

## OPINION

# The triple helix of collagens – an ancient protein structure that enabled animal multicellularity and tissue evolution

Aaron L. Fidler<sup>1,2,\*</sup>, Sergei P. Boudko<sup>1,2,\*</sup>, Antonis Rokas<sup>3</sup> and Billy G. Hudson<sup>1,2,4,5,6,7,8,9,‡</sup>

## ABSTRACT

The cellular microenvironment, characterized by an extracellular matrix (ECM), played an essential role in the transition from unicellularity to multicellularity in animals (metazoans), and in the subsequent evolution of diverse animal tissues and organs. A major ECM component are members of the collagen superfamily – comprising 28 types in vertebrates – that exist in diverse supramolecular assemblies ranging from networks to fibrils. Each assembly is characterized by a hallmark feature, a protein structure called a triple helix. A current gap in knowledge is understanding the mechanisms of how the triple helix encodes and utilizes information in building scaffolds on the outside of cells. Type IV collagen, recently revealed as the evolutionarily most ancient member of the collagen superfamily, serves as an archetype for a fresh view of fundamental structural features of a triple helix that underlie the diversity of biological activities of collagens. In this Opinion, we argue that the triple helix is a protein structure of fundamental importance in building the extracellular matrix, which enabled animal multicellularity and tissue evolution.

**KEY WORDS:** Cell biology, Collagen, Evolution, Extracellular matrix, Multicellularity, Triple helix

## Introduction

The extracellular matrix (ECM) played an essential role during the transition from unicellular organisms to multicellular animals (metazoans). The ECM comprises a basement membrane (BM) that underlies epithelia cells, and an interstitial matrix (IM) that is positioned between cells in the intercellular spaces and undergoes continuous controlled remodeling (Hynes, 2012; Bonnans et al., 2014; Nelson and Bissell, 2006; Inman et al., 2015). Yet, a major gap in cell biology is to understand how cells generate and interact with the ECM (Sherwood, 2015; Jayadev and Sherwood, 2017).

The collagen superfamily of proteins is a major component of ECMs, which – in vertebrates – comprises 28 types (I–XXVIII) that

are derived from a total of 46  $\alpha$ -chains across the superfamily (Fig. 1) (Ricard-Blum, 2011; Kadler et al., 2007; Ricard-Blum and Ruggiero, 2005). Invertebrates generally contain collagen IV, XV or XVIII, some fibrillar collagens, as well as some fibril-associated collagens with interrupted triple helices (FACITs) (Fidler et al., 2014, 2017; Fahey and Degnan, 2010; Meyer and Moussian, 2009; Boot-Handford and Tuckwell, 2003; Whittaker et al., 2006; Kadler et al., 2007). Among these collagens, type IV is the evolutionarily most ancient, based on recent studies of non-bilaterian animals (sponges, ctenophores, placozoans and cnidarians) and unicellular groups (Fidler et al., 2017; Grau-Bove et al., 2017) (Fig. 1).

Collagens are the most abundant protein in the human body (Kadler et al., 2007; Shoulders and Raines, 2009). They occur as diverse supramolecular assemblies, ranging from networks to fibrils, and broadly function in structural, mechanical and organizational roles that define tissue architecture and influence cellular behavior (Shoulders and Raines, 2009; Ricard-Blum, 2011; Ricard-Blum and Ruggiero, 2005). Defects in collagens underlie the cause of almost 40 human genetic diseases, affecting numerous organs and tissues in millions of people worldwide (summarized in Table 1).

Disease pathogenesis typically involves genetic alterations of the triple helix, a unique structure that is a hallmark feature common to all collagens. The triple helix bestows exceptional mechanical resistance to tensile forces and a capacity to bind a plethora of macromolecules. Yet, there is a gap in our current knowledge in understanding the mechanisms of how a triple helix encodes and utilizes information in building supramolecular assemblies on the outside of cells. Here, we present collagen IV, the most ancient of the collagen superfamily, and argue that it is ideally suited to serve as an archetype for investigating and describing core functions of a triple helix.

## The triple helix – assembly and structural features that encode information

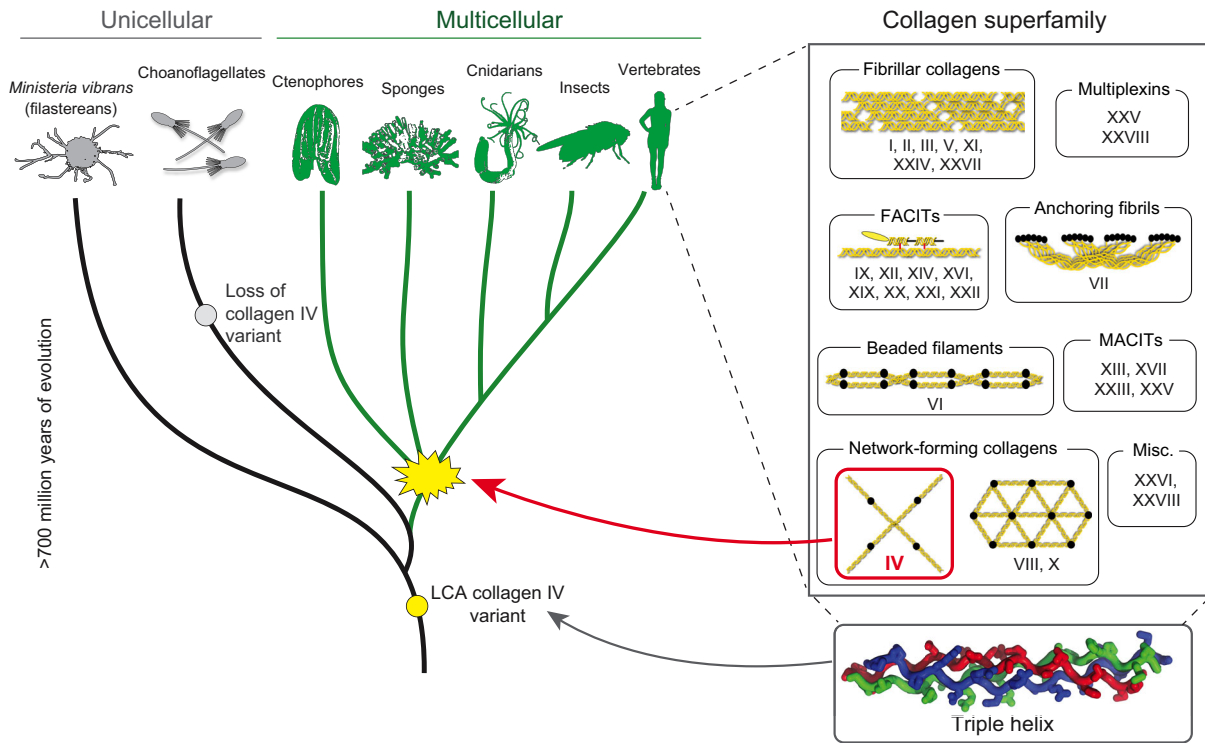
The chemical structure of the triple helix was determined through the seminal work of structural biologists and chemists over the last century (see Box 1 in supplementary material). Its unique structure bestows upon collagens an exceptional mechanical resistance to tensile forces and a plethora of organizing information for building an ECM (Fig. 2). The triple helix presents all residues, except glycine (Gly), on its surface, which is the most economical and robust way to encode binding motifs of any protein structure. Moreover, the triple helix exhibits extensive post-translational modifications (PTMs), such as hydroxylation, glycosylation and phosphorylation, adding – in tandem – a secondary layer of information in addition to its amino acid (aa) code (Yamauchi and Shiiba, 2008). These PTMs confer even more diversity with tissue-specific and disease-specific variations, even amongst identical types of collagen (Pokidysheva et al., 2013). Furthermore, additional collagen modifications are mediated by specific extracellular enzymes, such as peroxidase and lysyl oxidase-like

