

Lawrence Berkeley National Laboratory

LBL Publications

Title

Identification of unique $\alpha 4$ chain structure and conserved antiangiogenic activity of $\alpha 3\text{NC1}$ type IV collagen in zebrafish.

Permalink

<https://escholarship.org/uc/item/43c4v4bm>

Journal

Developmental Dynamics, 252(7)

Authors

LeBleu, Valerie
Dai, Jianli
MacDonald, Brian
[et al.](#)

Publication Date

2023-07-01

DOI

10.1002/dvdy.590

Peer reviewed



Published in final edited form as:

Dev Dyn. 2023 July ; 252(7): 1046–1060. doi:10.1002/dvdy.590.

Identification of unique $\alpha 4$ chain structure and conserved anti-angiogenic activity of $\alpha 3\text{NC1}$ type IV collagen in zebrafish

Valerie S. LeBleu^{1,2,3,*,#}, Jianli Dai^{1,#}, Susan Tsutakawa⁴, Brian A. MacDonald², Joseph L. Alge^{1,5}, Malin Sund³, Liang Xie³, Hikaru Sugimoto^{1,3}, John Tainer^{1,7}, Leonard I. Zon⁶, Raghu Kalluri^{1,3,8,9,*}

¹Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, TX

²Feinberg School of Medicine and Kellogg School of Management, Northwestern University, Chicago, IL

³Division of Matrix Biology, Beth Israel Deaconess Medical Center, Boston, MA

⁴Lawrence Berkeley National Laboratory, University of California, Berkeley, CA

⁵Department of Pediatrics, Baylor College of Medicine, Houston, TX

⁶Department of Hematology/Oncology, Children's Hospital, Boston, MA

⁷Department of Molecular and Cellular Oncology, University of Texas MD Anderson Cancer Center, Houston, TX

⁸Department of Bioengineering, Rice University, Houston, TX

⁹Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX

Abstract

Background: Type IV collagen is an abundant component of basement membranes in all multicellular species and is essential for the extracellular scaffold supporting tissue architecture and function. Lower organisms typically have two type IV collagen genes, encoding $\alpha 1$ and $\alpha 2$ chains, in contrast with the six genes in humans, encoding $\alpha 1$ to $\alpha 6$ 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 remains to be studied.

Results: We report on the molecular evolution of type IV collagen genes. The zebrafish $\alpha 4$ non-collagenous (NC1) domain, in contrast with its human ortholog, contains an additional cysteine

*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.
#co-first authors

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 the initial homology analysis. VSL, ST, JD, BAM prepared the figures. RK oversaw the scientific design of the project and edited the manuscript.

Grant support: This study was partially funded by a research grant from Emerald Foundation, NIH research grant DK55001 to Raghu Kalluri, Cancer Prevention and Research Institute of Texas, 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

residue and lacks the M93 and K211 residues involved in sulfilimine bond formation between adjacent protomers. This may alter $\alpha 4$ 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 $\alpha 3$ NC1 domain (endogenous angiogenesis inhibitor, Tumstatin), the zebrafish $\alpha 3$ NC1 domain exhibits conserved anti-angiogenic activity in human endothelial cells.

Conclusions: Our work supports type IV collagen is largely conserved between zebrafish and humans, with a possible difference involving the $\alpha 4$ chain.

Keywords

collagen IV; basement membrane; Tumstatin; angiogenesis; zebrafish

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 network⁷. 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 $\alpha 1$ and $\alpha 2$, whereas mammals express six distinct chains named $\alpha 1$ through $\alpha 6$, 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 $\alpha 1\alpha 2\alpha 1$, $\alpha 3\alpha 4\alpha 5$, and $\alpha 5\alpha 6\alpha 5$ have been identified in basement membranes^{16,17}. The collagen IV $\alpha 1\alpha 2\alpha 1$ network, considered to be more primordial, is ubiquitously found in basement membranes of all tissues. The $\alpha 1\alpha 2\alpha 1$ network is thought to be critical to the evolution of multicellular life^{18–20}. Indeed, genetic deletion of collagen IV $\alpha 1$ and/or $\alpha 2$ genes in mice results in a loss of basement membrane integrity and embryonic lethality^{21–23}, and loss of the collagen IV $\alpha 1$ alone in *C. elegans* and *Drosophila* leads to developmental arrest and embryonic lethality^{1,24–26}. It is also reported that collagen IV $\alpha 1$ and $\alpha 2$ mutations correlate with diseases in humans²⁷. In contrast, collagen IV $\alpha 3$ - $\alpha 6$ chains, only found in higher-order organisms, have restricted tissue distribution during development, and present in more specialized basement membranes¹⁷. For instance, the $\alpha 3\alpha 4\alpha 5$ (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 $\alpha 3$ -, $\alpha 4$ -, and $\alpha 5$ -chain encoding genes leads to the absence of $\alpha 3\alpha 4\alpha 5$ (IV) protomer formation in GBM, which is associated with several nephropathies, including Alport syndrome, thin basement membrane nephropathy, and focal segmental glomerulosclerosis^{2,28–31}. The $\alpha 3\alpha 4\alpha 5$ 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 $\alpha 3\alpha 4\alpha 5$ network composition in these basement membranes is associated with multiple organ impairment in Alport syndrome³⁹. In addition, autoantibodies against the NC1 domain of the collagen IV $\alpha 3$ 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 membranes¹³, the cleavage of the NC1 domain releases the bioactive $\alpha 3$ NC1 fragment, Tumstatin, which exerts anti-angiogenic and pro-apoptotic functions that significantly influences tumor growth^{42–44}. Interestingly, zebrafish, like mammals, express six collagen IV α chains⁴⁵. The development of the collagen IV $\alpha 3\alpha 4\alpha 5$ and $\alpha 5\alpha 6\alpha 5$ networks are thought to result from more recent evolutionary events, and these networks may not be as highly conserved as the primordial $\alpha 1\alpha 2\alpha 1$ (IV) network. Indeed, the $\alpha 3$ -chain cryptic epitope/s targeted by the autoantibodies in Goodpasture syndrome do not appear to be conserved in zebrafish⁴⁶. Little is otherwise known about the evolutionary divergence of collagen IV basement membrane composition from the zebrafish to mammals^{47,48}. 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 type collagen IV chains in zebrafish development at the mRNA level. We also demonstrate that the zebrafish $\alpha 3$ chain NC1 domain (zTumstatin) is present with conserved anti-angiogenic activity when compared to human Tumstatin.

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^{49–52}, and the head-to-head organization of the Col4 $\alpha 1$ -Col4 $\alpha 2$ pairs, Col4 $\alpha 3$ -Col4 α pairs and Col4 $\alpha 5$ -Col4 $\alpha 6$ pairs was found in human, mouse and zebrafish species (Fig. 1a). Each gene encodes for a protein with a characteristic N-terminal 7S domain, a collagenous domain, and a C-terminal non-collagenous domain (NC1) (Fig. 1b, Fig. 2). A high degree of amino acid sequence homology between the NC1 domains of zebrafish $\alpha 1$, $\alpha 2$, $\alpha 5$, and $\alpha 6$ chains and the mouse and human orthologs was noted, with more pronounced dissimilarity for $\alpha 3$ and $\alpha 4$ chains (Fig. 1c, 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. 3a). All six zebrafish Col4 chains contain the HSQ tripeptides that signal the beginning of the NC1 subdomain I and II (Fig. 3a, cyan). Each zebrafish α chain also contains twelve conserved cysteine residues (Fig. 3a, 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 residues⁹. These residues are mostly conserved in all six zebrafish α chains, except for R179 in the $\alpha 2$ chain, for which a R179K substitution is noted (Fig. 3a, red arrow). In addition, the zebrafish $\alpha 3$ and $\alpha 4$ NC1 domains lack the N187 residue (Fig. 3a, 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 peroxidase enzyme^{10,11,53}. Five of the six zebrafish α chains (except for the $\alpha 4$ chain) contain the M93 and K211 residues associated with the formation of the sulfilimine bond and protomer dimerization (Fig. 3a, yellow)¹⁰.

