hairpin β6–β7) and SM3' subdomains.

Amino acid sequence alignment for human, mouse and zebrafish NC1 domain, encircling key residues in the SM2 (hairpin β6–β7) and SM3' subdomains that indicates less conservation between the α3, α4 and α6 chains compared to the α1, α2 and α5 chains.

Supplementary Figure 4: Structural evidence for Zebrafish-specific α1α4α1 oligomer.

a. Cartoon schematic of NC1 heterohexamer of α1 and α2 NC1 chains (PDB 5NAX). Arrow shows viewpoint for b. **b.** Bottom-up view of heterotrimer. Inset shows close-up view of salt bridges that help form the heterotrimer interface. In the α1 and α2 chains, a glutamate (E1592 in α1 and E1633 in α2) forms a salt bridge with a basic residue (R1481 in α1 and K1525 in α2).

References

1. LeBleu VS, MacDonald B, Kalluri R. Structure and function of basement membranes. Maywood NJ. 2007;232(9):1121-9.
2. Kalluri R. Basement membranes: structure, assembly and role in tumour angiogenesis. Nat Rev Cancer. 2003;3(6):422-33. Epub 2003/06/05. doi: 10.1038/nrc1094. PubMed PMID: 12778132.
3. Pozzi A, Yurchenco PD, Iozzo RV. The nature and biology of basement membranes. Matrix Biol. 2017;57-58:1-11. Epub 2017/01/04. doi: 10.1016/j.matbio.2016.12.009. PubMed PMID: 28040522; PubMed Central PMCID: PMCPMC5387862.
4. Borchiellini C, Coulon J, Le Parco Y. The function of type IV collagen during Drosophila embryogenesis. Roux Arch Dev Biol. 1996;205(7-8):468-75. Epub 1996/05/01. doi: 10.1007/bf00377228. PubMed PMID: 28306099.
5. Cosgrove D, Liu S. Collagen IV diseases: A focus on the glomerular basement membrane in Alport syndrome. Matrix Biol. 2017;57-58:45-54. Epub 2016/08/31. doi: 10.1016/j.matbio.2016.08.005. PubMed PMID: 27576055; PubMed Central PMCID: PMCPMC5328792.
6. Yurchenco PD, Amenta PS, Patton BL. Basement membrane assembly, stability and activities observed through a developmental lens. Matrix Biol. 2004;22(7):521-38. Epub 2004/03/05. doi: 10.1016/j.matbio.2003.10.006. PubMed PMID: 14996432.
7. Timpl R, Wiedemann H, Van Delden V, Furthmayr H, Kuhn K. A network model for the organization of type IV collagen molecules in basement membranes. Eur J Biochem. 1981;120(2):203-11.
8. Brown KL, Cummings CF, Vanacore RM, Hudson BG. Building collagen IV smart scaffolds on the outside of cells. Protein Sci. 2017;26(11):2151-61.
9. Fidler AL, Vanacore RM, Chetyrkin SV, Pedchenko VK, Bhave G, Yin VP, et al. A unique covalent bond in basement membrane is a primordial innovation for tissue evolution. Proc Natl Acad Sci U S A. 2014;111(1):331-6. Epub 2013/12/18. doi: 10.1073/pnas.1318499111. PubMed PMID: 24344311; PubMed Central PMCID: PMCPMC3890831.
10. Cummings CF, Pedchenko V, Brown KL, Colon S, Rafi M, Jones-Paris C, et al. Extracellular chloride signals collagen IV network assembly during basement membrane formation. J Cell Biol. 2016;213(4):479-94. Epub 2016/05/25. doi: 10.1083/jcb.201510065. PubMed PMID: 27216258; PubMed Central PMCID: PMCPMC4878091.
11. Vanacore R, Ham AJ, Voehler M, Sanders CR, Conrads TP, Veenstra TD, et al. A sulfilimine bond identified in collagen IV. Science. 2009;325(5945):1230-4. Epub 2009/09/05. doi: 10.1126/science.1176811. PubMed PMID: 19729652; PubMed Central PMCID: PMCPMC2876822.
12. Hudson BG, Reeders ST, Tryggvason K. Type IV collagen: structure, gene organization, and role in human diseases. Molecular basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis. J Biol Chem. 1993;268(35):26033-6. Epub 1993/12/15. PubMed PMID: 8253711.
13. LeBleu V, Sund M, Sugimoto H, Birrane G, Kanasaki K, Finan E, et al. Identification of the NC1 domain of {alpha}3 chain as critical for {alpha}3{alpha}4{alpha}5 type IV collagen network assembly. J Biol Chem. 2010;285(53):41874-85. Epub 2010/09/18. doi: 10.1074/jbc.M110.149534. PubMed PMID: 20847057; PubMed Central PMCID: PMCPMC3009915.
14. Ries A, Engel J, Lustig A, Kuhn K. The function of the NC1 domains in type IV collagen. J Biol Chem. 1995;270(40):23790-4. Epub 1995/10/06. doi: 10.1074/jbc.270.40.23790. PubMed PMID: 7559554.
15. Borza DB, Bondar O, Ninomiya Y, Sado Y, Naito I, Todd P, et al. The NC1 domain of collagen IV encodes a novel network composed of the alpha 1, alpha 2, alpha 5, and alpha 6 chains in smooth muscle basement membranes. J Biol Chem. 2001;276(30):28532-40. Epub 2001/05/29. doi: 10.1074/jbc.M103690200. PubMed PMID: 11375996.
16. Khoshnoodi J, Pedchenko V, Hudson BG. Mammalian collagen IV. Microsc Res Tech. 2008;71(5):357-70. Epub 2008/01/26. doi: 10.1002/jemt.20564. PubMed PMID: 18219669; PubMed Central PMCID: PMCPMC4788096.

17. Hudson BG, Kalluri R, Gunwar S, Noelken ME. Structure and organization of type IV collagen of renal glomerular basement membrane. Contrib Nephrol. 1994;107:163-7. Epub 1994/01/01. PubMed PMID: 8004963.
18. Fidler AL, Darris CE, Chetyrkin SV, Pedchenko VK, Boudko SP, Brown KL, et al. Collagen IV and basement membrane at the evolutionary dawn of metazoan tissues. Elife. 2017;6. Epub 2017/04/19. doi: 10.7554/eLife.24176. PubMed PMID: 28418331; PubMed Central PMCID: PMCPMC5395295.
19. Grau-Bove X, Torruella G, Donachie S, Suga H, Leonard G, Richards TA, et al. Dynamics of genomic innovation in the unicellular ancestry of animals. Elife. 2017;6. Epub 2017/07/21. doi: 10.7554/eLife.26036. PubMed PMID: 28726632; PubMed Central PMCID: PMCPMC5560861.
20. Kuhn K. Basement membrane (type IV) collagen. Matrix Biol. 1995;14(6):439-45. Epub 1995/02/01. doi: 10.1016/0945-053x(95)90001-2. PubMed PMID: 7795882.
21. Poschl E, Schlotzer-Schrehardt U, Brachvogel B, Saito K, Ninomiya Y, Mayer U. Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development. Development. 2004;131(7):1619-28. Epub 2004/03/05. doi: 10.1242/dev.01037. PubMed PMID: 14998921.
22. Gould DB, Phalan FC, Breedveld GJ, van Mil SE, Smith RS, Schimenti JC, et al. Mutations in Col4a1 cause perinatal cerebral hemorrhage and porencephaly. Science. 2005;308(5725):1167-71. Epub 2005/05/21. doi: 10.1126/science.1109418. PubMed PMID: 15905400.
23. Favor J, Gloeckner CJ, Janik D, Klempt M, Neuhauser-Klaus A, Pretsch W, et al. Type IV procollagen missense mutations associated with defects of the eye, vascular stability, the brain, kidney function and embryonic or postnatal viability in the mouse, Mus musculus: an extension of the Col4a1 allelic series and the identification of the first two Col4a2 mutant alleles. Genetics. 2007;175(2):725-36. Epub 2006/12/21. doi: 10.1534/genetics.106.064733. PubMed PMID: 17179069; PubMed Central PMCID: PMCPMC1800636.
24. Pastor-Pareja JC, Xu T. Shaping cells and organs in Drosophila by opposing roles of fat body-secreted Collagen IV and perlecan. Dev Cell. 2011;21(2):245-56. Epub 2011/08/16. doi: 10.1016/j.devcel.2011.06.026. PubMed PMID: 21839919; PubMed Central PMCID: PMCPMC4153364.
25. Sessions AO, Kaushik G, Parker S, Raedschelders K, Bodmer R, Van Eyk JE, et al. Extracellular matrix downregulation in the Drosophila heart preserves contractile function and improves lifespan. Matrix Biol. 2017;62:15-27. Epub 2016/10/30. doi: 10.1016/j.matbio.2016.10.008. PubMed PMID: 27793636; PubMed Central PMCID: PMCPMC5405015.
26. Kelemen-Valkony I, Kiss M, Csiha J, Kiss A, Bircher U, Szidonya J, et al. Drosophila basement membrane collagen col4a1 mutations cause severe myopathy. Matrix Biol. 2012;31(1):29-37. Epub 2011/11/01. doi: 10.1016/j.matbio.2011.09.004. PubMed PMID: 22037604.
27. Jeanne M, Gould DB. Genotype-phenotype correlations in pathology caused by collagen type IV alpha 1 and 2 mutations. Matrix Biol. 2017;57-58:29-44. Epub 2016/10/31. doi: 10.1016/j.matbio.2016.10.003. PubMed PMID: 27794444; PubMed Central PMCID: PMCPMC5328961.
28. Savige J. Alport syndrome: its effects on the glomerular filtration barrier and implications for future treatment. J Physiol. 2014;592(18):4013-23. Epub 2014/08/12. doi: 10.1113/jphysiol.2014.274449. PubMed PMID: 25107927; PubMed Central PMCID: PMCPMC4198011.
29. Gast C, Pengelly RJ, Lyon M, Bunyan DJ, Seaby EG, Graham N, et al. Collagen (COL4A) mutations are the most frequent mutations underlying adult focal segmental glomerulosclerosis. Nephrol Dial Transplant. 2016;31(6):961-70. Epub 2015/09/09. doi: 10.1093/ndt/gfv325. PubMed PMID: 26346198.
30. Gatseva A, Sin YY, Brezzo G, Van Agtmael T. Basement membrane collagens and disease mechanisms. Essays Biochem. 2019. Epub 2019/08/08. doi: 10.1042/EBC20180071. PubMed PMID: 31387942.
31. Iozzo RV, Gubbiotti MA. Extracellular matrix: The driving force of mammalian diseases. Matrix Biol. 2018;71-72:1-9. Epub 2018/04/07. doi: 10.1016/j.matbio.2018.03.023. PubMed PMID: 29625183; PubMed Central PMCID: PMCPMC6146050.