<sup>1</sup>Department of Medicine, Division of Nephrology and Hypertension, Vanderbilt University Medical Center, Nashville, TN, 37232, USA. <sup>2</sup>Center for Matrix Biology, Vanderbilt University Medical Center, Nashville, TN, 37232, USA. <sup>3</sup>Department of Biological Sciences, Vanderbilt University Medical Center, Nashville, TN, 37232, USA. <sup>4</sup>Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN, 37232, USA. <sup>5</sup>Department of Medical Education and Administration, Vanderbilt University Medical Center, Nashville, TN, 37232, USA. <sup>6</sup>Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN, 37232, USA. <sup>7</sup>Department of Biochemistry, Vanderbilt University Medical Center, Nashville, TN, 37232, USA. <sup>8</sup>Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, 37232, USA. <sup>9</sup>Vanderbilt Institute of Chemical Biology, Vanderbilt University Medical Center, Nashville, TN, 37232, USA.

\*These authors contributed equally to this work

‡Author for correspondence (billy.hudson@vanderbilt.edu)

© A.L.F., 0000-0002-2519-8864; A.R., 0000-0002-7248-6551; B.G.H., 0000-0002-5420-4100



**Fig. 1. Collagen IV is the evolutionarily most ancient of the vertebrate collagen superfamily.** The collagen superfamily has 28 members in vertebrates, each comprising three of 46  $\alpha$ -chains – the basis of the diverse suprastructures distributed across different tissues – invertebrates have less members. A hallmark feature of all collagens is the triple helix, which is characterized by three intertwined polypeptide chains. A collagen IV-like gene probably first appeared in the last common ancestor (LCA) to filastereans, choanoflagellates and animals. The phylogenetic distribution suggests that collagen IV played a critical role in the transition of unicellular organisms to multicellular animals. FACITs, fibril-associated collagens with interrupted triple helices; MACITs, membrane-associated collagens with interrupted triple helices; Misc., miscellaneous.

proteins (LOXLs) for crosslinking, and Goodpasture antigen-binding protein (COL4A3BP, hereafter referred to as GPBP) and other extracellular kinases for phosphorylation (Bhave et al., 2012; Añazco et al., 2016; Revert et al. 1995, Raya et al. 1999; Yalak and Olsen, 2015). Non-enzymatic modifications, such as glycation, oxidation or chlorination, add even more complexity (Brown et al., 2015). Together, these modifications may serve as regulatory mechanisms on the outside of cells that may instruct cell behavior and influence tissue architecture and stability (Yalak and Olsen, 2015; Pedchenko et al., 2010).

To fully appreciate the capacity and versatility of the triple helix, one should consider the building blocks (polypeptide chains) of this structure. These chains are simple ‘rope-like’ structures built from proline residues. In aqueous solutions, such a polyproline chain adopts a polyproline type II (PPII) helix in a left-handed conformation (Adzhubei et al., 2013) (Fig. 2A). A unique feature of this structure is that it is non-extensible (Okuyama et al., 1981). Such simple ‘ropes’ occur in the extracellular space of plants (Lampert, 1974) and have been identified as extensions of triple helices within mini-collagens of cnidarians (e.g. *Hydra*) (Holstein et al., 1994), where they perform structural and mechanical roles. To allow for the formation of a triple helix from this simple rope, every third proline residue of the PPII helix is replaced with a glycine residue – the smallest aa residue as it lacks a side chain. This endows the structure with flexibility and sufficient space to tightly pack three chains (Fig. 2A) that are each a left-handed helix, wind around each other with a shift in one residue, so that the smaller glycine residues are buried inside the triple helix, whereas only proline residues are exposed on the surface (Fig. 2A) (Ramachandran and Kartha, 1954;

Rich and Crick, 1955; Okuyama et al., 1981). Moreover, as these three chains wind together forming a triple helix, each individual left-handed chain adopts a right-handed superhelix. The unique role of the glycine residues in the packing of the collagen triple helix explains the adverse effects of mutations at these positions, as any other residue in place of glycine will distort its tight packing (Bella et al., 1994).

Three main changes are achieved upon the transition from a single polyproline helix to a triple helix: (i) higher bending rigidity, (ii) a less accessible chain backbone that is, thus, less prone to proteolysis and, (iii) the ability to essentially accept any aa in place of the proline residues at position X and Y without any significant destabilization to the triple helix structure (Fig. 2B). Moreover, proline and other aa within the triple helix remain accessible to solvent and, thus, their numerous known PTMs are possible without disturbing the native helix structure. For example, probably the most important PTM is the 4(R)-hydroxylation of proline residues position Y, because it substantially increases the thermal stability of the collagen triple helix (Sakakibara et al., 1973). Although all proline and hydroxyproline residues are essential for the stability of a PPII helix, ~65% of aa at these positions can vary in the native triple helix, while still maintaining its stability. Thus, the triple helix, in contrast to a single PPII helix, confers additional capacity to specify binding partners.