In contrast with other collagen IV α chains, the zebrafish $\alpha 4$ chain has two important divergent features. First, the zebrafish collagen IV $\alpha 4$ 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. 3a, magenta). The Cys-3 to Cys-4 disulfide bond forms a β -hairpin structure in the SM2 subdomain ($\beta 6$ - $\beta 7$) in the NC1 domain that inserts into the SM3' subdomain of an adjacent NC1 domain^{8,54-56}. 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 $\alpha 4$ -chain shorter than in other species; this deletion, coupled with the natural flexibility in the peptide backbone of the domain (asterisk, Fig. 3b), 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 $\alpha 4$ -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. 3b-c)⁵⁷. Second, the zebrafish $\alpha 4$ chain lacks the M93 and K211 residues required for the peroxidase-catalyzed sulfilimine bond formation that stabilizes end-to-end protomer dimerization¹⁰. Such deficit would prevent the covalent linking of $\alpha 4$ containing-protomers at the C terminus. These results suggest that zebrafish collagen IV networks containing the $\alpha 4$ chain may be weaker than their human homolog.

Unique temporal expression patterns of type IV collagen chains in zebrafish embryonic 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. 4a). Expression of $\alpha 1$ and $\alpha 2$ chains mRNA was not restricted to a particular developmental stage, and the synchronous expression patterns of the $\alpha 1$ and $\alpha 2$ chains supports that zebrafish synthesize the $\alpha 1\alpha 2\alpha 1$ (IV) protomer throughout embryogenesis (Fig. 4a). In addition, transcription of genes encoding the $\alpha 5$ and $\alpha 6$ chains was synchronously detected in post-gastrulation embryos, suggesting that synthesis of the $\alpha 5\alpha 6\alpha 5$ (IV) protomer does not occur in the pre-gastrulation stages of embryogenesis (Fig. 4a). Finally, in late stage of embryos, when there was also a high level of $\alpha 5$ chain gene transcription, transcription of $\alpha 3$ and $\alpha 4$ chains was observed, supporting that the possible synthesis of the $\alpha 3\alpha 4\alpha 5$ (IV) protomer in later stages of zebrafish embryonic development (Fig. 4a). 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 $\alpha 1$ and $\alpha 4$ chains are synchronously expressed in the absence of other chains at the shield time point (Fig. 4a, lane 3). In addition, there is no detectable $\alpha 3$ chain expression despite persistent expression of $\alpha 4$ at the Prim-11 stage (Fig. 4a, lane 12). This raises the possible existence of novel protomer involving the $\alpha 4$ chain other than the $\alpha 3\alpha 4\alpha 5$ (IV) protomer at this stage, such as $\alpha 1\alpha 4\alpha 1$. In zebrafish, expression of proteins before 6 hours post hybridization (hpf) stage is maternal. Only the $\alpha 5$ chain is not maternally contributed, which also supports that the $\alpha 3$ - and $\alpha 4$ -chains could assemble protomers with other composition than the $\alpha 3\alpha 4\alpha 5$ (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 $\alpha 1$ - $\alpha 4$ heterohexamer is not formed in human, we postulated that mutation in zebrafish $\alpha 4$ allow for this formation. Targeting regions identified to be critical for oligomerization⁵⁸, we identified any changes in zebrafish $\alpha 4$ NC1 that made it similar to $\alpha 1$ and $\alpha 2$ NC1. Although there were several mutations that may render zebrafish $\alpha 4$ to be similar to $\alpha 2$, only one change made it similar to both $\alpha 1$ and $\alpha 2$ in one region, SM1' (structural motif SM1': hairpin $\beta 3'$ - $\beta 4'$ ⁵⁸) (Fig. 2). Examination of the $\alpha 1/\alpha 2$ heterohexamer structure from human placenta⁵⁵ revealed that the mutation in zebrafish $\alpha 4$ to glutamate (E1589) would make it analogous to E1592 and E1633 in $\alpha 1$ and $\alpha 2$ respectively (Fig. 4b-c). These analyses raise the possibility for $\alpha 4$ chain containing (IV) protomer that include either the $\alpha 1$ or $\alpha 2$ 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 study⁵⁹, our results show that the $\alpha 1$ and $\alpha 2$ chains were transcribed ubiquitously (Fig. 4a, d). Although the caudal expression level was low, no specific organ displayed a differential distribution of $\alpha 1$ or $\alpha 2$ chain. The $\alpha 3$ and $\alpha 4$ 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. 4e). At 36 hpf, $\alpha 6$ chain expression in the otic vesicle, gills (g), and pronephric ducts (pn) could offer support for the synthesis of the $\alpha 5\alpha 6\alpha 5$ protomer in these tissues (Fig 4a, Fig. 5a).

In humans and mice, organs such as intestine and skin are strictly composed of the collagen IV $\alpha 1\alpha 2\alpha 1$ and $\alpha 5\alpha 6\alpha 5$ protomers, yet other organs display a developmental restricted expression of specific protomers^{35,60,61}. For example, during human nephrogenesis, the GBM is initially comprised of the collagen IV $\alpha 1\alpha 2\alpha 1$ protomer. As the glomerulus matures, the GBM collagen IV undergoes an isoform switch from the $\alpha 1\alpha 2\alpha 1$ to the $\alpha 3\alpha 4\alpha 5$ protomer^{62,63}. Our results show that the expression of the more specialized $\alpha 3\alpha 4\alpha 5$ protomer was not detected by *in situ* hybridization in many of the zebrafish organs during embryonic development. We posit that the absence of the $\alpha 3\alpha 4\alpha 5$ 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 $\alpha 1$, $\alpha 2$, $\alpha 5$ and $\alpha 6$ chains were ubiquitously expressed in the adult zebrafish (although $\alpha 5$ chain expression was low in the cardiac tissue), but the $\alpha 3$ and $\alpha 4$ chains

exhibited a more restricted expression pattern (Fig. 5b). The $\alpha 3$ 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. 5b). The $\alpha 4$ chain had a similar expression pattern, but it was also expressed in cardiac and hepatic tissue (Fig. 5b). These results support a protomer isoform switch in the post-embryonic zebrafish development, similarly to what is observed in mammals. The $\alpha 4$ chain expression in the adult cardiac tissue, concurrent with absent $\alpha 3$ and low $\alpha 5$ chain expression, indicates a more promiscuous $\alpha 4$ chain expression in zebrafish compared to human, and supports potential formation of novel protomers composition.

Conserved anti-angiogenic function of the zebrafish collagen IV $\alpha 3$ NC1 domain, zTumstatin

Given that Tumstatin, the NC1 domain of collagen IV $\alpha 3$ chain, is a known anti-angiogenic cleavage product of Col4 $\alpha 3$, we wondered whether zebrafish $\alpha 3$ -chain NC1 domain, zTumstatin, would also exert anti-angiogenic functions, similarly to its human homolog^{42,43,64}. To test this hypothesis, we produced recombinant zTumstatin using a eukaryotic expression system, as previously described⁴⁶. Human umbilical vein endothelial cells (HUVEC) proliferation was inhibited in a dose-dependent manner by eukaryotic zTumstatin (Fig. 6a). zTumstatin also inhibited HUVEC migration in a Boyden chamber assay (Fig. 6b). 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. 6c). Microscopic angiogenesis, evaluated by H&E staining of Matrigel™ plugs, also revealed suppressed angiogenesis (54% reduction) in zTumstatin-containing plugs compared to control (Fig. 6d–e). These results support antiangiogenic function of the zTumstatin.