32. Sanes JR, Engvall E, Butkowski R, Hunter DD. Molecular heterogeneity of basal laminae: isoforms of laminin and collagen IV at the neuromuscular junction and elsewhere. J Cell Biol. 1990;111(4):1685-99. Epub 1990/10/01. doi: 10.1083/jcb.111.4.1685. PubMed PMID: 2211832; PubMed Central PMCID: PMCPMC2116223.
33. Frojdman K, Pelliniemi LJ, Virtanen I. Differential distribution of type IV collagen chains in the developing rat testis and ovary. Differentiation. 1998;63(3):125-30. Epub 1998/08/11. doi: 10.1046/j.1432-0436.1998.6330125.x. PubMed PMID: 9697306.
34. Kalluri R, Gattone VH, 2nd, Hudson BG. Identification and localization of type IV collagen chains in the inner ear cochlea. Connect Tissue Res. 1998;37(1-2):143-50. Epub 1998/06/27. PubMed PMID: 9643653.
35. Zehnder AF, Adams JC, Santi PA, Kristiansen AG, Wacharasindhu C, Mann S, et al. Distribution of type IV collagen in the cochlea in Alport syndrome. Arch Otolaryngol Head Neck Surg. 2005;131(11):1007-13. Epub 2005/11/23. doi: 10.1001/archotol.131.11.1007. PubMed PMID: 16301374.
36. Mariyama M, Leinonen A, Mochizuki T, Tryggvason K, Reeders ST. Complete primary structure of the human alpha 3(IV) collagen chain. Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in human tissues. J Biol Chem. 1994;269(37):23013-7. Epub 1994/09/16. PubMed PMID: 8083201.
37. Urabe N, Naito I, Saito K, Yonezawa T, Sado Y, Yoshioka H, et al. Basement membrane type IV collagen molecules in the choroid plexus, pia mater and capillaries in the mouse brain. Arch Histol Cytol. 2002;65(2):133-43. Epub 2002/08/08. PubMed PMID: 12164337.
38. Ninomiya Y, Kagawa M, Iyama K, Naito I, Kishiro Y, Seyer JM, et al. Differential expression of two basement membrane collagen genes, COL4A6 and COL4A5, demonstrated by immunofluorescence staining using peptide-specific monoclonal antibodies. J Cell Biol. 1995;130(5):1219-29. Epub 1995/09/01. doi: 10.1083/jcb.130.5.1219. PubMed PMID: 7657706; PubMed Central PMCID: PMCPMC2120565.
39. Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG. Alport's syndrome, Goodpasture's syndrome, and type IV collagen. N Engl J Med. 2003;348(25):2543-56. Epub 2003/06/20. doi: 10.1056/NEJMra022296. PubMed PMID: 12815141.
40. Pedchenko V, Bondar O, Fogo AB, Vanacore R, Voziyan P, Kitching AR, et al. Molecular architecture of the Goodpasture autoantigen in anti-GBM nephritis. N Engl J Med. 2010;363(4):343-54. Epub 2010/07/28. doi: 10.1056/NEJMoa0910500. PubMed PMID: 20660402; PubMed Central PMCID: PMCPMC4144421.
41. Pedchenko V, Kitching AR, Hudson BG. Goodpasture's autoimmune disease - A collagen IV disorder. Matrix Biol. 2018;71-72:240-9. Epub 2018/05/16. doi: 10.1016/j.matbio.2018.05.004. PubMed PMID: 29763670; PubMed Central PMCID: PMCPMC6482832.
42. Maeshima Y, Sudhakar A, Lively JC, Ueki K, Kharbanda S, Kahn CR, et al. Tumstatin, an endothelial cell-specific inhibitor of protein synthesis. Science. 2002;295(5552):140-3. Epub 2002/01/05. doi: 10.1126/science.1065298. PubMed PMID: 11778052.
43. Hamano Y, Zeisberg M, Sugimoto H, Lively JC, Maeshima Y, Yang C, et al. Physiological levels of tumstatin, a fragment of collagen IV alpha3 chain, are generated by MMP-9 proteolysis and suppress angiogenesis via alphaV beta3 integrin. Cancer Cell. 2003;3(6):589-601. Epub 2003/07/05. PubMed PMID: 12842087; PubMed Central PMCID: PMCPMC2775452.
44. Driever W, Stemple D Fau - Schier A, Schier A Fau - Solnica-Krezel L, Solnica-Krezel L. Zebrafish: genetic tools for studying vertebrate development. Trend Genet. 1994;10(5):152-9.
45. MacDonald BA, Sund M, Grant MA, Pfaff KL, Holthaus K, Zon LI, et al. Zebrafish to humans: evolution of the alpha3-chain of type IV collagen and emergence of the autoimmune epitopes associated with Goodpasture syndrome. Blood. 2006;107(5):1908-15. Epub 2005/10/29. doi: 10.1182/blood-2005-05-1814. PubMed PMID: 16254142; PubMed Central PMCID: PMCPMC1895704.
46. Drummond IA. Kidney development and disease in the zebrafish. J Am Soc Nephrol. 2005;16(2):299-304. Epub 2005/01/14. doi: 10.1681/asn.2004090754. PubMed PMID: 15647335.

47. Morales EE, Wingert RA. Zebrafish as a Model of Kidney Disease. Results Probl Cell Differ. 2017;60:55-75. Epub 2017/04/15. doi: 10.1007/978-3-319-51436-9_3. PubMed PMID: 28409342.
48. Thisse C, Thisse B, Schilling TF, Postlethwait JH. Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos. Development. 1993;119(4):1203-15. Epub 1993/12/01. PubMed PMID: 8306883.
49. Maeshima Y, Yerramalla UL, Dhanabal M, Holthaus KA, Barbashov S, Kharbanda S, et al. Extracellular matrix-derived peptide binds to alpha(v)beta(3) integrin and inhibits angiogenesis. J Biol Chem. 2001;276(34):31959-68. Epub 2001/06/16. doi: 10.1074/jbc.M103024200. PubMed PMID: 11399763.
50. Soder S, Poschl E. The NC1 domain of human collagen IV is necessary to initiate triple helix formation. Biochem Biophys Res Commun. 2004;325(1):276-80. Epub 2004/11/04. doi: 10.1016/j.bbrc.2004.10.034. PubMed PMID: 15522229.
51. Colorado PC, Torre A, Kamphaus G, Maeshima Y, Hopfer H, Takahashi K, et al. Anti-angiogenic cues from vascular basement membrane collagen. Cancer Res. 2000;60(9):2520-6. Epub 2000/05/16. PubMed PMID: 10811134.
52. Marneros AG, Olsen BR. The role of collagen-derived proteolytic fragments in angiogenesis. Matrix Biol. 2001;20(5-6):337-45. Epub 2001/09/22. doi: 10.1016/s0945-053x(01)00151-2. PubMed PMID: 11566268.
53. Brodsky B, Ramshaw JA. The collagen triple-helix structure. Matrix Biol. 1997;15(8-9):545-54. Epub 1997/03/01. doi: 10.1016/s0945-053x(97)90030-5. PubMed PMID: 9138287.
54. Ero-Tolliver IA, Hudson BG, Bhave G. The Ancient Immunoglobulin Domains of Peroxidasin Are Required to Form Sulfilimine Cross-links in Collagen IV. J Biol Chem. 2015;290(35):21741-8. Epub 2015/07/17. doi: 10.1074/jbc.M115.673996. PubMed PMID: 26178375; PubMed Central PMCID: PMCPMC4571896.
55. Khoshnoodi J, Sigmundsson K, Cartailler JP, Bondar O, Sundaramoorthy M, Hudson BG. Mechanism of chain selection in the assembly of collagen IV: a prominent role for the alpha2 chain. J Biol Chem. 2006;281(9):6058-69. Epub 2005/12/24. doi: 10.1074/jbc.M506555200. PubMed PMID: 16373348.
56. Than ME, Henrich S, Huber R, Ries A, Mann K, Kuhn K, et al. The 1.9-A crystal structure of the noncollagenous (NC1) domain of human placenta collagen IV shows stabilization via a novel type of covalent Met-Lys cross-link. Proc Natl Acad Sci U S A. 2002;99(10):6607-12. Epub 2002/05/16. doi: 10.1073/pnas.062183499. PubMed PMID: 12011424; PubMed Central PMCID: PMCPMC124450.
57. Vanacore RM, Friedman DB, Ham AJ, Sundaramoorthy M, Hudson BG. Identification of S-hydroxylysyl-methionine as the covalent cross-link of the noncollagenous (NC1) hexamer of the alpha1alpha1alpha2 collagen IV network: a role for the post-translational modification of lysine 211 to hydroxylysine 211 in hexamer assembly. J Biol Chem. 2005;280(32):29300-10. Epub 2005/06/14. doi: 10.1074/jbc.M502752200. PubMed PMID: 15951440.
58. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9. Epub 2010/04/01. doi: 10.1038/nmeth0410-248. PubMed PMID: 20354512; PubMed Central PMCID: PMCPMC2855889.
59. Casino P, Gozalbo-Rovira R, Rodriguez-Diaz J, Banerjee S, Boutaud A, Rubio V, et al. Structures of collagen IV globular domains: insight into associated pathologies, folding and network assembly. IUCrJ. 2018;5(Pt 6):765-79. Epub 2018/11/18. doi: 10.1107/S2052252518012459. PubMed PMID: 30443360; PubMed Central PMCID: PMCPMC6211539.
60. Carrara N, Weaver M, Piedade WP, Vocking O, Famulski JK. Temporal characterization of optic fissure basement membrane composition suggests nidogen may be an initial target of remodeling. Dev Biol. 2019;452(1):43-54. Epub 2019/04/30. doi: 10.1016/j.ydbio.2019.04.012. PubMed PMID: 31034836.
61. Tanaka K, Iyama K, Kitaoka M, Ninomiya Y, Oohashi T, Sado Y, et al. Differential expression of alpha 1(IV), alpha 2(IV), alpha 5(IV) and alpha 6(IV) collagen chains in the basement membrane of basal cell carcinoma. Histochem J. 1997;29(7):563-70. Epub 1997/07/01. PubMed PMID: 9279559.