Once assembled from three superhelical  $\alpha$ -chains, the collagen triple helix provides several distinct sites to tether macromolecules through three modes of binding motifs (single, double and triple  $\alpha$ -chain) with further three levels of variations to these modes (mode 1, 2 and 3) (Fig. 2B). The three binding modes operate by utilizing

**Table 1. Human genetic diseases/disorders due to mutation(s) in collagens and experimental collagen knockout studies**

| Collagen type                         | Human genetic diseases/disorders due to mutation(s) in collagens  | Loss of protein   |
|---------------------------------------|---|---|
| <b>I</b>                              | (G) <b>Osteogenesis imperfecta</b> (Wenstrup et al., 1988; Vogel et al., 1987; Peltonen et al., 1980; Lamande et al., 1989; Labhard et al., 1988; Cohn et al., 1986; Baldwin et al., 1989)<br>(N) <b>Ehlers–Danlos syndrome</b> (Wirtz et al., 1987; Malfait et al., 2006; Nuytinck et al., 2000)<br><b>Caffey disease</b> (Gensure et al., 2005) (for all, see <a href="http://www.le.ac.uk/ge/collagen/">http://www.le.ac.uk/ge/collagen/</a> )   | Embryonic lethality during organogenesis before E15, embryonic growth arrest at E11.5–E12 (Schnieke et al., 1983; Harbers et al., 1984)   |
| <b>II</b>                             | (G) <b>Spondyloepiphyseal dysplasias</b> (Lee et al., 1989; Tiller et al., 1990; Chan et al., 1993; Jurgens et al., 2015; Barat-Houari et al., 2016)<br>(G) <b>Spondyloepimetaphyseal dysplasias</b> (Tiller et al., 1995; Sulko et al., 2005)<br>(G) <b>Achondrogenesis 2</b> (Vissing et al., 1989; Mortier et al., 1995, 2000; Körkkö et al., 2000; Chan et al., 1995; Barat-Houari et al., 2016)<br>(G) <b>Legg–Calvé–Perthes disease</b> (Miyamoto et al., 2007)<br>(G) <b>Kniest dysplasia</b> (Wilkin et al., 1994; Winterpacht et al., 1996)<br>(G) <b>Avascular necrosis of femoral head</b> (Liu et al., 2005; Kannu et al., 2011)<br><b>Platyspondylic lethal skeletal dysplasia, Torrance type</b> (Nishimura et al., 2004; Zankl et al., 2005)<br><b>Spondyloperipheral dysplasia</b> (Zankl et al., 2005; Nishimura et al., 2004)<br>(G) <b>Stickler syndrome 1</b> (Richards et al., 2000, 2010; McAlinden et al., 2008; Körkkö et al., 1993; Barat-Houari et al., 2016)<br>(N) <b>Osteoarthritis with mild chondrodysplasia</b> (Ala-Kokko et al., 1990)<br>(N) <b>Czech dysplasia</b> (Williams et al., 1993; Tzschach et al., 2008; Matsui et al., 2009; Barat-Houari et al., 2016) | Perinatal lethality, abnormal skeleton, abnormal cranium, abnormal respiratory system (Li et al., 1995a).<br>IIA isoform-specific knockout: prenatal lethality, holoprosencephal loss of head tissues and mid-facial structures (Leung et al., 2010). |
| <b>III</b>                            | (G) <b>Ehlers–Danlos syndrome, vascular types</b> (Narcisi et al., 1994; Tromp et al., 1989a,b; Pepin et al., 1993) ( <a href="http://www.le.ac.uk/ge/collagen/">http://www.le.ac.uk/ge/collagen/</a> )   | Embryonic lethality before E9.5, pre- to mid-gestation lethality (Smith et al., 2011)   |
| <b>IV</b> ( $\alpha1\alpha2$ )        | (G) <b>Cerebral small vessel disease</b> (Gould et al., 2006; Shah et al., 2012; Sibon et al., 2007)<br>(G) <b>Hereditary angiopathy with nephropathy aneurysms and muscle cramps (HANAC)</b> (Plaisier et al., 2007, 2010)<br>(G) <b>Porencephaly 1</b> (Breedveld et al., 2006; de Vries et al., 2009; Gould et al., 2005; Yoneda et al., 2013)<br>(G) <b>Porencephaly 2</b> (Yoneda et al., 2012)<br>(G) <b>Tortuosity of retinal arteries</b> (Zenteno et al., 2014)<br>(G) <b>Schizencephaly</b> (Yoneda et al., 2013)<br>(N) <b>Intracerebral hemorrhage</b> (Jeanne et al., 2012; Weng et al., 2012)   | Embryonic lethality at E10.5–E11.5 (Poschl et al., 2004)  |
| <b>IV</b> ( $\alpha3\alpha4\alpha5$ ) | (G) <b>Alport syndrome</b> (Barker et al., 1990; Boye et al., 1998; Heidet et al., 2001; Kashtan, 2017; Knebelmann et al., 1992; Mehta and Jim, 2017; Mochizuki et al., 1994; Nagel et al., 2005; Renieri et al., 1992; Zhou et al., 1991; Van Der Loop et al., 2000; Zhou et al., 1992) ( <a href="http://databases.lovd.nl/shared/genes/COL4A5">http://databases.lovd.nl/shared/genes/COL4A5</a> )<br>(G) <b>Thin basement membrane nephropathy (TBMD)</b> (Badenas et al., 2002; Buzza et al., 2003; Lemmink et al., 1996; Kashtan, 2004)  | Decreased survival rate, premature death, proteinuria, glomerulosclerosis, glomerulonephritis, renal interstitial fibrosis, hearing loss (Cosgrove et al., 1996, 1998; Rheault et al., 2004; LeBleu et al., 2010)                                     |
| <b>IV</b> ( $\alpha5\alpha6$ )        | (G) <b>X-linked deafness 6</b> (Rost et al., 2014)  | No abnormal phenotype detected (Fox et al., 2007)   |
| <b>V</b>                              | (G) <b>Ehlers–Danlos syndrome</b> (Watanabe et al., 2016; Symoens et al., 2009; Richards et al., 1998; Michalickova et al., 1998; Malfait et al., 2005; Giunta and Steinmann, 2000; Giunta et al., 2002; De Paepe et al., 1997) ( <a href="http://www.le.ac.uk/ge/collagen/">http://www.le.ac.uk/ge/collagen/</a> )   | Embryonic lethality at E10–E12 (Wenstrup et al., 2004; Park et al., 2015)   |
| <b>VI</b>                             | (G) <b>Bethlem myopathy 1</b> (Scacheri et al., 2002; Pepe et al., 1999; Pan et al., 1998; Luciola et al., 2005; Lampe et al., 2005; Jöbsis et al., 1996; Baker et al., 2007)<br>(G) <b>Ullrich congenital muscular dystrophy 1</b> (Lampe et al., 2005; Kawahara et al., 2007; Giusti et al., 2005; Baker et al., 2005)<br><b>Dystonia 27</b> (Zech et al., 2015)  | Abnormalities in morphology of I muscle, impaired skeletal muscle contractility, myopathy (Bonaldo et al., 1998)  |
| <b>VII</b>                            | (G) <b>Epidermolysis bullosa dystrophica, multiple types</b> (Shinkuma, 2015; Fine et al., 1989; Christiano et al., 1993, 1994, 1995, 1996; Woodley et al., 2004)   | Postnatal lethality within the first week from complications of blistering (Heinonen et al., 1999)  |
| <b>VIII</b>                           | (N) <b>Corneal dystrophies</b> (Biswas et al., 2001)  | Dysgenesis of the anterior segment of the eye (Hopfer et al., 2005)   |
| <b>IX</b>                             | (G) <b>Multiple epiphyseal dysplasia 2, 3 and 6</b> (Paassilta et al., 1999; Jeong et al., 2014; Holden et al., 1999; Czarny-Ratajczak et al., 2001)<br><b>Stickler Syndrome 4 and 5</b> (Van Camp et al., 2006; Baker et al., 2011)<br>(N) <b>Intervertebral disc disease</b> (Paassilta et al., 2001; Annunen et al., 1999b)  | Osteoarthritis, abnormalities in morphology of bone and cartilage, abnormal bone healing, impaired hearing (Fassler et al., 1994; Asamura et al., 2005; Ting et al., 1999)  |