Discussion

Collagen IV exists in all metazoans, including sponges¹⁸. Type IV collagen has pivotal function in shaping tissue and organs, transducing signaling pathways and maintaining normal development in multicellular organisms^{3,18}. 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 $\alpha 3$ chain with $\alpha 1$ and $\alpha 5$ chain, and the increased similarities in the $\alpha 4$ chain with the $\alpha 2$ and $\alpha 6$ chains^{18,65}. 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, peroxidase and lysyl oxidase-like 2^{11,56,66}, are also conserved. We note that the zebrafish α 4-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^{10,55}, suggesting distinct adjacent protomer interaction for the zebrafish α 4-containing protomer. 3D remodeling in our study also supports that the additional cysteine residue alters the zebrafish α 4-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: α 1 α 2 α 1, α 3 α 4 α 5 and α 5 α 6 α 5. Similarly, in mice, spatial and temporal distribution of collagen IV isoforms in the developing eye also reveals these protomers⁶⁷. In our study, we note that the transcriptional pattern of α 1 and α 2 chain is widely distributed, whereas which of the α 3- α 6 chain exhibits a restricted, organ specificity pattern, consistent with recent findings⁵⁹. Synchronous expression of α 3-, α 4- and α 5- chains in the brain suggests that α 3 α 4 α 5 (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 development⁶⁵. Furthermore, synchronous expression of these α chains in the otic vesicle and pronephros support an evolutionary conserved anatomic specificity for α 3 α 4 α 5 synthesis, given that this protomer is the major component of the cochlear and glomerular basement membranes in humans. The chains expression enabling α 3 α 4 α 5 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^{62,63}.

Our findings unexpectedly yielded complementary evidence (homology and temporospatial expression) that the zebrafish collagen IV protomer repertoire may include novel protomers. For instance, zebrafish α 4 chain was synchronously expressed with the α 1 chain in the absence of α 2, α 3 and α 5 chains at the shield time point and in cardiac tissue of adult zebrafish, wherein α 3 and α 5 chain expression were lacking or were minimal. These findings suggest the potential existence of a putative α 1 α 4 α 1 protomer. The zebrafish α 4 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 protomers⁵⁴⁻⁵⁶. Lack of M93 and K211 residues and an additional cysteine residue may trigger distinct adjacent protomer interaction for the zebrafish α 4-containing protomer. Collectively, our results suggest that apart from α 3 α 4 α 5, other potential collagen IV protomers (α 1 α 4 α 1) may present in zebrafish. Although, our analyses rely on gene transcription findings, we previously showed protein-level distribution of collagen IV α 3⁴⁶. Collagen IV α 3 protein expression in the adult zebrafish is in the gills and kidney⁴⁶, which is in agreement with our RNA analyses. Collagen IV α 4, using protein trap insertional mutagenesis was detected in

the caudal vascular plexus and myotomes of the developing zebrafish⁶⁸. 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 α 3 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 conserved⁴⁶. This suggests a distinct immune system maturation and regulation of the zebrafish compared to human, with respect to the autoimmune response to collagen IV α 3 chain in Goodpasture syndrome⁶⁹. 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 organisms⁷⁰.

Experimental Procedures

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 Omega and MView. 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 $[(\alpha 1)2(\alpha 2)]2$ 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 described⁴⁶. 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 TRIzolTM (Invitrogen) reagent and immediately homogenized. Generation of cDNA was performed using Superscript IITM 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.

z α 1 NC1 Forward: 5'-TCCTCAATGGACCACGGCTTCCTT-3'

za.1 NC1 Reverse: 5'-CCTGCTGATGTGAGTTCTTAG-3'
 za.2 NC1 Forward: 5'-CGCAGTGTGAACGTTGGCTAT-3'
 za.2 NC1 Reverse: 5'-TTGGCAGCGGATGATGCGGGA-3'
 za.3 NC1 Forward: 5'-AGAGACGGATTCTTTTCACC-3'
 za.3 NC1 Reverse: 5'-AGGCTGTGCTTCATACAGAC-3'
 za.4 NC1 Forward: 5'-GGAAGAAATGAGACACTTTCT-3'
 za.4 NC1 Reverse: 5'-TCATTGGTTGGGTCATTCAT-3'
 za.5 NC1 Forward: 5'-ACTTTGTCTGGTGCACACGGT-3'
 za.5 NC1 Reverse: 5'-TTACGTCCTTTCATACACAC-3'
 za.6 NC1 Forward: 5'-ACGCGCAGCACTGGCATTGGT-3'
 za.6 NC1 Reverse: 5'-ACTGAGACGCGAGCGGAACTG-3'

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

Whole-mount *in situ* hybridization was performed as described⁷¹. Digoxigenin-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 ($\alpha 1$ - $\alpha 6$) in zebrafish are:

za.1 NC1 Forward: 5'-TCCTGCCTCAGCGCTCTTTCTCAT-3'
 za.1 NC1 Reverse: 5'-CATAAGTTTTTCATAAATAACTAAG-3'
 za.2 NC1 Forward: 5'-CTCCTGTCCCGCATCAGCCGCTGC-3'
 za.2 NC1 Reverse: 5'-GCAAGATAAAACCCCTGTTGCC-3'
 za.3 NC1 Forward: 5'-CAACAGCCTCAAGGTGACCTTGCT-3'
 za.3 NC1 Reverse: 5'-GGGCATGCCTTGCTTACAATGGAAACCGCC-3'
 za.4 NC1 Forward: 5'-TGATAAAGCCAATGGTATTTTCT-3'
 za.4 NC1 Reverse: 5'-TTACTATTGTAGGTGAAATATCCC-3'
 za.6 NC1 Forward: 5'-AGCTTGTGTAGATGCACTGCAGGA-3'
 za.6 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^{46,72}.

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 μ M 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 μ m 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 μ mol/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. 6b–c, compared to positive control. Two-tailed unpaired Student's t test was used in Fig. 6e. The data is presented as the mean \pm the standard error of the mean. Significance of statistical tests is reported in all graphs as follows: ****, $p < 0.0001$; ***, $p < 0.001$; **, $p < 0.01$.

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.

Abbreviations:

NC1 domain	non-collagenous 1 domain
GBM	glomerular basement membrane
hpf	hour post fertilization