62. Kelley PB, Sado Y, Duncan MK. Collagen IV in the developing lens capsule. Matrix Biol. 2002;21(5):415-23. Epub 2002/09/13. doi: 10.1016/s0945-053x(02)00014-8. PubMed PMID: 12225806.
63. Kuroda N, Yoshikawa N, Nakanishi K, Iijima K, Hanioka K, Hayashi Y, et al. Expression of type IV collagen in the developing human kidney. Pediatr Nephrol. 1998;12(7):554-8. Epub 1998/10/07. PubMed PMID: 9761353.
64. Kalluri R, Shield CF, Todd P, Hudson BG, Neilson EG. Isoform switching of type IV collagen is developmentally arrested in X-linked Alport syndrome leading to increased susceptibility of renal basement membranes to endoproteolysis. J Clin Invest. 1997;99(10):2470-8. Epub 1997/05/15. doi: 10.1172/jci119431. PubMed PMID: 9153291; PubMed Central PMCID: PMCPMC508088.
65. Eikesdal HP, Sugimoto H, Birrane G, Maeshima Y, Cooke VG, Kieran M, et al. Identification of amino acids essential for the antiangiogenic activity of tumstatin and its use in combination antitumor activity. Proc Natl Acad Sci U S A. 2008;105(39):15040-5. Epub 2008/09/27. doi: 10.1073/pnas.0807055105. PubMed PMID: 18818312; PubMed Central PMCID: PMCPMC2567489.
66. Takeuchi M, Yamaguchi S, Yonemura S, Kakiguchi K, Sato Y, Higashiyama T, et al. Type IV Collagen Controls the Axogenesis of Cerebellar Granule Cells by Regulating Basement Membrane Integrity in Zebrafish. PLoS Genet. 2015;11(10):e1005587. Epub 2015/10/10. doi: 10.1371/journal.pgen.1005587. PubMed PMID: 26451951; PubMed Central PMCID: PMCPMC4599943.
67. Anazco C, Lopez-Jimenez AJ, Rafi M, Vega-Montoto L, Zhang MZ, Hudson BG, et al. Lysyl Oxidase-like-2 Cross-links Collagen IV of Glomerular Basement Membrane. J Biol Chem. 2016;291(50):25999-6012. Epub 2016/10/23. doi: 10.1074/jbc.M116.738856. PubMed PMID: 27770022; PubMed Central PMCID: PMCPMC5207071.
68. Bai X, Dilworth DJ, Weng YC, Gould DB. Developmental distribution of collagen IV isoforms and relevance to ocular diseases. Matrix Biol. 2009;28(4):194-201. Epub 2009/03/12. doi: 10.1016/j.matbio.2009.02.004. PubMed PMID: 19275937; PubMed Central PMCID: PMCPMC4749355.
69. Westcot SE, Hatzold J, Urban MD, Richetti SK, Skuster KJ, Harm RM, et al. Protein-Trap Insertional Mutagenesis Uncovers New Genes Involved in Zebrafish Skin Development, Including a Neuregulin 2a-Based ErbB Signaling Pathway Required during Median Fin Fold Morphogenesis. PLoS One. 2015;10(6):e0130688. Epub 2015/06/26. doi: 10.1371/journal.pone.0130688. PubMed PMID: 26110643; PubMed Central PMCID: PMCPMC4482254.
70. Foster MH. Basement membranes and autoimmune diseases. Matrix Biol. 2017;57-58:149-68. Epub 2016/08/09. doi: 10.1016/j.matbio.2016.07.008. PubMed PMID: 27496347; PubMed Central PMCID: PMCPMC5290253.
71. Heljasvaara R, Aikio M, Ruotsalainen H, Pihlajaniemi T. Collagen XVIII in tissue homeostasis and dysregulation - Lessons learned from model organisms and human patients. Matrix Biol. 2017;57-58:55-75. Epub 2016/10/26. doi: 10.1016/j.matbio.2016.10.002. PubMed PMID: 27746220.

Figure 1

```
                1450      1460      1470      1480      1490      1500      1510      1520
                          |         |         |SM1      |         |         |SM2      |
A1_Hs GFLVTRHSQTIDDPQCPSGTKILYHGYSLLYVQGNERAHGQDLGTAGSCLRKFSTMPFLFCNINNVCNFASRNDYSYWLS
A1_Mm GFLVTRHSQTTDDPLCPPGTKILYHGYSLLYVQGNERAHGQDLGTAGSCLRKFSTMPFLFCNINNVCNFASRNDYSYWLS
A1_Dr GFLVTRHSQTTDDPLCPPGTKILYHGYSLLYVQGNERAHGQDLGTAGSCLRKFSTMPFLFCNINNVCNFASRNDYSYWLS
A2_Hs GYLLVKHSQTDQEPMCPVGMNKLWSGYSLLYFEGQEKAHNQDLGLAGSCLARFSTMPFLYCNPGDVCYYASRNDKSYWLS
A2_Mm GYLLVKHSQTDQEPMCPVGMNKLWSGYSLLYFEGQEKAHNQDLGLAGSCLARFSTMPFLYCNPGDVCYYASRNDKSYWLS
A2_Dr GYLLVKHSQSEEIPMCPQGMSLLWMGYSLLYFEGQEKAHNQDLGLAGSCLPRFNTMPFLYCNPGDICYYASRNDKSYWLS
A3_Hs GFVFTRHSQTTAIPSCPEGTVPLYSGFSFLFVQGNQRAHGQDLGTLGSCLQRFTTMPFLFCNVNDVCNFASRNDYSYWLS
A3_Mm GFIFTRHSQTTAIPSCPEGTQPLYSGFSLLFVQGNKRAHGQDLGTLGSCLQRFTTMPFLFCNINNVCNFASRNDYSYWLS
A3_Dr GFLFTRHSQTTVIPECPAGSKRLYTGYSLLFINGNNRGHGQDLGTLGSCLPMFNTMPFMVCNRDETCRYASRNDYSYWLS
A4_Hs GFLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQVCHYAQRNDrSYWLA
A4_Mm GFLLVLHSQTDQEPACPVGMPRLWTGYSLLYMEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQVCHYAQRNDrSYWLS
A4_Dr GFLLVMHSQSRYVPTCPAGLTQLWNGYSLLYLEGQERAHTQDLGQAGSCLPVFSTMPFSCCNM-DTCDYASRNDKSYWLS
A5_Hs GFLITRHSQTTDAPQCPQGTLQVYEGFSLLYVQGNKRAHGQDLGTAGSCLRRFSTMPFMFCNINNVCNFASRNDYSYWLS
A5_Mm GFLITRHSQTTEAPQCPRGTVHIYEGFSLLYVQGNKRAHGQDLGTAGSCLRRFSTMPFMFCNINNVCNFASRNDYSYWLS
A5_Dr GFLITRHSQTIEVPSCPDGTSPIYEGYSLLYVQGNERAHGQDLGTAGSCLRRFSTMPFMFCNINNVCNFASRNDYSYWLS
A6_Hs GYTLVKHSQSEQVPPCPIGMSQLWVGYSLLFVEGQEKAHNQDLGFAGSCLPRFSTMPFIYCNINEVCHYARRNDKSYWLS
A6_Mm GYTLVKHSQSEQVPLCPTGMSRLWVGYSLLFVEGQEKAHNQDLGFAGSCLPRFSTMPFIYCNINEVCHYARRNDKSYWLS
A6_Dr GYTLVKHSQSTRVPMCPEGMTRLWEGYSLLYVEGQEKAHNQDLGLPGSCLPHFSTIPFLYCTANEICYYASRNDKSYWLS

                1530      1540      1550      1560      1570      1580      1590      1600
                          |         |         |         |         |         |SM1'     |
A1_Hs TPEPMPMSMAPITGENIRPFISRCAVCEAPAMVMAVHSQTIQIPPCPSGWSSLWIGYSFVMHTSAGAEGSGQALASPGSC
A1_Mm TPEPMPMSMAPISGDNIRPFISRCAVCEAPAMVMAVHSQTIQIPQCPNGWSSLWIGYSFVMHTSAGAEGSGQALASPGSC
A1_Dr TPEPMPMSMAPISGDNIRPFISRCAVCEAPAMVMAVHSQTIQIPQCPNGWSSLWIGYSFVMHTSAGAEGSGQALASPGSC
A2_Hs TTAPLPM--MPVAEDEIKPYISRCSVCEAPAIAIAVHSQDVSIPHCPAGWRSLWIGYSFLMHTAAGDEGGGQSLVSPGSC
A2_Mm TTAPLPM--MPVAEEEIKPYISRCSVCEAPAVAIAVHSQDTSIPHCPAGWRSLWIGYSFLMHTAAGDEGGGQSLVSPGSC
A2_Dr TTQPIPM--MPVEESEIKPYISRCSVCEAPSVAIAVHSQDTTIPNCPVGWRSLWIGYSFLMHTAAGDEGGGQSLASPGSC
A3_Hs TPALMPMNMAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTGQALASPGSC
A3_Mm TPALMPMDMAPISGRALEPYISRCTVCEGPAMAIAVHSQTTAIPPCPQDWVSLWKGFSFIMFTSAGSEGAGQALASPGSC
A3_Dr TDTPMLPDQQMMSGEILKWYISRCSVCEAIANVIAIHSQTINIPQCPVGWLSLWEGYSFVMQTGVGAEGSGQPLVSPGSC
A4_Hs SAAPLPM--MPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSIPPCPQTWRSLWIGYSFLMHTGAGDQGGGQALMSPGSC
A4_Mm SAAPLPM--MPLSEEEIRSYISRCAVCEAPAQAVAVHSQDQSIPPCPRTWRSLWIGYSFLMHTGAGDQGGGQALMSPGSC
A4_Dr TNAPIPN--KPLKGQDIEEHISRCVVCEAPTPTIAIHSQDrLDPVCPPKWRSLWTGFSFMMYTGSGDEGGGQSLTSTGSC
A5_Hs TPEPMPMSMQPLKGQSIQPFISRCAVCEAPAVVIAVHSQTIQIPHCPQGWDSLWIGYSFMMHTSAGAEGSGQALASPGSC
A5_Mm TPEPMPMNMEPLKGQSIQPFISRCAVCEAPAVVIAVHSQTIQIPHCPQGWDSLWIGYSFMMHTSAGAEGSGQALASPGSC
A5_Dr TPQPMPMTMAPITGENIKPYISRCSVCEAPAMVIAVHSQTIQLPACPENWDTLWIGYSFMMHTSAGAEGSGQALASPGSC
A6_Hs TTAPIPM--MPVSQTQIPQYISRCSVCEAPSQAIAVHSQDITIPQCPLGWRSLWIGYSFLMHTAAGAEGGGQSLVSPGSC
A6_Mm TTAPIPM--MPVGQTQIPQYISRCSVCEAPSQAIAVHSQDIIVPQCPLGWHSLWIGYSFLMHTAAGAEGGGQSLVSPGSC
A6_Dr TTAAIPM--MPVANQEISPYISRCSVCEVPSQPIAVHSQDLTIPTCPQGWRSLWIGYSFLMHTGAGGEGGGQSLMSPGSC

                1610      1620      1630      1640      1650      1660
                          |         CI|         |SM3      |         |
A1_Hs LEEFRSAPFIECHG-RGTCNYYANAYSFWLATIERSEMF-KKPTPSTLKAGEL-RTHVSRCQVCMRRT
A1_Mm LEEFRSAPFIECHG-RGTCNYYANAYSFWLATIERSEMF-KKPTPSTLKAGEL-RTHVSRCQVCMRRT
A1_Dr LEEFRSAPFIECHG-RGTCNYYANAYSFWLATIERSEMF-KKPTPSTLKAGEL-RTHVSRCQVCMRRT
A2_Hs LEDFRATPFIECNGGRGTCHYYANKYSFWLTTIPE-QSFQGSPSADTLKAGLI-RTHISRCQVCMKNL
A2_Mm LEDFRATPFIECNGGRGTCHYFANKYSFWLTTIPE-QNFQSTPSADTLKAGLI-RTHISRCQVCMKNL
A2_Dr LEDFRTTPFIECNGAKGTCHYFANKHSFWLTSID--ESFQSSPSSETLKAGQL-LSRISRCQVCMKNL
A3_Hs LEEFRASPFLECHG-RGTCNYYSNSYSFWLASLNPERMF-RKPIPSTVKAGEL-EKIISRCQVCMKKRH
A3_Mm LEEFRASPFIECHG-RGTCNYYSNSYSFWLASLNPERMF-RKPIPSTVKAGDL-EKIISRCQVCMKKR
A3_Dr LEQFRKIPFIECHG-RGTCNFYPDSYSYWLASLDHTNMF-SMPNRQTAKQ----KEIISRCQVCMKQQ
A4_Hs LEDFRAAPFLECQGRQGTCHFFANKYSFWLTTVKADLQFSSAPAPDTLKESQAQRQKISRCQVCVKYS
A4_Mm LEDFRAAPFVECQGRQGTCHFFANEYSFWLTTVNPDLQFASGPSPDTLKEVQAQRRKISRCQVCMKHS
A4_Dr LQDFRSQPFVECQGPRGTCSYFASIYSFWMTIDMEH--NDSSPHGPVITEERQQRDSTSRCSICMMND
A5_Hs LEEFRSAPFIECHG-RGTCNYYANSYSFWLATVDVSDMF-SKPQSETLKAGDL-RTRISRCQVCMKRT
A5_Mm LEEFRSAPFIECHG-RGTCNYYANSYSFWLATVDMSDMF-NKPQSETLKAGDL-RTRISRCQVCMKRT
A5_Dr LEEFRSAPFIECHG-RGTCNYYGNSYSFWLATVDVNEMF-RKPQSETLKAGNL-RTRVSRCVVCMKRT
A6_Hs LEDFRATPFIECSGARGTCHYFANKYSFWLTTVEERQQFGELPVSETLKAGQL-HTRVSRCQVCMKSL
A6_Mm LEDFRATPFIECSGARGTCHYFANKYSFWLTTVEERGQFREEPVSETLKTGQL-HTRVSRCQVCMKTP
A6_Dr LEDFRATPFIECNGARGTCHYFANKYSFWLTTVDVNEQFVQGPTQETLKGDQF-RSRLSRCQVCMKIL
```