Continued

Table 1. Continued

| Collagen type     | Human genetic diseases/disorders due to mutation(s) in collagens   | Loss of protein  |
|-------------------|--|--|
| <b>X</b>          | <b>Schmid metaphyseal chondrodysplasia</b> (Wallis et al., 1994; McIntosh et al., 1994; Bateman et al., 2005)  | Abnormal spleen morphology, abnormal hematopoietic system (Rosati et al., 1994)  |
| <b>XI</b>         | (G) <b>Stickler syndrome 2</b> (Annunen et al., 1999a; Richards et al., 1996, 2010),<br><b>Marshall syndrome</b> (Annunen et al., 1999a)<br>(G) <b>Fibrochondrogenesis 1</b> (Tompson et al., 2010)<br>(G) <b>Otospondylomegapiphyseal dysplasia (OSMED)</b> (Melkonieni et al., 2000; Pihlajamaa et al., 1998; Sirko-Osadsa et al., 1998; Vikkula et al., 1995)<br>(G) <b>Deafness</b> (Chakchouk et al., 2015; Chen et al., 2005; McGuirt et al., 1999)<br><b>Fibrochondrogenesis 2</b> (Tompson et al., 2012) | Neonatal lethality; abnormalities in cartilage of limbs, ribs, mandible and trachea (Li et al., 1995b)   |
| <b>XII</b>        | <b>Ullrich congenital muscular dystrophy 2</b> (Zou et al., 2014)<br>(G) <b>Bethlem myopathy 2</b> (Zou et al., 2014; Hicks et al., 2014)  | Increased perinatal lethality, skeletal abnormalities, decreased body weight, abnormal skeletal muscle morphology and physiology (Izu et al., 2011; Zou et al., 2014)  |
| <b>XIII*</b>      | <b>Congenital myasthenic syndrome type 19</b> (Logan et al., 2015)   | Deletion of cytosolic and transmembrane regions, abnormal skeletal muscle morphology, muscle degeneration, myopathy, impaired exercise endurance, increased susceptibility to injury (Kvist et al., 2001)    |
| <b>XIV</b>        | unknown  | Decreased tendon stiffness, decreased skin tensile strength (Ansoorge et al., 2009)  |
| <b>XV</b>         | unknown  | Abnormal muscle morphology, increased muscle regeneration, myopathy (Eklund et al., 2001)  |
| <b>XVI</b>        | unknown  | unknown  |
| <b>XVII</b>       | (G) <b>Generalized atrophic benign epidermolysis bullosa</b> (McGrath et al., 1996; Schumann et al., 1997; Tasanen et al., 2000a,b; Wu et al., 2002)<br><b>Epithelial recurrent erosion dystrophy</b> (Jonsson et al., 2015)   | Significant postnatal lethality, sterility, postnatal growth retardation, blistering, abnormal renal glomerulus morphology (Hurskainen et al., 2012; Nishie et al., 2007)                                    |
| <b>XVIII</b>      | <b>Knobloch syndrome</b> (Aldahmesh et al., 2013; Sertie et al., 2000)   | Eye abnormalities, intracranial hemorrhage, abnormalities in morphology of brain, cranium and renal glomerulus; abnormal lipid metabolism (Bishop et al., 2010; Fukai et al., 2002; Utraiainen et al., 2004) |
| <b>XIX</b>        | unknown  | ~95% postnatal lethality, dysphagia, abnormal esophagus morphology, weight loss (Sumiyoshi et al., 2004)   |
| <b>XX to XXIV</b> | unknown  | unknown  |
| <b>XXV</b>        | (G) <b>Congenital cranial dysinnervation disorder</b> (Shinwari et al., 2015)  | Neonatal lethality; abnormalities in morphology of motor neurons, skeletal muscles; respiratory failure; cyanosis (Tanaka et al., 2014)  |
| <b>XXVI</b>       | unknown  | unknown  |
| <b>XXVII*</b>     | (G) <b>Steel syndrome</b> (Gonzaga-Jauregui et al., 2015)  | 87-residue deletion of collagenous domain; neonatal lethality; chondrodysplasia; respiratory failure at birth; abnormalities in morphology of lung and skeleton; decreased fetal size (Plumb et al., 2011)   |
| <b>XXVIII</b>     | unknown  | unknown  |