References

1. Borchellini C, Coulon J, Le Parco Y. The function of type IV collagen during *Drosophila* embryogenesis. *Roux Arch Dev Biol*. May 1996;205(7–8):468–475. 10.1007/bf00377228. [PubMed: 28306099]
2. Cosgrove D, Liu S. Collagen IV diseases: A focus on the glomerular basement membrane in Alport syndrome. *Matrix Biol*. Jan 2017;57–58:45–54. 10.1016/j.matbio.2016.08.005. [PubMed: 27751945]
3. Kalluri R. Basement membranes: structure, assembly and role in tumour angiogenesis. *Nat Rev Cancer*. Jun 2003;3(6):422–33. 10.1038/nrc1094. [PubMed: 12778132]
4. LeBleu VS, Macdonald B, Kalluri R. Structure and function of basement membranes. *Exp Biol Med* (Maywood). Oct 2007;232(9):1121–9. 10.3181/0703-MR-72. [PubMed: 17895520]
5. Pozzi A, Yurchenco PD, Iozzo RV. The nature and biology of basement membranes. *Matrix Biol*. Jan 2017;57–58:1–11. 10.1016/j.matbio.2016.12.009. [PubMed: 27751945]
6. Yurchenco PD, Amenta PS, Patton BL. Basement membrane assembly, stability and activities observed through a developmental lens. *Matrix Biol*. Jan 2004;22(7):521–38. 10.1016/j.matbio.2003.10.006. [PubMed: 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*. Nov 1981;120(2):203–11. 10.1111/j.1432-1033.1981.tb05690.x. [PubMed: 6274634]
8. Brown KL, Cummings CF, Vanacore RM, Hudson BG. Building collagen IV smart scaffolds on the outside of cells. *Protein Sci*. 2017;26(11):2151–2161. [PubMed: 28845540]
9. Cummings CF, Pedchenko V, Brown KL, et al. Extracellular chloride signals collagen IV network assembly during basement membrane formation. *J Cell Biol*. May 23 2016;213(4):479–94. 10.1083/jcb.201510065. [PubMed: 27216258]
10. Fidler AL, Vanacore RM, Chetyrkin SV, et al. A unique covalent bond in basement membrane is a primordial innovation for tissue evolution. *Proc Natl Acad Sci U S A*. Jan 7 2014;111(1):331–6. 10.1073/pnas.1318499111. [PubMed: 24344311]
11. Vanacore R, Ham AJ, Voehler M, et al. A sulfilimine bond identified in collagen IV. *Science*. Sep 4 2009;325(5945):1230–4. 10.1126/science.1176811. [PubMed: 19729652]
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*. Dec 15 1993;268(35):26033–6. [PubMed: 8253711]
13. LeBleu V, Sund M, Sugimoto H, 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*. Dec 31 2010;285(53):41874–85. 10.1074/jbc.M110.149534. [PubMed: 20847057]
14. Ries A, Engel J, Lustig A, Kuhn K. The function of the NC1 domains in type IV collagen. *J Biol Chem*. Oct 6 1995;270(40):23790–4. 10.1074/jbc.270.40.23790. [PubMed: 7559554]
15. Borza DB, Bondar O, Ninomiya Y, et al. The NC1 domain of collagen IV encodes a novel network composed of the α_1 , α_2 , α_5 , and α_6 chains in smooth muscle basement membranes. *J Biol Chem*. Jul 27 2001;276(30):28532–40. 10.1074/jbc.M103690200. [PubMed: 11375996]
16. 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. 10.1159/000422975. [PubMed: 8004963]
17. Khoshnoodi J, Pedchenko V, Hudson BG. Mammalian collagen IV. *Microsc Res Tech*. May 2008;71(5):357–70. 10.1002/jemt.20564. [PubMed: 18219669]
18. Fidler AL, Darris CE, Chetyrkin SV, et al. Collagen IV and basement membrane at the evolutionary dawn of metazoan tissues. *Elife*. Apr 18 2017;6. 10.7554/eLife.24176.
19. Grau-Bove X, Torruella G, Donachie S, et al. Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife*. Jul 20 2017;6. 10.7554/eLife.26036.
20. Kuhn K. Basement membrane (type IV) collagen. *Matrix Biol*. Feb 1995;14(6):439–45. 10.1016/0945-053x(95)90001-2. [PubMed: 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*. Apr 2004;131(7):1619–28. 10.1242/dev.01037. [PubMed: 14998921]
22. Gould DB, Phalan FC, Breedveld GJ, et al. Mutations in Col4a1 cause perinatal cerebral hemorrhage and porencephaly. *Science*. May 20 2005;308(5725):1167–71. 10.1126/science.1109418. [PubMed: 15905400]
23. Favor J, Gloeckner CJ, Janik D, 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*. Feb 2007;175(2):725–36. 10.1534/genetics.106.064733. [PubMed: 17179069]
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*. Aug 16 2011;21(2):245–56. 10.1016/j.devcel.2011.06.026. [PubMed: 21839919]
25. Sessions AO, Kaushik G, Parker S, et al. Extracellular matrix downregulation in the *Drosophila* heart preserves contractile function and improves lifespan. *Matrix Biol*. Oct 2017;62:15–27. 10.1016/j.matbio.2016.10.008. [PubMed: 27793636]
26. Kelemen-Valkony I, Kiss M, Csiha J, et al. *Drosophila* basement membrane collagen col4a1 mutations cause severe myopathy. *Matrix Biol*. Jan 2012;31(1):29–37. 10.1016/j.matbio.2011.09.004. [PubMed: 22037604]
27. Jeanne M, Gould DB. Genotype-phenotype correlations in pathology caused by collagen type IV alpha 1 and 2 mutations. *Matrix Biol*. Jan 2017;57–58:29–44. 10.1016/j.matbio.2016.10.003. [PubMed: 27751945]
28. Savige J Alport syndrome: its effects on the glomerular filtration barrier and implications for future treatment. *J Physiol*. Sep 15 2014;592(18):4013–23. 10.1113/jphysiol.2014.274449. [PubMed: 25107927]
29. Gast C, Pengelly RJ, Lyon M, et al. Collagen (COL4A) mutations are the most frequent mutations underlying adult focal segmental glomerulosclerosis. *Nephrol Dial Transplant*. Jun 2016;31(6):961–70. 10.1093/ndt/gfv325. [PubMed: 26346198]
30. Gatseva A, Sin YY, Brezzo G, Van Agtmael T. Basement membrane collagens and disease mechanisms. *Essays Biochem*. Aug 6 2019. 10.1042/EBC20180071.
31. Iozzo RV, Gubbio MA. Extracellular matrix: The driving force of mammalian diseases. *Matrix Biol*. Oct 2018;71–72:1–9. 10.1016/j.matbio.2018.03.023.
32. Frojzman K, Pelliniemi LJ, Virtanen I. Differential distribution of type IV collagen chains in the developing rat testis and ovary. *Differentiation*. Jul 1998;63(3):125–30. 10.1046/j.1432-0436.1998.6330125.x. [PubMed: 9697306]
33. 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. 10.3109/03008209809028906. [PubMed: 9643653]
34. 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*. Sep 16 1994;269(37):23013–7. [PubMed: 8083201]
35. Ninomiya Y, Kagawa M, Iyama K, 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*. Sep 1995;130(5):1219–29. 10.1083/jcb.130.5.1219. [PubMed: 7657706]
36. 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*. Oct 1990;111(4):1685–99. 10.1083/jcb.111.4.1685. [PubMed: 2211832]
37. Urabe N, Naito I, Saito K, et al. Basement membrane type IV collagen molecules in the choroid plexus, pia mater and capillaries in the mouse brain. *Arch Histol Cytol*. Jun 2002;65(2):133–43. 10.1679/aohc.65.133. [PubMed: 12164337]