Hs Human
Mm Mouse
Dr Zebrafish

Figure 2

a

zCol4α1
GFLVTRHSQTTDDPLCPPGTKILYHGYSLLYVQGNERAHGQDLGTAGSCLRKFSTMPFLFCNINNVCNFASRNDYSYWLST
PEPMPMSMAPISGDNIRPFISRCAVCEAPAMVMAVHSQTIQIPQCPNGWSSLWIGYSFVMHTSAGAEGSGQALASPGSCLE
EFRSAPFIECHGRGTCNYYANAYSFWLATIERSEMFKKPTPSTLKAGELRTHVSRCQVCMRRT

zCol4α2
GYLLVKHSQSEEIPMCPQGMSLLWMGYSLLYFEGQEKAHNQDLGLAGSCLPRFNTMPFLYCNPGDICYYASRNDKSYWLS
TTQPIPMMPVEESEIKPYISRCSVCEAPSVAIAVHSQDTTIPNCPVGWRSLWIGYSFLMHTAAGDEGGGQSLASPGSCLEDF
RTTPFIECNGAKGTCHYFANKHSFWLTSIDESFQSSPSSETLKAGQLLSRISRCQVCMKNL

zCol4α3
GFLFTRHSQTTVIPECPAGSKRLYTGYSLLFINGNNRGHGQDLGTLGSCLPMFNTMPFMVCNRDETCRYASRNDYSYWLST
DTPMLPDQQMMSGEILKWYISRCSVCEAIANVIAIHSQTINIPQCPVGWLSLWEGYSFVMQTGVGAEGSGQPLVSPGSCLE
QFRKIPFIECHGRGTCNFYPDSYSYWLASLDHTNMFSMPNRQTAKQKEIISRCQVCMKQQ

zCol4α4
GFLLVMHSQSRYVPTCPAGLTQLWNGYSLLYLEGQERAHTQDLGQAGSCLPVFSTMPFSCCNMDTCDYASRNDKSYWLS
TNAPIPNKPLKGQDIEEHISRCVVCEAPTPTIAIHSQDRLDPVCPPKWRSLWTGFSFMMYTGSGDEGGGQSLTSTGSCLQD
FRSQPFVECQGPRGTCSYFASIYSFWMTIDMEHNDSSPHGPVITEERQQRDSTSRCSICMMND

zCol4α5
GFLITRHSQTIEVPSCPDGTSPIYEGYSLLYVQGNERAHGQDLGTAGSCLRRFSTMPFMFCNINNVCNFASRNDYSYWLSTP
QPMPMTMAPITGENIKPYISRCSVCEAPAMVIAVHSQTIQLPACPENWDTLWIGYSFMMHTSAGAEGSGQALASPGSCLEE
FRSAPFIECHGRGTCNYYGNSYSFWLATVDVNEMFRKPQSETLKAGNLRTRVSRCVVCMKRT

zCol4α6
GYTLVKHSQSTRVPMCPEGMTRLWEGYSLLYVEGQEKAHNQDLGLPGSCLPHFSTIPFLYCTANEICYYASRNDKSYWLST
TAAIPMMPVANQEISPYISRCSVCEVPSQPIAVHSQDLTIPTCPQGWRSLWIGYSFLMHTGAGGEGGGQSLMSPGSCLEDF
RATPFIECNGARGTCHYFANKYSFWLTTVDVNEQFVQGPTQETLKGDQFRSRLSRCQVCMKIL