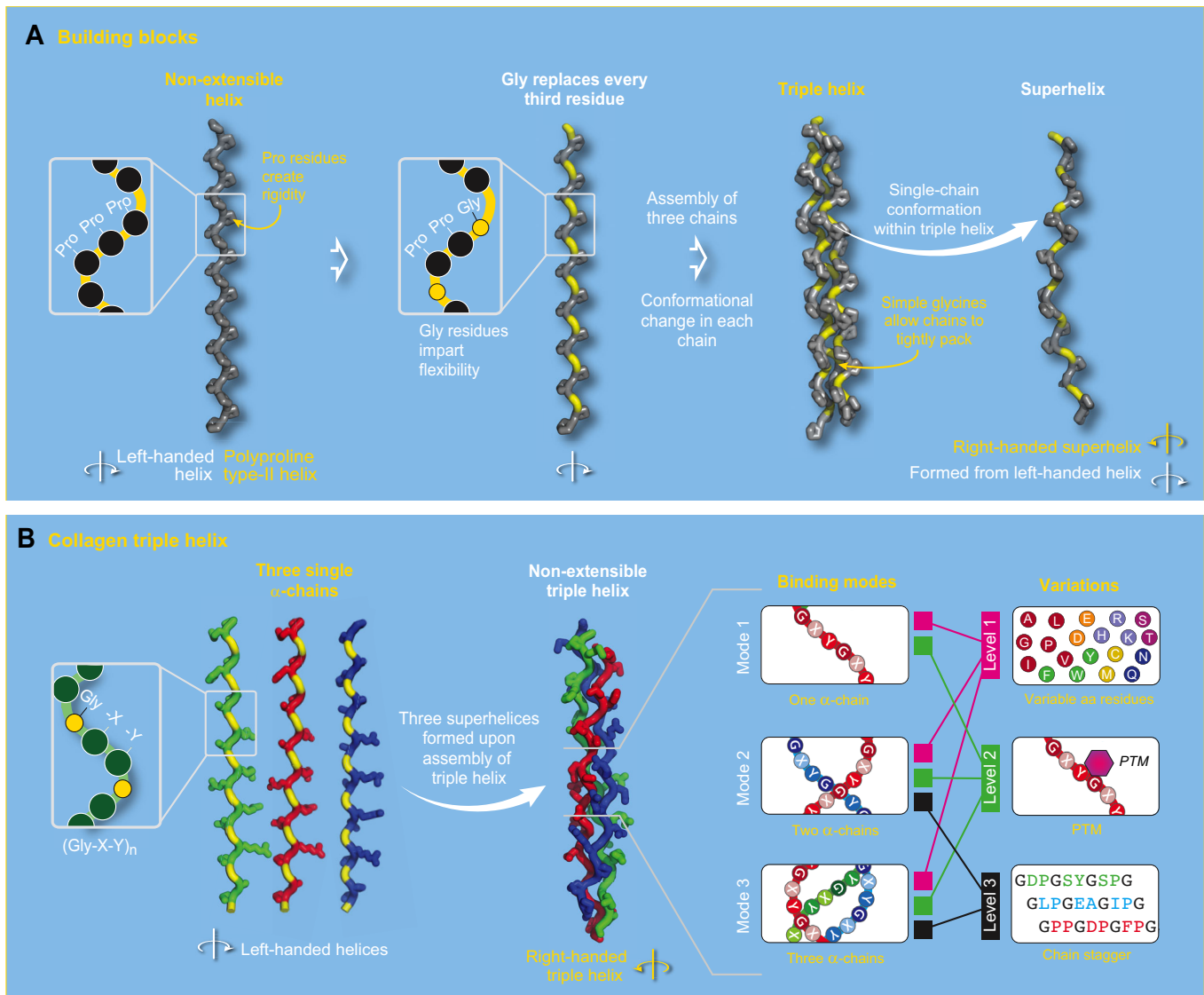
G, Gly and non-Gly mutations reported in triple helix; N, only non-Gly mutations reported in triple helix; \*deletion mutants for collagens XIII and XXVII

one, two or all three distinct chains of the triple helix in different combinations to directly bind a molecule (Fig. 2B; modes 1–3). The three levels of variation to these modes add additional diversity to binding specificity on top of involving either one, two or three chains. The first level stems from the variation that is provided by the  $\leq 20$  possible aa residues that can occupy the variable positions X and Y of each tripeptide; this results in extensive structural variability across the three chains, as well as within each Gly-X-Y tripeptide of the same chain (Fig. 2B; level 1). The second level confers even more diversity because of the proclivity of collagens to become post-translationally modified and change their structure (Fig. 2B; level 2). The third level utilizes chain staggering that occurs during triple helix assembly between either two or three chains and enables new combinations of chains that can give rise to additional binding motifs (Fig. 2B; level 3).

In addition to these binding specificities, the triple helix also possesses other structural features that underlie its biological function. It is non-stretchable along its longitudinal axis,

providing great tensile strength to withstand all physiological mechanical loads and stresses in our body. It also confers resistance to proteases – making collagens some of the most long-lived proteins – as well as to a wide range of pH values in order to withstand adverse conditions (Steven, 1965; Hafter and Höermann, 1963; Grant and Alburn, 1960; Drake et al., 1966; Uzawa et al., 1998; Pokidysheva et al., 2013; Hudson et al., 2017; Eyre et al., 2011). Furthermore, the triple helix exhibits variable lengths among collagen suprastructures, such as networks, fibrils and filaments (Ricard-Blum, 2011). For example, the triple helix length of fibrillar collagens results from multiple duplications of exons that are either 54 or 45 base pairs in length and encode Gly-X-Y (Yamada et al., 1980; Exposito et al., 2010). Moreover, the terminal and internal incorporation of non-triple helical sequences (e.g. interruptions) in all collagen types extends their functional capacity (Ricard-Blum, 2011; Khoshnoodi et al., 2006). Collectively, the binding motifs (see above) together with these features of a collagen triple helix underlie the diversity of





**Fig. 2. Building a collagen triple helix and encoding information.** (A) The building blocks (polypeptide chains) of the triple helix are left-handed polyproline type-II helices, i.e. non-extensible structures that have structural and mechanical roles in the ECM of plants and some animals. Replacement of every third proline (Pro) residue with a glycine (Gly) residue results in increased flexibility. (B) Three left-handed superhelices wind together and pack tightly owing to these Gly residues, thereby forming a right-handed triple helix. Once associated, the non-extensibility of the structure is restored, and the combination of three chains results in three binding modes (one, two or three chains) with three levels of variation to these modes, i.e. level 1: 20 variable aa residues, level 2: post-translational modifications (PTMs) and, level 3: chain stagger. Together, these specify the numerous possible binding motifs that are positioned along the length of the triple helix of any of the 28 collagen types to bind various macromolecules.

biological activities of the diverse supramolecular assemblies of collagens.

**The roles of the ECM and collagen IV in the transition of the Urmetazoan to multicellular animals**

The last common ancestor to animals, the Urmetazoan, almost certainly reproduced by gametogenesis, underwent gastrulation during early development, had the ability for cells to differentiate both during development and as stem cells, and comprised an epithelial layer of cells forming the body of the animal – features that are still fundamental to extant animals (Richter and King, 2013; King and Rokas, 2017). Importantly, these cellular activities ultimately required the invention of an ECM to provide a substrate for attachment and signaling cues to regulate cell behavior and function in tissue genesis and homeostasis (Abedin and King, 2010). The appearance of a specialized form of ECM, the

BM, coincided with the transition to multicellularity. The BM functions in several cellular activities, including migration, adhesion, delineation of apical–basal polarity and modulation of differentiation during development (Petersen et al., 1992; Lukashev and Werb, 1998; Daley and Yamada, 2013; Hynes, 2009, 2012; Yurchenco, 2011; Ozbek et al., 2010; Henry and Campbell, 1998).