38. Zehnder AF, Adams JC, Santi PA, et al. Distribution of type IV collagen in the cochlea in Alport syndrome. *Arch Otolaryngol Head Neck Surg*. Nov 2005;131(11):1007–13. 10.1001/archotol.131.11.1007. [PubMed: 16301374]
39. Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG. Alport's syndrome, Goodpasture's syndrome, and type IV collagen. *N Engl J Med*. Jun 19 2003;348(25):2543–56. 10.1056/NEJMra022296. [PubMed: 12815141]
40. Pedchenko V, Bondar O, Fogo AB, et al. Molecular architecture of the Goodpasture autoantigen in anti-GBM nephritis. *N Engl J Med*. Jul 22 2010;363(4):343–54. 10.1056/NEJMoa0910500. [PubMed: 20660402]
41. Pedchenko V, Kitching AR, Hudson BG. Goodpasture's autoimmune disease - A collagen IV disorder. *Matrix Biol*. Oct 2018;71–72:240–249. 10.1016/j.matbio.2018.05.004.
42. Hamano Y, Zeisberg M, Sugimoto H, 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*. Jun 2003;3(6):589–601. 10.1016/s1535-6108(03)00133-8. [PubMed: 12842087]
43. Maeshima Y, Sudhakar A, Lively JC, et al. Tumstatin, an endothelial cell-specific inhibitor of protein synthesis. *Science*. Jan 4 2002;295(5552):140–3. 10.1126/science.1065298. [PubMed: 11778052]
44. Mundel TM, Kalluri R. Type IV collagen-derived angiogenesis inhibitors. *Microvasc Res*. Sep-Nov 2007;74(2–3):85–9. 10.1016/j.mvr.2007.05.005. [PubMed: 17602710]
45. 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):152–159.
46. MacDonald BA, Sund M, Grant MA, 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*. Mar 1 2006;107(5):1908–15. 10.1182/blood-2005-05-1814. [PubMed: 16254142]
47. Drummond IA. Kidney development and disease in the zebrafish. *J Am Soc Nephrol*. Feb 2005;16(2):299–304. 10.1681/asn.2004090754. [PubMed: 15647335]
48. Morales EE, Wingert RA. Zebrafish as a Model of Kidney Disease. *Results Probl Cell Differ*. 2017;60:55–75. 10.1007/978-3-319-51436-9_3. [PubMed: 28409342]
49. Soder S, Poschl E. The NC1 domain of human collagen IV is necessary to initiate triple helix formation. *Biochem Biophys Res Commun*. Dec 3 2004;325(1):276–80. 10.1016/j.bbrc.2004.10.034. [PubMed: 15522229]
50. Colorado PC, Torre A, Kamphaus G, et al. Anti-angiogenic cues from vascular basement membrane collagen. *Cancer Res*. May 1 2000;60(9):2520–6. [PubMed: 10811134]
51. Marneros AG, Olsen BR. The role of collagen-derived proteolytic fragments in angiogenesis. *Matrix Biol*. Sep 2001;20(5–6):337–45. 10.1016/s0945-053x(01)00151-2. [PubMed: 11566268]
52. Brodsky B, Ramshaw JA. The collagen triple-helix structure. *Matrix Biol*. Mar 1997;15(8–9):545–54. 10.1016/s0945-053x(97)90030-5. [PubMed: 9138287]
53. Ero-Tolliver IA, Hudson BG, Bhawe G. The Ancient Immunoglobulin Domains of Peroxidase Are Required to Form Sulfilimine Cross-links in Collagen IV. *J Biol Chem*. Aug 28 2015;290(35):21741–8. 10.1074/jbc.M115.673996. [PubMed: 26178375]
54. Khoshnoodi J, Sigmundsson K, Cartailier 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*. Mar 3 2006;281(9):6058–69. 10.1074/jbc.M506555200. [PubMed: 16373348]
55. Than ME, Henrich S, Huber R, 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*. May 14 2002;99(10):6607–12. 10.1073/pnas.062183499. [PubMed: 12011424]
56. 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*. Aug 12 2005;280(32):29300–10. 10.1074/jbc.M502752200. [PubMed: 15951440]

57. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. Apr 2010;7(4):248–9. 10.1038/nmeth0410-248. [PubMed: 20354512]
58. Casino P, Gozalbo-Rovira R, Rodriguez-Diaz J, et al. Structures of collagen IV globular domains: insight into associated pathologies, folding and network assembly. *IUCrJ*. Nov 1 2018;5(Pt 6):765–779. 10.1107/S2052252518012459.
59. 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*. Aug 1 2019;452(1):43–54. 10.1016/j.ydbio.2019.04.012. [PubMed: 31034836]
60. Tanaka K, Iyama K, Kitaoka M, 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*. Jul 1997;29(7):563–70. [PubMed: 9279559]
61. Kelley PB, Sado Y, Duncan MK. Collagen IV in the developing lens capsule. *Matrix Biol*. Aug 2002;21(5):415–23. 10.1016/s0945-053x(02)00014-8. [PubMed: 12225806]
62. 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*. May 15 1997;99(10):2470–8. 10.1172/jci119431. [PubMed: 9153291]
63. Kuroda N, Yoshikawa N, Nakanishi K, et al. Expression of type IV collagen in the developing human kidney. *Pediatr Nephrol*. Sep 1998;12(7):554–8. [PubMed: 9761353]
64. Eikesdal HP, Sugimoto H, Birrane G, 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*. Sep 30 2008;105(39):15040–5. 10.1073/pnas.0807055105. [PubMed: 18818312]
65. Takeuchi M, Yamaguchi S, Yonemura S, et al. Type IV Collagen Controls the Axogenesis of Cerebellar Granule Cells by Regulating Basement Membrane Integrity in Zebrafish. *PLoS Genet*. Oct 2015;11(10):e1005587. 10.1371/journal.pgen.1005587. [PubMed: 26451951]
66. Anazco C, Lopez-Jimenez AJ, Rafi M, et al. Lysyl Oxidase-like-2 Cross-links Collagen IV of Glomerular Basement Membrane. *J Biol Chem*. Dec 9 2016;291(50):25999–26012. 10.1074/jbc.M116.738856. [PubMed: 27770022]
67. Bai X, Dilworth DJ, Weng YC, Gould DB. Developmental distribution of collagen IV isoforms and relevance to ocular diseases. *Matrix Biol*. May 2009;28(4):194–201. 10.1016/j.matbio.2009.02.004. [PubMed: 19275937]
68. Westcot SE, Hatzold J, Urban MD, 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. 10.1371/journal.pone.0130688. [PubMed: 26110643]
69. Foster MH. Basement membranes and autoimmune diseases. *Matrix Biol*. Jan 2017;57–58:149–168. 10.1016/j.matbio.2016.07.008. [PubMed: 27751945]
70. 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*. Jan 2017;57–58:55–75. 10.1016/j.matbio.2016.10.002. [PubMed: 27751945]
71. 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*. Dec 1993;119(4):1203–15. [PubMed: 8306883]
72. Maeshima Y, Yerramalla UL, Dhanabal M, et al. Extracellular matrix-derived peptide binds to alpha(v)beta(3) integrin and inhibits angiogenesis. *J Biol Chem*. Aug 24 2001;276(34):31959–68. 10.1074/jbc.M103024200. [PubMed: 11399763]