b

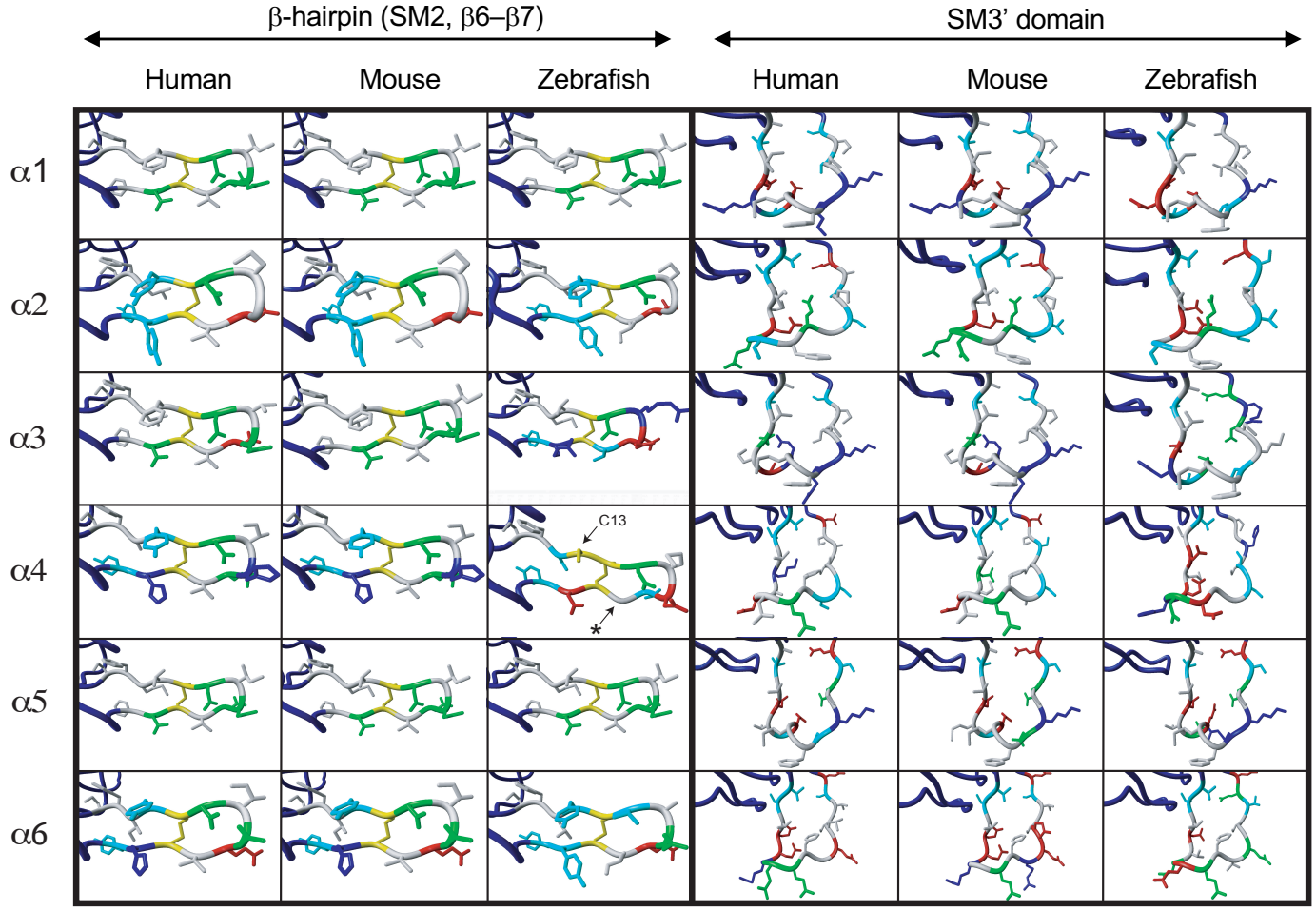
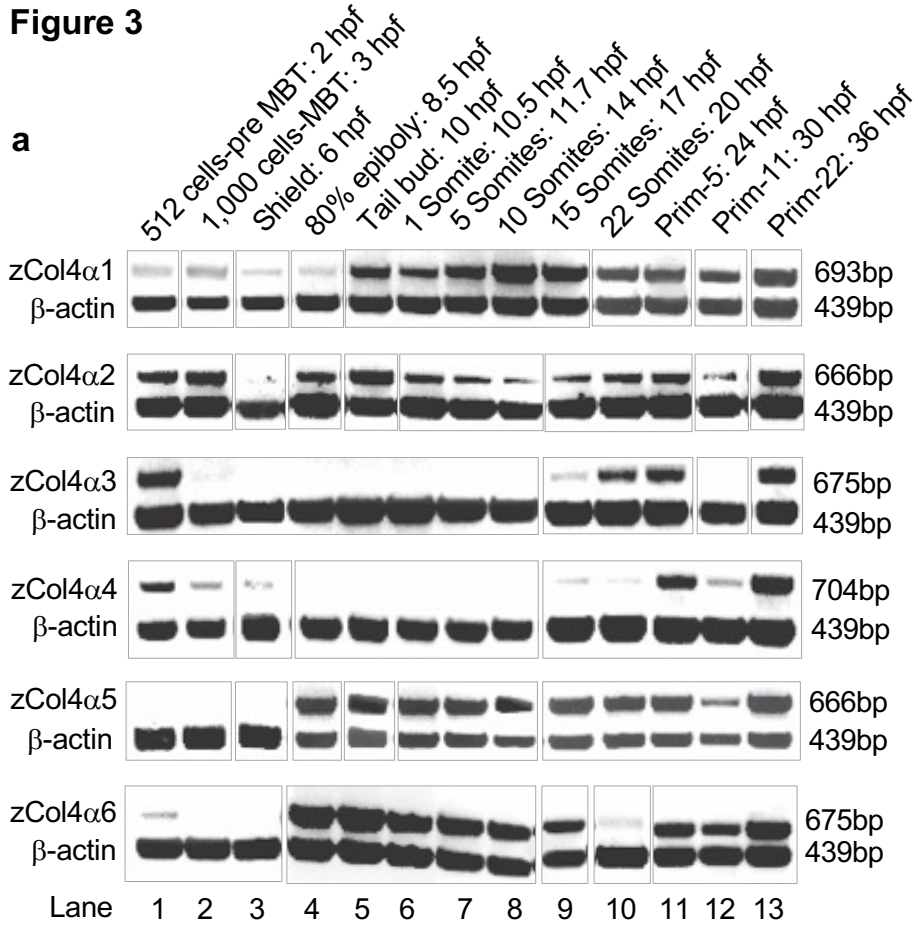
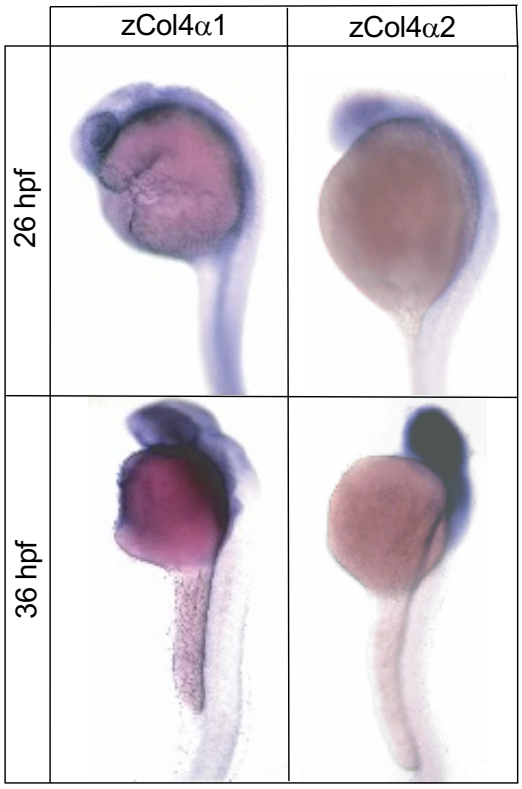


Figure 3

a


512 cells-pre MBT: 2 hpf
1,000 cells-MBT: 3 hpf
Shield: 6 hpf
80% epiboly: 8.5 hpf
Tail bud: 10 hpf
1 Somite: 10.5 hpf
5 Somites: 11.7 hpf
10 Somites: 14 hpf
15 Somites: 17 hpf
22 Somites: 20 hpf
Prim-5: 24 hpf
Prim-11: 30 hpf
Prim-22: 36 hpf
zCol4α1
β-actin
693bp
439bp
zCol4α2
β-actin
666bp
439bp
zCol4α3
β-actin
675bp
439bp
zCol4α4
β-actin
704bp
439bp
zCol4α5
β-actin
666bp
439bp
zCol4α6
β-actin
675bp
439bp
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13