Importantly, understanding the makeup of BMs between different animals sheds light on the functions of proteins in the evolution of animal multicellularity and tissues. BMs are composed of numerous proteins, vary between animals. The BM of bilaterian animals (e.g. human, fly, *C. elegans*, sea urchin) is composed of several proteins, including collagen IV, laminin, perlecan, nidogen, fibronectin, proteoglycans, peroxidasin, GPBP, and collagens XV and XVIII (Hynes, 2012; Yurchenco, 2011; Jayadev and Sherwood, 2017). Potentially, there are many more components of the BM (Chew and Lennon, 2018). Among these, collagen IV is a major

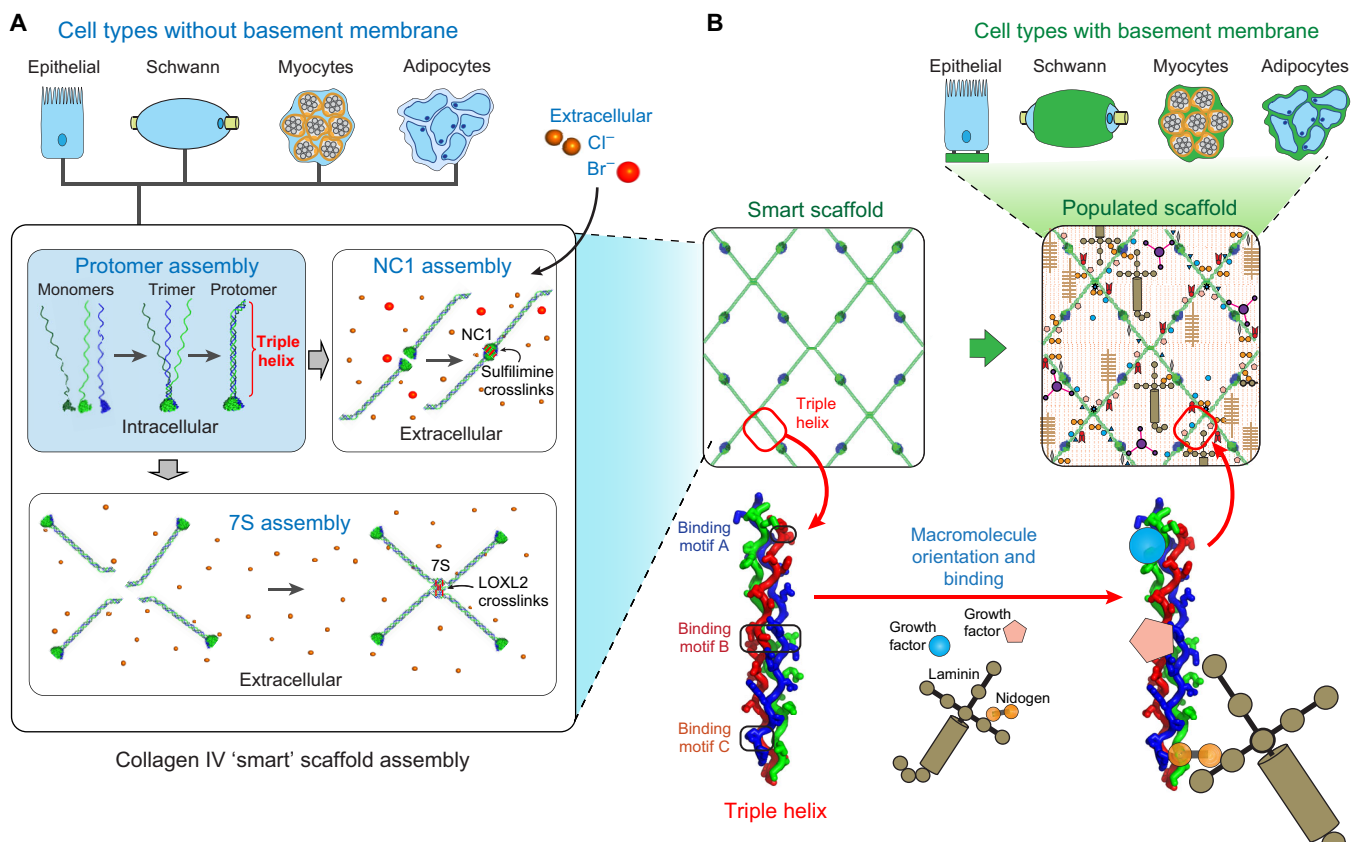
component that is conserved among animal phyla (Fidler et al., 2014). In a recent study, we described collagen IV at the origins of animal multicellularity and in tissue evolution, as revealed by close examination of sponges, ctenophores and other non-bilaterally symmetrical animals (Fidler et al., 2017). Ctenophores and sponges have been established as the two most likely candidates to be the sister-groups to the rest of animals, based on phylogenetic analyses of genomic and transcriptomic data and cell-type evolution (Ryan et al., 2013; Moroz et al., 2014; Pisani et al., 2015; Whelan et al., 2015; Telford et al., 2016; King and Rokas, 2017; Feuda et al., 2017). Our genomic analysis of the extracellular matrix components within the ctenophores and Homoscleromorph sponges revealed a BM ‘toolkit’ consisting of just collagen IV and laminin (Fidler et al., 2017). However, the demosponges, a sponge class, lack both laminin and classic collagen IV but do contain spongins, which are short collagen IV variants (Exposito et al., 1991; Aouacheria et al., 2006; Fidler et al., 2017). The order in which these variants and collagen IV first appeared is still unknown. Despite containing fewer BM proteins than bilaterians, many sponges and ctenophores form classic BMs (Fidler et al., 2017; Boute et al., 1996; Leys et al., 2009).

Importantly, comparison of BM components of animals with those of unicellular lineages is key to determining their importance during the transition to multicellularity. Within choanoflagellates,

the closest relatives of animals, no complete ECM proteins exist; yet, domains that are characteristic of laminins and short collagenous Gly-X-Y repeats are present (King et al., 2008; Fahey and Degnan, 2012; Fidler et al., 2017). Interestingly, a recent study reported the discovery of a collagen IV-like gene in the filasterean, *Ministeria vibrans*, a unicellular lineage that diverged prior to choanoflagellates and animals (Grau-Bove et al., 2017). This finding indicates that collagen IV has a premetazoan ancestry and a function for single cells. Collectively, these findings suggest that collagen IV played a role in the transition from unicellular organisms to multicellular animals (Grau-Bove et al., 2017; Fidler et al., 2017) (Fig. 1). Therefore, we consider collagen IV as an archetype of collagens to describe the fundamental features of a triple helix that underlie biological functions.