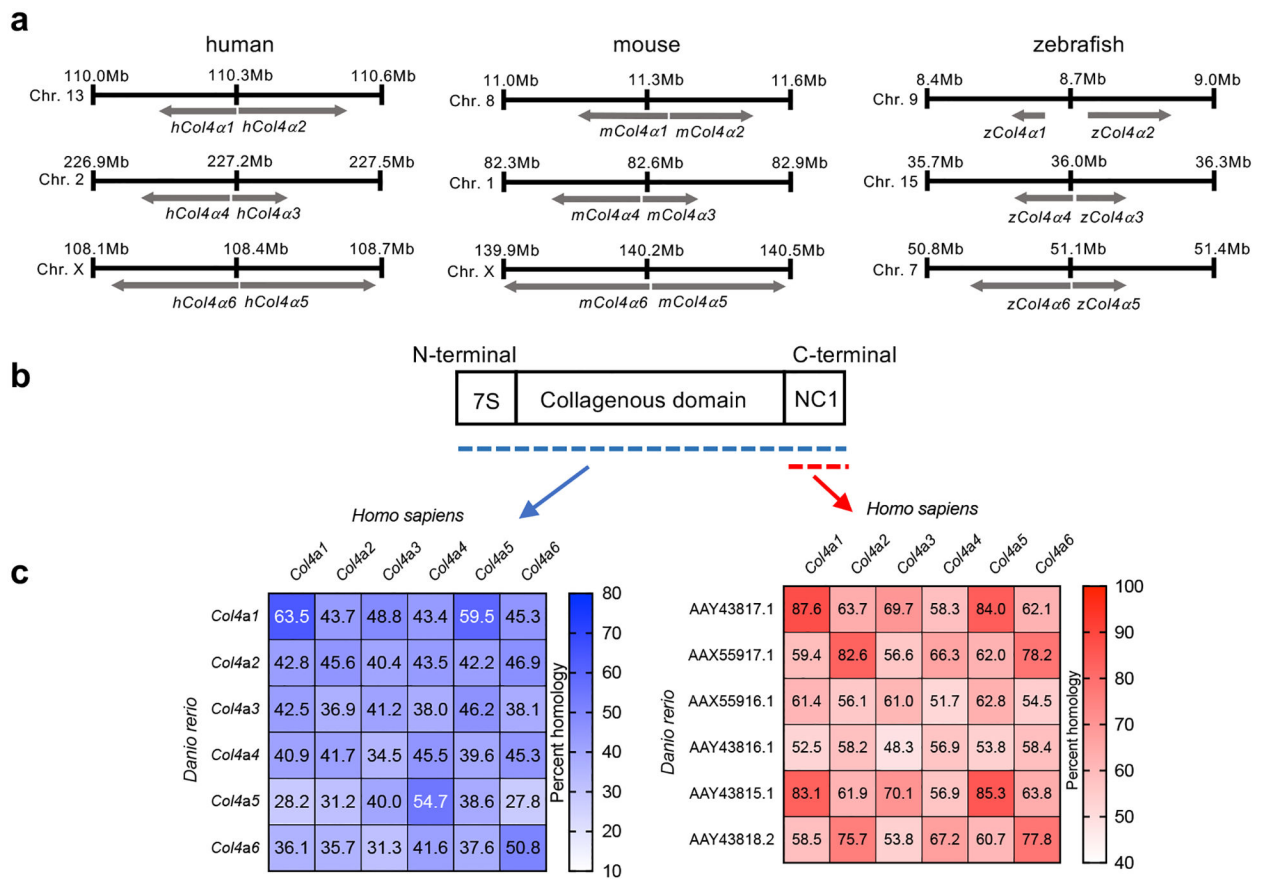


Figure 1: Chromosome mapping for α chains between human, mouse, and zebrafish and amino acid sequence homology.

a. Chromosome mapping of human, mouse, and zebrafish collagen IV α chains. **b.** Schematic representation of the domains of Collagen IV chains, and **c.** percent homology of amino acids in full length collagen IV (blue) or NC1 domains (red).

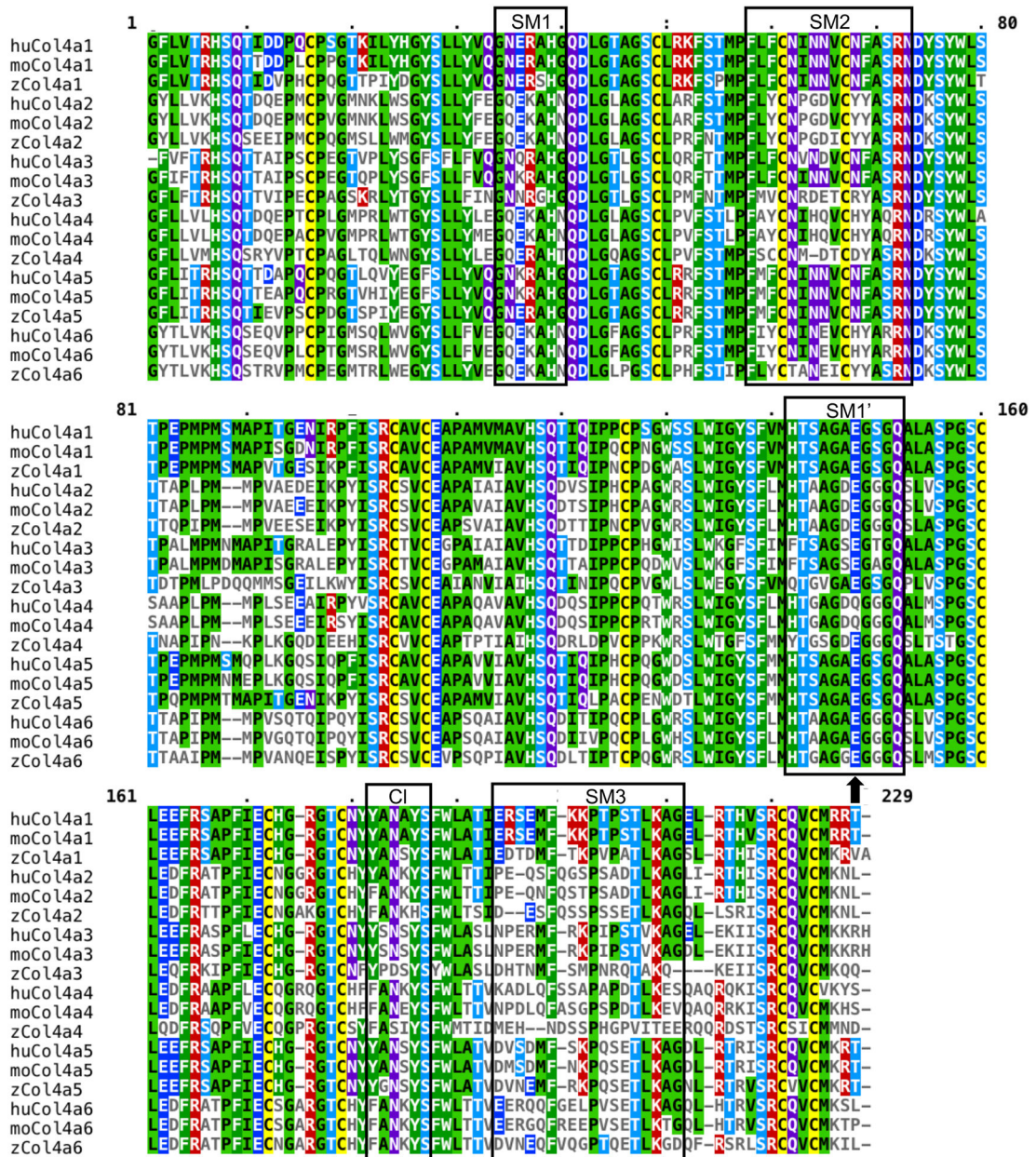


Figure 2: Sequence comparison of the human, mouse, and zebrafish NC1 domain.

Comparison of amino acid sequence shows conservation among different species.

Subdomains of interests are boxed, structural motifs SM1 (hairpin $\beta 3$ – $\beta 4$), SM2 (hairpin $\beta 6$ – $\beta 7$), SM1' (hairpin $\beta 3'$ – $\beta 4'$), SM3 ($\beta 9$ and preceding loop), CI (chloride ion). Arrow points to amino acid residue similar in zCo14a4 and hCo14a1 and hCo14a2 NC1 domain. hu, human; mo, mouse; z, zebrafish.