b


zCol4α1
zCol4α2
26 hpf
36 hpf

c

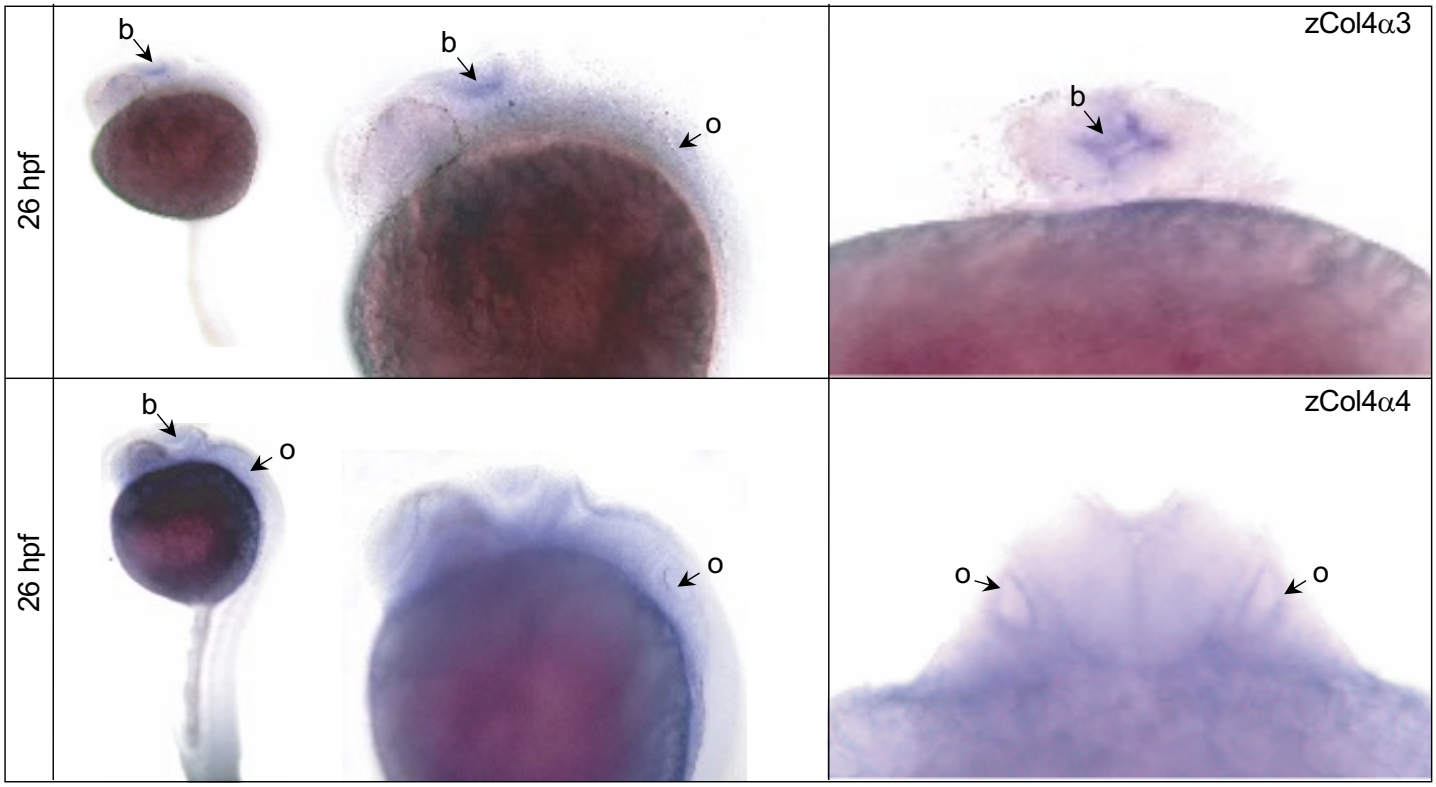

26 hpf
b
b
o
zCol4α3
b
26 hpf
b
o
o
zCol4α4
o
o

Figure 4

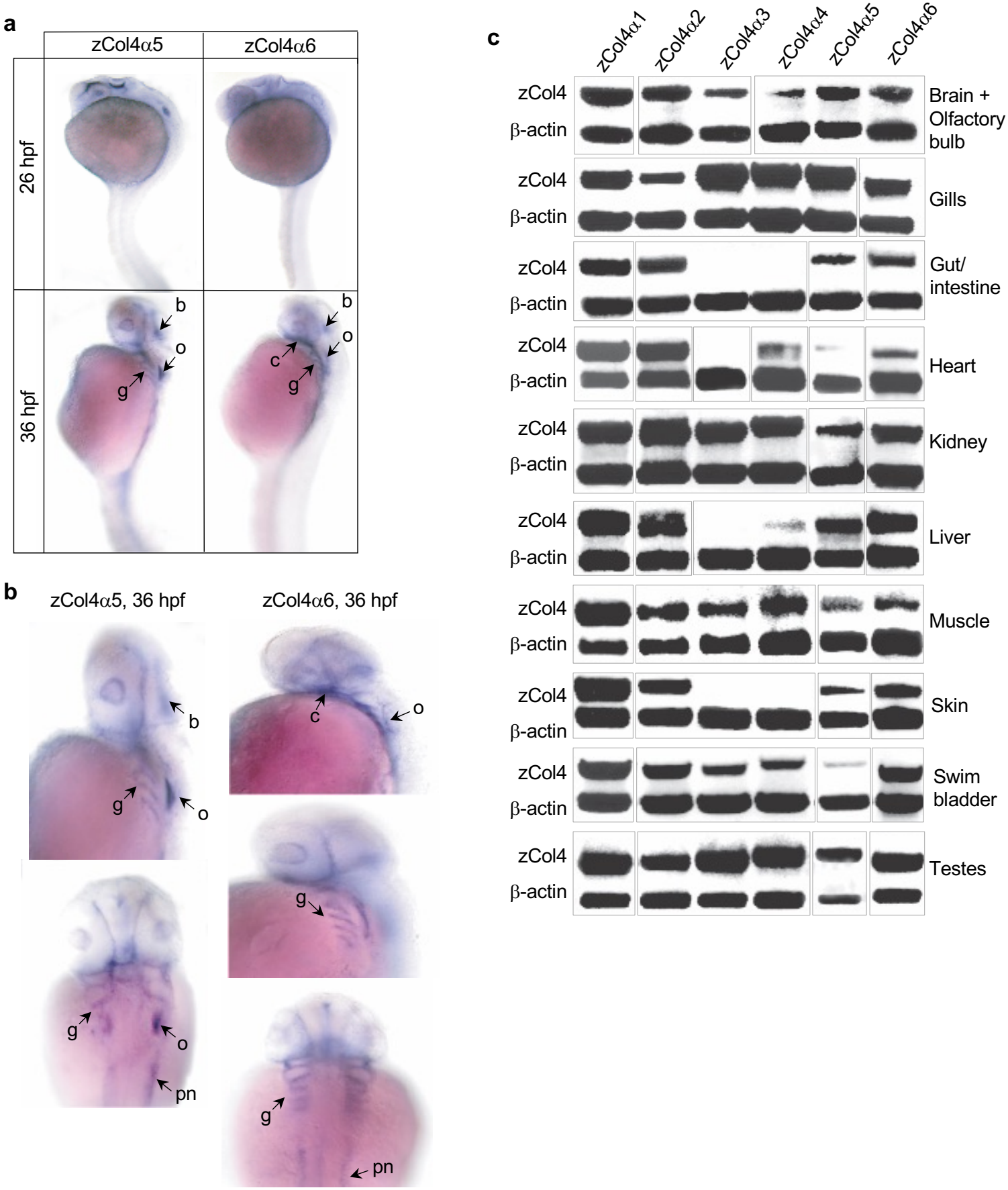
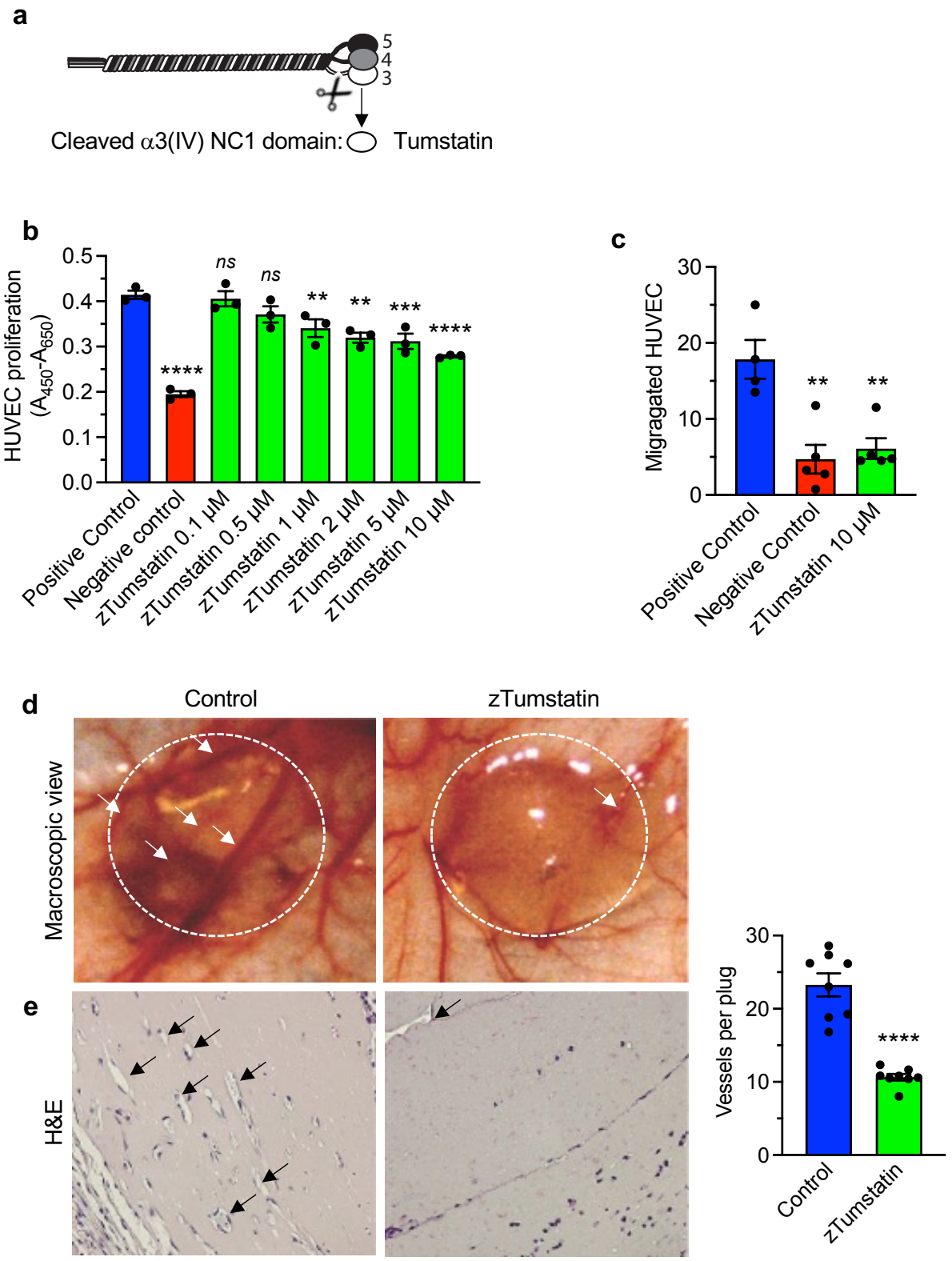
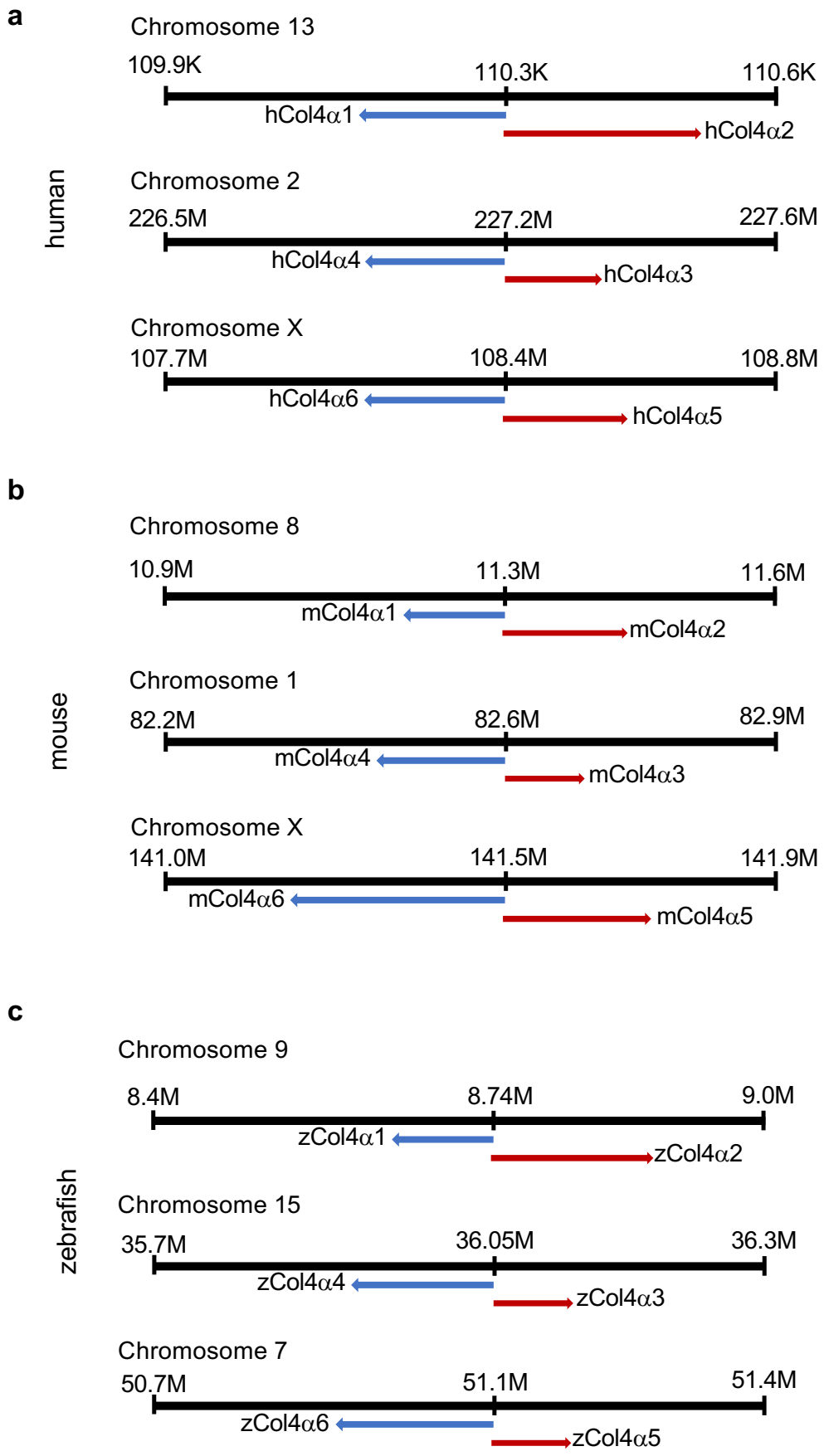