**The triple helix of collagen IV scaffolds**

Collagen IV forms a network that functions as a scaffold within BMs (Fig. 3). These scaffolds provide tensile strength, connect adjacent cells and organize supramolecular protein assemblies that are able to influence cell behavior (Wang et al., 2008; Emsley et al., 2000; Parkin et al., 2011; Cummings et al., 2016; Vanacore et al., 2009). Network assembly begins with the intracellular formation of triple-helical protomers comprising three  $\alpha$ -chains (Brown et al.,



**Fig. 3. The triple helix of collagen IV scaffolds encodes information for the tethering of macromolecules.** (A) Assembly begins with collagen IV monomers ( $\alpha$ -chains) transcribed and assembled into protomers inside of many cell types. Protomers are composed of three intertwining  $\alpha$ -chains that form a triple helix. On the outside of cells, the NC1- and 7S-domains direct the assembly of protomers into network structures of higher order crosslinked by sulfilimine and lysyl oxidase-like protein 2 (LOXL2). The higher  $Cl^-$  concentration of the ECM induces NC1-domain-directed oligomerization of protomers that form ‘smart’ scaffolds. (B) Once assembled, collagen IV networks function as smart scaffolds, bestowing BMs that underlie and surround cells of several capabilities. Vast amounts of structural information is encoded in motifs located at specific sites along the surface of the triple helix to tether macromolecules. The mode of variation is based on 20 variable aa residues on a single chain, or the combination of one, two or three chains, post-translational modifications (PTMs) and chain stagger (see Fig. 2). The tethering at specific sites spatially organizes molecules along the triple helix, resulting in a populated scaffold within the BM that provides tensile strength to tissues, and influences cell behavior, adhesion and migration during tissue development and regeneration. Partly modified, with permission (*Protein Science*) from Brown et al., 2017.

2017). Protomer assembly is directed and regulated by the non-collagenous (NC)1 recognition modules, which are located at the C-terminus of each  $\alpha$ -chain; this is followed by the twisting together of the collagenous domains – the Gly-X-Y repeats – into a triple helix. In mammals, three distinct protomers ( $\alpha 1(I)$ ,  $\alpha 2(I)$  and  $\alpha 1(II)$ ) are formed from six genetically distinct  $\alpha$ -chains ( $\alpha 1(I)$ – $\alpha 6(I)$ ), thereby forming three distinct networks (Khoshnoodi et al., 2008). Once secreted into the extracellular space, protomers adjoin via their NC1 domains and the N-terminal 7S domains (Fig. 3A). NC1-domain association is mediated by extracellular  $Cl^-$ , which activates a molecular switch that enables adjacent protomers to adjoin two NC1 domain trimers (Fig. 3A) (Cummings et al., 2016). 7S domains assemble into a complex of four independent protomer 7S domains (Añazco et al., 2016). Upon association mediated through NC1- and 7S-domains, the collagen IV networks are reinforced through covalent crosslinks at both the NC1- and 7S-domain interfaces. NC1-domain hexamers are stabilized through sulfilimine (–S=N–) double bonds, crosslinks that are induced by peroxidasin (PXDN) – an animal heme peroxidase embedded in the BM – and by  $Br^-$  (Vanacore et al., 2009; Bhave et al., 2012; McCall et al., 2014). Concurrently, 7S dodecamers are crosslinked by lysyl oxidase-like protein 2 (LOXL2) (Añazco et al., 2016).

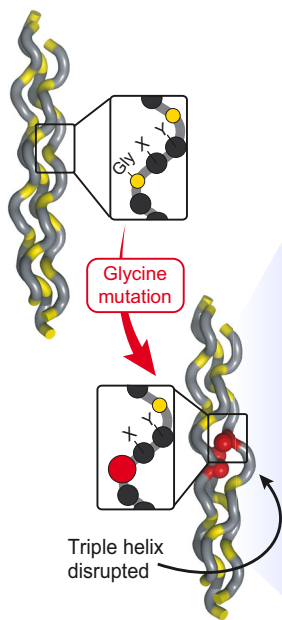
Collagen IV networks function as smart scaffolds, bestowing BMs with several capabilities (Fig. 3B). Via the triple helix, scaffolds tether different extracellular molecules, i.e. laminins, proteoglycans, perlecan, nidogens, growth factors and extracellular enzymes (such as peroxidasin and lysyl oxidase-like protein 2) (Parkin et al., 2011; Hynes, 2012; Bhave et al., 2012; Añazco et al.,

2016). The information for the tethering of these molecules is encoded at sites within the triple helix, and depends on the 20 variable aa residues within a single chain, or a combination of one, two or three chains, chain stagger and post-translational modifications (see Fig. 2). The tethering of these molecules at specific sites along the triple helix spatially organizes binding partners (Fig. 3B), which – in turn – forms a diverse multi-protein complex that represents a BM. The distribution of binding partners within a BM is not a static arrangement, and can be dynamically regulated throughout early development and beyond (Inman et al., 2015; Jones-Paris et al., 2016). The resulting scaffold, populated with macromolecules bound to the triple helix, provides tensile strength to tissues, attaches to cells through cell-surface receptors and influences cell behavior in tissue development, function and regeneration (Eble et al., 1993; Emsley et al., 2000; Khoshnoodi et al., 2008; Valiathan et al., 2012; Fu et al., 2013).

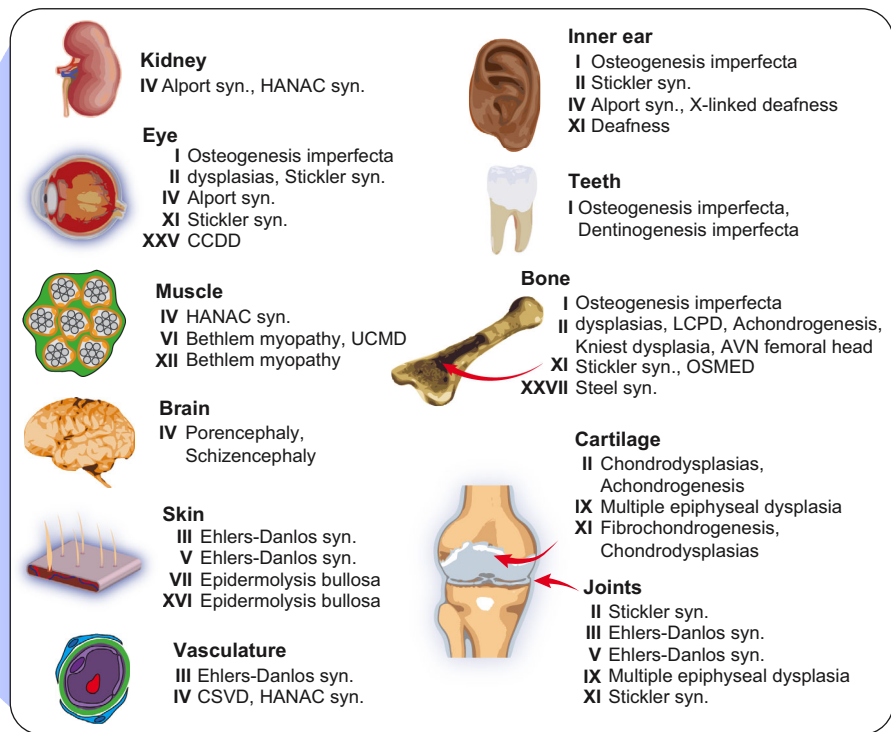
**Evidence for essentiality of the triple helix**