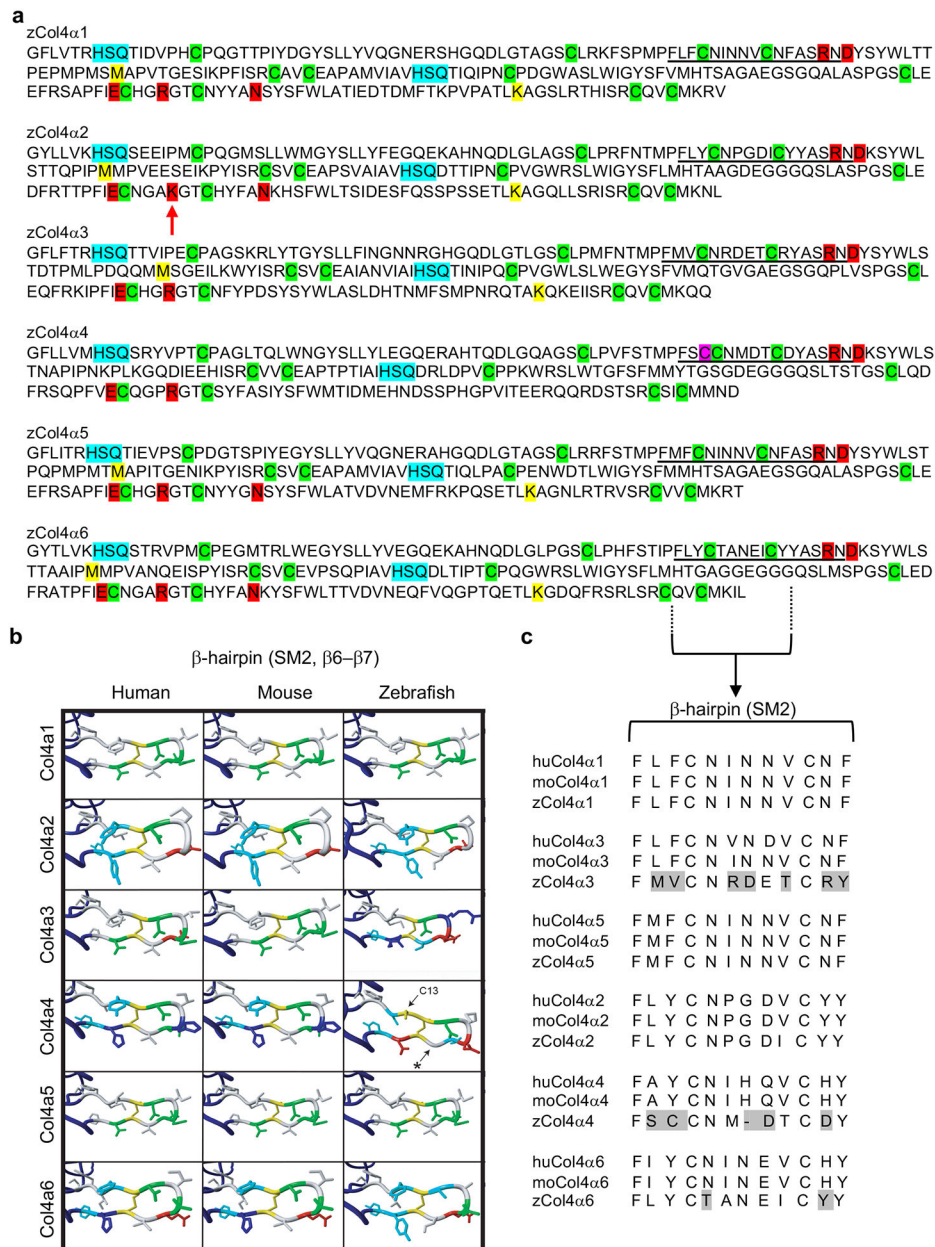


Figure 3: Structure of the NC1 domain of zebrafish collagen IV α -chain proteins.

a. Sequence comparison of the zebrafish α 1- α 6 NC1 domains. Cyan: HSQ tripeptide that begin each NC1. Green: twelve conserved cysteine residues. Magenta: additional cysteine (Cysteine 13, Cys-13) in zCol4 α 4. Red: residues involved in salt bridges formed between the R76 D78, E175-R179, and N187. The red arrow points to the R179K substitution in zCol4 α 2. Yellow: M93 and K211 residues involved in the peroxidase-catalyzed sulfilimine bond formation. **b.** Representation of the 3D molecular remodeling of human, mouse and zebrafish NC1 domain β -hairpin SM2 subdomain. The additional cysteine (C13) in NC1 β -hairpin of zCol4 α 4 is indicated by the arrow. The asterisk points to the area where the zebrafish protein backbone is shorter than the human and mouse homologues. **c.** Amino acid sequence alignment for human, mouse, and zebrafish NC1 domain, highlighting key

residues in the SM2 (hairpin $\beta 6$ – $\beta 7$) subdomain that indicates less conservation between the $\alpha 3$, $\alpha 4$ and $\alpha 6$ chains compared to the $\alpha 1$, $\alpha 2$ and $\alpha 5$ chains.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

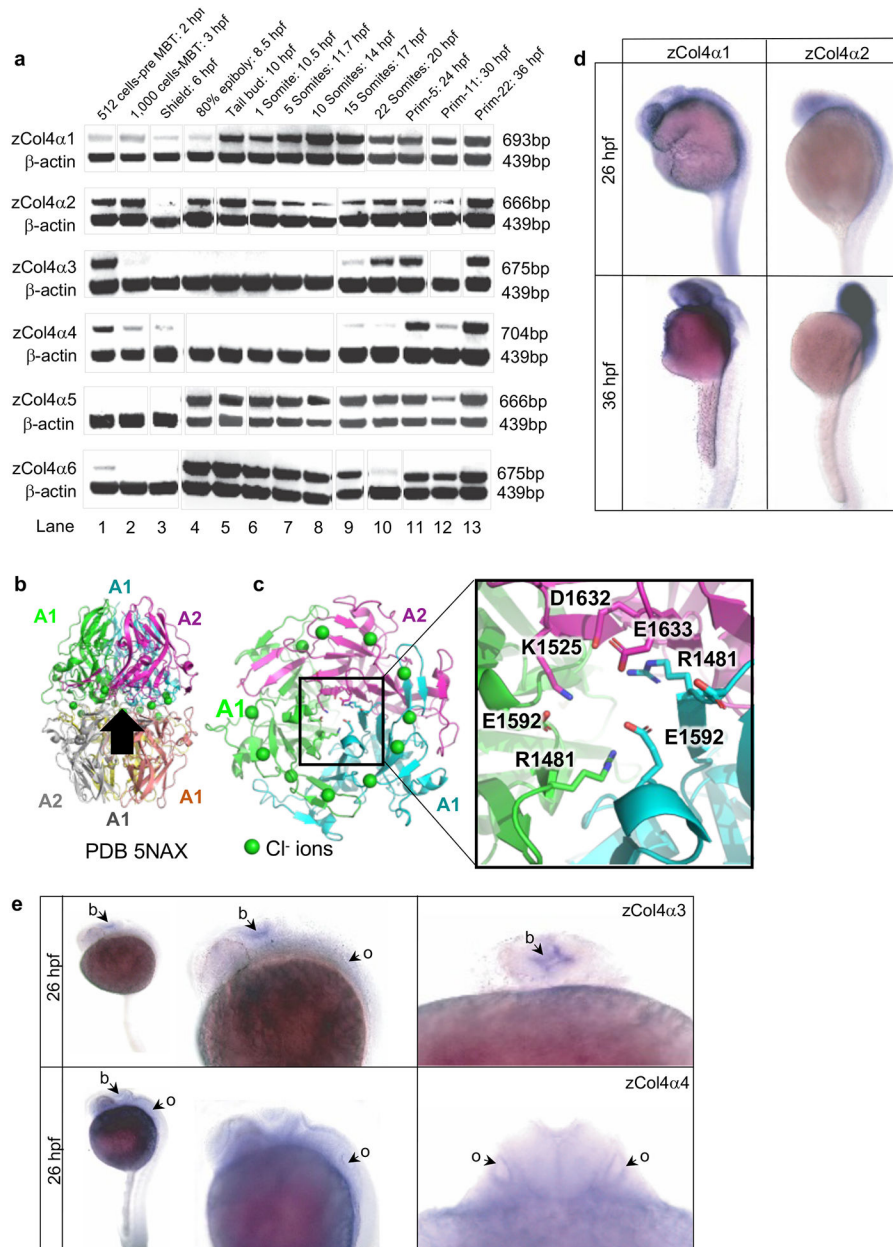


Figure 4: Embryonic expression of the six collagen IV α chains in zebrafish.