a
zCol4α5
zCol4α6
26 hpf
36 hpf
b
o
g
c
b
zCol4α5, 36 hpf
zCol4α6, 36 hpf
pn
c
zCol4α1
zCol4α2
zCol4α3
zCol4α4
zCol4α5
zCol4α6
zCol4
β-actin
Brain + Olfactory bulb
Gills
Gut/ intestine
Heart
Kidney
Liver
Muscle
Skin
Swim bladder
Testes

Figure 5


a
5
4
3
Cleaved α3(IV) NC1 domain:
Tumstatin
b
HUVEC proliferation
(A450-A650)
0.5
0.4
0.3
0.2
0.1
0.0
ns
ns
**
**
***
****
****
Positive Control
Negative control
zTumstatin 0.1 μM
zTumstatin 0.5 μM
zTumstatin 1 μM
zTumstatin 2 μM
zTumstatin 5 μM
zTumstatin 10 μM
c
Migragated HUVEC
30
20
10
0
**
**
Positive Control
Negative Control
zTumstatin 10 μM
d
Control
zTumstatin
Macroscopic view
e
H&E
Vessels per plug
30
20
10
0
****
Control
zTumstatin

Supplementary Figure 1



Supplementary Figure 2

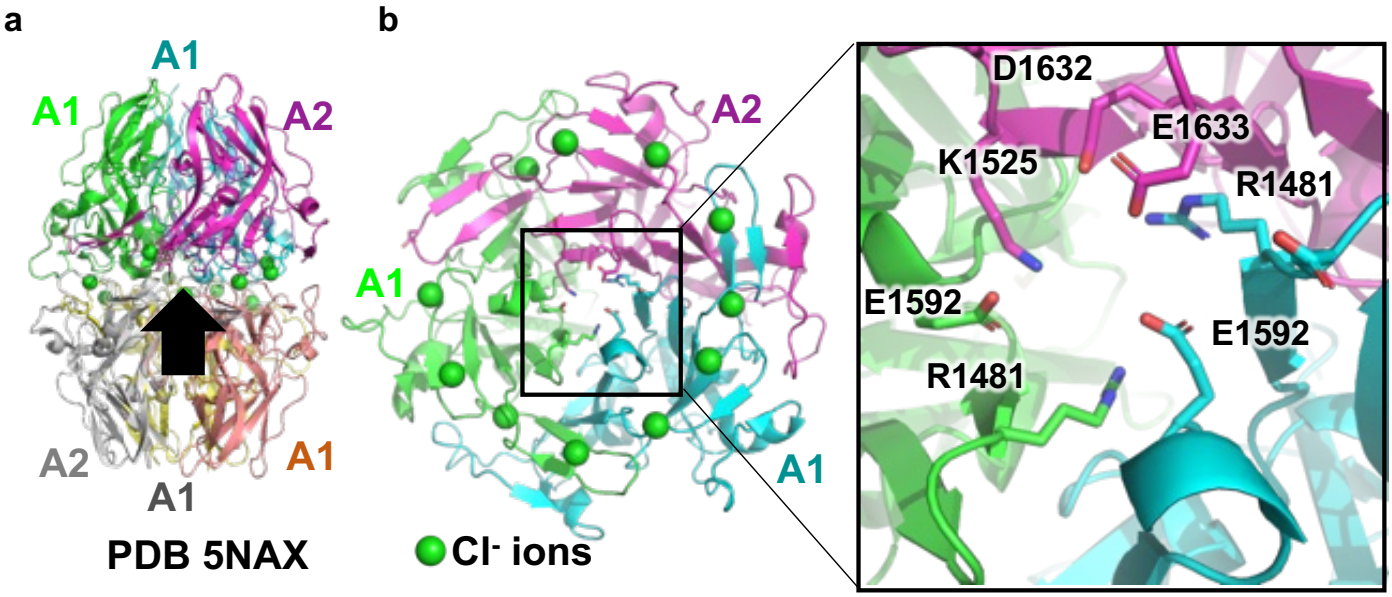
```
hCol4α1_NC1 GFLVTRHSQTIDDPQCPSGTKILYHGYSLLYVQGNERAHGQDLGTAGSCLRKFSTMPFLFCNINNVCNFASRNDYSYW  78
mCol4α1_NC1 GFLVTRHSQTTDDPLCPPGTKILYHGYSLLYVQGNERAHGQDLGTAGSCLRKFSTMPFLFCNINNVCNFASRNDYSYW  78
zCol4α1_NC1 GFLVTRHSQTTDDPLCPPGTKILYHGYSLLYVQGNERAHGQDLGTAGSCLRKFSTMPFLFCNINNVCNFASRNDYSYW  78

hCol4α1_NC1 LSTPEPMPMSMAPITGENIRPFISRCAVCEAPAMVMAVHSQTIQIPPCPSGWSSLWIGYSFVMHTSAGAEGSGQALAS 156
mCol4α1_NC1 LSTPEPMPMSMAPISGDNIRPFISRCAVCEAPAMVMAVHSQTIQIPQCPNGWSSLWIGYSFVMHTSAGAEGSGQALAS 156
zCol4α1_NC1 LSTPEPMPMSMAPISGDNIRPFISRCAVCEAPAMVMAVHSQTIQIPQCPNGWSSLWIGYSFVMHTSAGAEGSGQALAS 156

hCol4α1_NC1 PGSCLEEFRSAPFIECHGRGTCNYYANAYSFWLATIERSEMFKKPTPSTLKAGELRTHVSRCQVCMRRT          225
mCol4α1_NC1 PGSCLEEFRSAPFIECHGRGTCNYYANAYSFWLATIERSEMFKKPTPSTLKAGELRTHVSRCQVCMRRT          225
zCol4α1_NC1 PGSCLEEFRSAPFIECHGRGTCNYYANAYSFWLATIERSEMFKKPTPSTLKAGELRTHVSRCQVCMRRT          225

hCol4α2_NC1 GYLLVKHSQTDQEPMCPVGMNKLWSGYSLLYFEGQEKAHNQDLGLAGSCLARFSTMPFLYCNPGDVCYYASRNDKSYW  78
mCol4α2_NC1 GYLLVKHSQTDQEPMCPVGMNKLWSGYSLLYFEGQEKAHNQDLGLAGSCLARFSTMPFLYCNPGDVCYYASRNDKSYW  78
zCol4α2_NC1 GYLLVKHSQSEEIPMCPQGMSLLWMGYSLLYFEGQEKAHNQDLGLAGSCLPRFNTMPFLYCNPGDICYYASRNDKSYW  78

hCol4α2_NC1 LSTTAPLPMMPVAEDEIKPYISRCSVCEAPAIAIAVHSQDVSIPHCPAGWRSLWIGYSFLMHTAAGDEGGGQSLVSPG 156
mCol4α2_NC1 LSTTAPLPMMPVAEEEIKPYISRCSVCEAPAVAIAVHSQDTSIPHCPAGWRSLWIGYSFLMHTAAGDEGGGQSLVSPG 156
zCol4α2_NC1 LSTTQPIPMMPVEESEIKPYISRCSVCEAPSVAIAVHSQDTTIPNCPVGWRSLWIGYSFLMHTAAGDEGGGQSLASPG 156

hCol4α2_NC1 SCLEDFRATPFIECNGGRGTCHYYANKYSFWLTTIPEQSFQGSPSADTLKAGLIRTHISRCQVCMKNL          224
mCol4α2_NC1 SCLEDFRATPFIECNGGRGTCHYFANKYSFWLTTIPEQNFQSTPSADTLKAGLIRTHISRCQVCMKNL          224
zCol4α2_NC1 SCLEDFRTTPFIECNGAKGTCHYFANKHSFWLTSIDE-SFQSSPSSETLKAGQLLSRISRCQVCMKNL          223

hCol4α3_NC1 PGLKGKRGDSGSPATWTTRGFVFTRHSQTTAIPSCPEGTVPLYSGFSFLFVQGNQRAHGQDLGTLGSCLQRFTTMPFL  78
mCol4α3_NC1 ------------------GFIFTRHSQTTAIPSCPEGTQPLYSGFSLLFVQGNKRAHGQDLGTLGSCLQRFTTMPFL  59
zCol4α3_NC1 ------------------GFLFTRHSQTTVIPECPAGSKRLYTGYSLLFINGNNRGHGQDLGTLGSCLPMFNTMPFM  59

hCol4α3_NC1 FCNVNDVCNFASRNDYSYWLSTPALMPMNMAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGF 156
mCol4α3_NC1 FCNINNVCNFASRNDYSYWLSTPALMPMDMAPISGRALEPYISRCTVCEGPAMAIAVHSQTTAIPPCPQDWVSLWKGF 137
zCol4α3_NC1 VCNRDETCRYASRNDYSYWLSTDTPMLPDQQMMSGEILKWYISRCSVCEAIANVIAIHSQTINIPQCPVGWLSLWEGY 137

hCol4α3_NC1 SFIMFTSAGSEGTGQALASPGSCLEEFRASPFLECHGRGTCNYYSNSYSFWLASLNPERMFRKPIPSTVKAGELEKII 234
mCol4α3_NC1 SFIMFTSAGSEGAGQALASPGSCLEEFRASPFIECHGRGTCNYYSNSYSFWLASLNPERMFRKPIPSTVKAGDLEKII 215
zCol4α3_NC1 SFVMQTGVGAEGSGQPLVSPGSCLEQFRKIPFIECHGRGTCNFYPDSYSYWLASLDHTNMFSMPNRQTAKQ---KEII 212

hCol4α3_NC1 SRCQVCMKKRH                                                                    245
mCol4α3_NC1 SRCQVCMKKR-                                                                    225
zCol4α3_NC1 SRCQVCMKQQ-                                                                    222

hCol4α4_NC1 GFLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQVCHYAQRNDRSYW  78
mCol4α4_NC1 GFLLVLHSQTDQEPACPVGMPRLWTGYSLLYMEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQVCHYAQRNDRSYW  78
zCol4α4_NC1 GFLLVMHSQSRYVPTCPAGLTQLWNGYSLLYLEGQERAHTQDLGQAGSCLPVFSTMPFSCCNMD-TCDYASRNDKSYW  77

hCol4α4_NC1 LASAAPLPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSIPPCPQTWRSLWIGYSFLMHTGAGDQGGGQALMSPG 156
mCol4α4_NC1 LSSAAPLPMMPLSEEEIRSYISRCAVCEAPAQAVAVHSQDQSIPPCPRTWRSLWIGYSFLMHTGAGDQGGGQALMSPG 156
zCol4α4_NC1 LSTNAPIPNKPLKGQDIEEHISRCVVCEAPTPTIAIHSQDRLDPVCPPKWRSLWTGFSFMMYTGSGDEGGGQSLTSTG 155

hCol4α4_NC1 SCLEDFRAAPFLECQGRQGTCHFFANKYSFWLTTVKADLQFSSAPAPDTLKESQAQRQKISRCQVCVKYS          226
mCol4α4_NC1 SCLEDFRAAPFVECQGRQGTCHFFANEYSFWLTTVNPDLQFASGPSPDTLKEVQAQRRKISRCQVCMKHS          226
zCol4α4_NC1 SCLQDFRSQPFVECQGPRGTCSYFASIYSFWMT-IDMEHNDSSPHGP-VITEERQQRDSTSRCSICMMND          223

hCol4α5_NC1 GFLITRHSQTTDAPQCPQGTLQVYEGFSLLYVQGNKRAHGQDLGTAGSCLRRFSTMPFMFCNINNVCNFASRNDYSYW  78
mCol4α5_NC1 GFLITRHSQTTEAPQCPRGTVHIYEGFSLLYVQGNKRAHGQDLGTAGSCLRRFSTMPFMFCNINNVCNFASRNDYSYW  78
zCol4α5_NC1 GFLITRHSQTIEVPSCPDGTSPIYEGYSLLYVQGNERAHGQDLGTAGSCLRRFSTMPFMFCNINNVCNFASRNDYSYW  78

hCol4α5_NC1 LSTPEPMPMSMQPLKGQSIQPFISRCAVCEAPAVVIAVHSQTIQIPHCPQGWDSLWIGYSFMMHTSAGAEGSGQALAS 156
mCol4α5_NC1 LSTPEPMPMNMEPLKGQSIQPFISRCAVCEAPAVVIAVHSQTIQIPHCPQGWDSLWIGYSFMMHTSAGAEGSGQALAS 156
zCol4α5_NC1 LSTPQPMPMTMAPITGENIKPYISRCSVCEAPAMVIAVHSQTIQLPACPENWDTLWIGYSFMMHTSAGAEGSGQALAS 156

hCol4α5_NC1 PGSCLEEFRSAPFIECHGRGTCNYYANSYSFWLATVDVSDMFSKPQSETLKAGDLRTRISRCQVCMKRT          225
mCol4α5_NC1 PGSCLEEFRSAPFIECHGRGTCNYYANSYSFWLATVDMSDMFNKPQSETLKAGDLRTRISRCQVCMKRT          225
zCol4α5_NC1 PGSCLEEFRSAPFIECHGRGTCNYYGNSYSFWLATVDVNEMFRKPQSETLKAGNLRTRVSRCVVCMKRT          225

hCol4α6_NC1 GYTLVKHSQSEQVPPCPIGMSQLWVGYSLLFVEGQEKAHNQDLGFAGSCLPRFSTMPFIYCNINEVCHYARRNDKSYW  78
mCol4α6_NC1 GYTLVKHSQSEQVPLCPTGMSRLWVGYSLLFVEGQEKAHNQDLGFAGSCLPRFSTMPFIYCNINEVCHYARRNDKSYW  78
zCol4α6_NC1 GYTLVKHSQSTRVPMCPEGMTRLWEGYSLLYVEGQEKAHNQDLGLPGSCLPHFSTIPFLYCTANEICYYASRNDKSYW  78

hCol4α6_NC1 LSTTAPIPMMPVSQTQIPQYISRCSVCEAPSQAIAVHSQDITIPQCPLGWRSLWIGYSFLMHTAAGAEGGGQSLVSPG 156
mCol4α6_NC1 LSTTAPIPMMPVGQTQIPQYISRCSVCEAPSQAIAVHSQDIIVPQCPLGWHSLWIGYSFLMHTAAGAEGGGQSLVSPG 156
zCol4α6_NC1 LSTTAAIPMMPVANQEISPYISRCSVCEVPSQPIAVHSQDLTIPTCPQGWRSLWIGYSFLMHTGAGGEGGGQSLMSPG 156

hCol4α6_NC1 SCLEDFRATPFIECSGARGTCHYFANKYSFWLTTVEERQQFGELPVSETLKAGQLHTRVSRCQVCMKSL          225
mCol4α6_NC1 SCLEDFRATPFIECSGARGTCHYFANKYSFWLTTVEERGQFREEPVSETLKTGQLHTRVSRCQVCMKTP          225
zCol4α6_NC1 SCLEDFRATPFIECNGARGTCHYFANKYSFWLTTVDVNEQFVQGPTQETLKGDQFRSRLSRCQVCMKIL          225
```