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 ($\alpha 1$ and $\alpha 2$ chains); post-gastrulation ($\alpha 5$ and $\alpha 6$ chains); and late-onset expression ($\alpha 3$ and $\alpha 4$ chains). Expression before 6 hours post fertilization (hpf) stage is maternal. **b.** Cartoon schematic of NC1 heterohexamer of $\alpha 1$ and $\alpha 2$ NC1 chains (PDB 5NAX). Arrow shows viewpoint for **c.** **c.** Bottom-up view of heterotrimer. Inset is close-up view of salt bridges that help form the heterotrimer interface. In the $\alpha 1$ and $\alpha 2$ chains, a glutamate (E1592 in $\alpha 1$ and E1633 in $\alpha 2$) forms a salt bridge with a basic residue (R1481 in $\alpha 1$ and K1525 in $\alpha 2$). **d.** In situ hybridization for zebrafish $\alpha 1$ and $\alpha 2$ chains at 26 hpf and 36 hpf. **e.** In situ

hybridization for zebrafish $\alpha 3$ and $\alpha 4$ chains at 26 hpf, dorsal and lateral views. The $\alpha 3$ and $\alpha 4$ chain are localized in the developing brain (b), with $\alpha 4$ chain also detected in the otic vesicle (o). hpf, hours post fertilization, MBT, mid-blastula transition.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

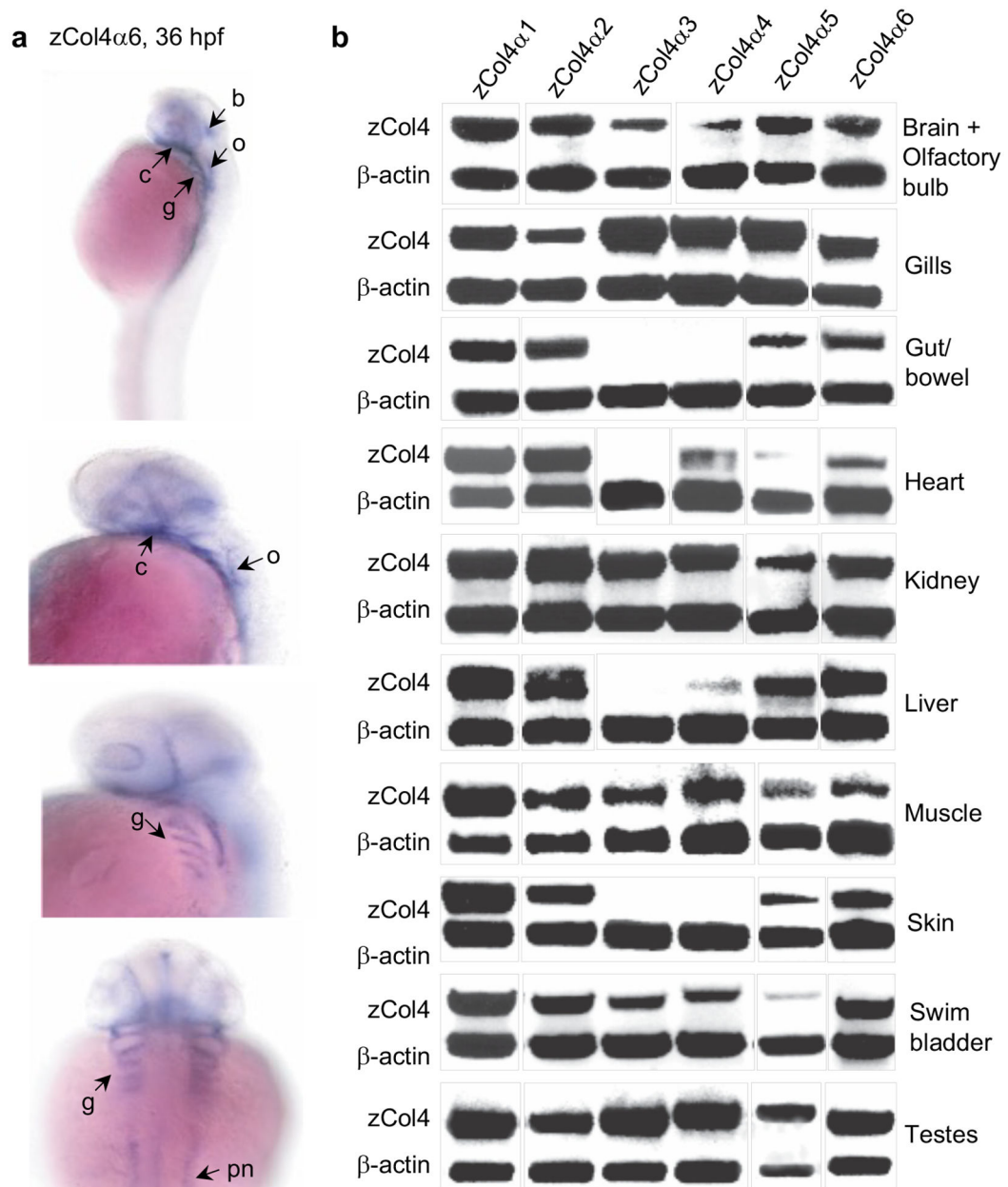


Figure 5: Tissue distribution of the six collagen IV α chains in zebrafish development.

a. In situ hybridization for zebrafish α 6 chains at 36 hpf in the brain (b), otic vesicle (o), gills (g), developing cardiac tissue (c), and the pronephric duct (pn) of the future kidney.

b. 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 α 1, α 2, α 5, and α 6 chains only. Liver and heart both display expression of all α chains except the α 3 chain.

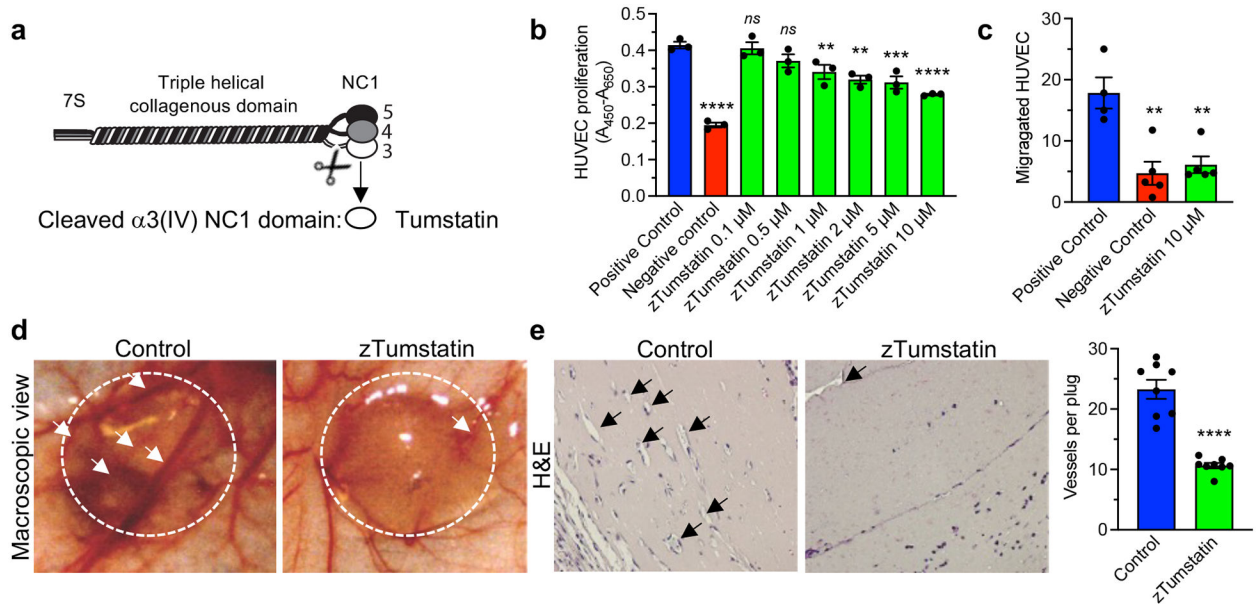


Figure 6: Recombinant zTumstain, the NC1 domain of the $\alpha 3$ chain of type IV collagen, shows conserved antiangiogenic activity on human endothelial cells.

a. Schematic depiction of Col4 $\alpha 3/\alpha 4/\alpha 5$ protomer and cleavage site releasing Tumstain ($\alpha 3$ NC1 domain). **b.** Relative HUVECs proliferation when treated with the indicated concentration of eukaryotically-produced zTumstain, $n = 3$ 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 zTumstain MatrigelTM plug (encircled) and blood vessels (arrows). **e.** Representative H&E section of control and zTumstain MatrigelTM 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, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.