Supplementary Figure 3

a β-hairpin domain (SM2, β6–β7)

		1	2	3	4	5	6	7	8	9	10	11	12	13
Human	α1	P	F	L	F	C	N	I	N	N	V	C	N	F
Mouse	α1	P	F	L	F	C	N	I	N	N	V	C	N	F
Zebrafish	α1	P	F	L	F	C	N	I	N	N	V	C	N	F
Human	α3	P	F	L	F	C	N	V	N	D	V	C	N	F
Mouse	α3	P	F	L	F	C	N	I	N	N	V	C	N	F
Zebrafish	α3	P	F	M	V	C	N	R	D	E	T	C	R	Y
Human	α5	P	F	M	F	C	N	I	N	N	V	C	N	F
Mouse	α5	P	F	M	F	C	N	I	N	N	V	C	N	F
Zebrafish	α5	P	F	M	F	C	N	I	N	N	V	C	N	F
Human	α2	P	F	L	Y	C	N	P	G	D	V	C	Y	Y
Mouse	α2	P	F	L	Y	C	N	P	G	D	V	C	Y	Y
Zebrafish	α2	P	F	L	Y	C	N	P	G	D	I	C	Y	Y
Human	α4	P	F	A	Y	C	N	I	H	Q	V	C	H	Y
Mouse	α4	P	F	A	Y	C	N	I	H	Q	V	C	H	Y
Zebrafish	α4	P	F	S	C	C	N	M	D	T	-	C	D	Y
Human	α6	P	F	I	Y	C	N	I	N	E	V	C	H	Y
Mouse	α6	P	F	I	Y	C	N	I	N	E	V	C	H	Y
Zebrafish	α6	P	F	L	Y	C	T	A	N	E	I	C	Y	Y

b SM3' domain

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Human	α1	A	T	I	E	R	S	E	M	F	K	K	-	P	T	P	S
Mouse	α1	A	T	I	E	R	S	E	M	F	K	K	-	P	T	P	S
Zebrafish	α1	A	T	I	E	R	S	E	M	F	K	K	-	P	T	P	S
Human	α3	A	S	L	N	P	E	R	M	F	R	K	-	P	I	P	S
Mouse	α3	A	S	L	N	P	E	R	M	F	R	K	-	P	I	P	S
Zebrafish	α3	A	S	L	D	H	T	N	M	F	S	M	-	P	N	R	Q
Human	α5	A	T	V	D	V	S	D	M	F	S	K	-	P	Q	S	E
Mouse	α5	A	T	V	D	M	S	D	M	F	N	K	-	P	Q	S	E
Zebrafish	α5	A	T	V	D	V	N	E	M	F	R	K	-	P	Q	S	E
Human	α2	T	T	I	P	E	-	Q	S	F	Q	G	S	P	S	A	D
Mouse	α2	T	T	I	P	E	-	Q	N	F	Q	S	T	P	S	A	D
Zebrafish	α2	T	S	I	D	E	-	-	S	F	Q	S	S	P	S	S	E
Human	α4	T	T	V	K	A	D	L	Q	F	S	S	A	P	A	P	D
Mouse	α4	T	T	V	N	P	D	L	Q	F	A	S	G	P	S	P	D
Zebrafish	α4	T	I	D	M	E	H	N	D	-	-	S	S	P	H	G	P
Human	α6	T	T	V	E	E	R	Q	Q	F	G	E	L	P	V	S	E
Mouse	α6	T	T	V	E	E	R	G	Q	F	R	E	E	P	V	S	E
Zebrafish	α6	T	T	V	D	V	N	E	Q	F	V	Q	G	P	T	Q	E

Supplementary Figure 4