The biological importance of the triple helix is displayed in several ways. It is an ancient structure that is conserved between animals and is expressed ubiquitously in their ECMs. The triple helix is a common protein structure of numerous and distinct collagen suprastructures with diverse biological activities, including the network-forming collagens (IV, VIII, X), the FACITs (IX, XII, XIV, XVI, XIX, XX, XXI, XXII), fibrils (I, II, III, V, XI, XXIV, XXVII), anchoring fibrils (VII) and beaded filaments (VI) (Fig. 1). There are almost 40 diseases wherein mutations of glycine residues affect multiple tissues and organs in millions of people (Fig. 4A,B,

**A** Glycine mutations in triple helix



**B** Diseases and disorders caused by glycine mutations in triple helix



**Fig. 4. Mutations in the triple helix can result in genetic diseases.** (A) A single missense mutation of a glycine to another residue results in disruption of the triple helix's function. (B) Glycine mutations are responsible for a number of genetic diseases involving the kidney, teeth, muscle, joints, cartilage, brain, skin, vasculature, bone, inner ear and eye in humans (see Table 1). (B) Diseases and disorders due to replacement of glycine residue(s) in the triple helix of collagens. The type of collagen is indicated in bold Roman numerals. AVN, avascular necrosis of femoral head; CCDD, congenital cranial dysinnervation disorder; CSVD, cerebral small-vessel disease; HANAC, hereditary angiopathy with nephropathy, aneurysms and muscle cramps; UCMD, Ullrich congenital muscular dystrophy; LCPD, Legg–Calvé–Perthes disease; OSMED, otospondylomegaepiphyseal dysplasia; syn, syndrome.

Table 1). Glycine residues are crucial for the structural integrity of the triple helix (see Figs 2 and 4A); therefore, mutated collagen molecules can assemble into faulty fibrils, networks, and other assemblies can cause tissue dysfunction.

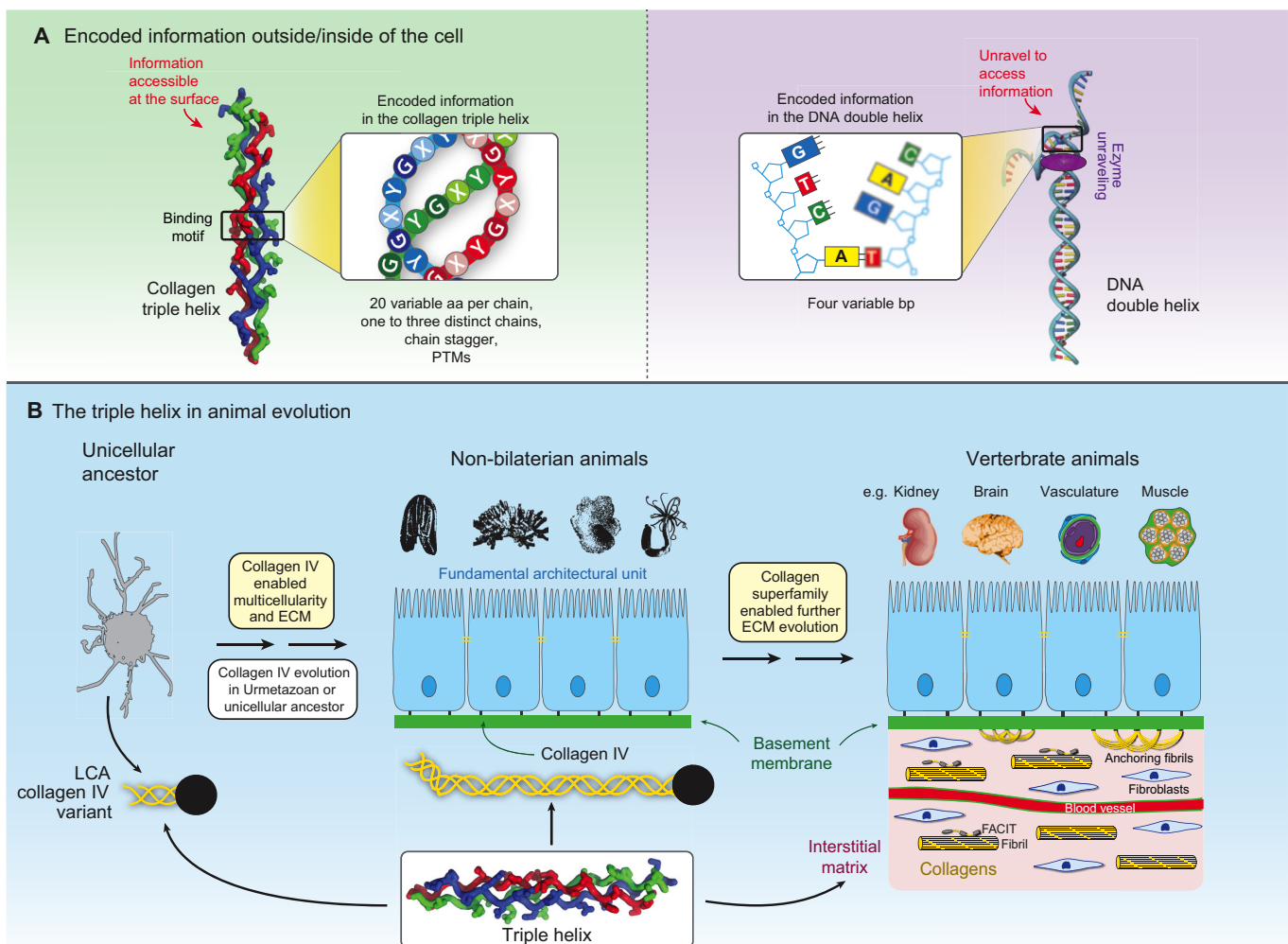
As examples for such glycine mutations, osteogenesis imperfecta (OI; also known as brittle bone disease), and Alport syndrome are two collagen-dependent genetic disorders that are well-studied (Fig. 4C, Table 1). For OI, over 800 mutations in collagen I have been described (Marini et al., 2007; Forlino et al., 2011; Forlino and Marini, 2016). Approximately 80% are glycine mutations that occur in the triple helix (Forlino et al., 2011). However, such mutations, depending on the nature of substitution as well as its location, lead to different degrees of post-translational modifications and structural destabilization of the triple helix. For example, substitutions in the first 200 residues of collagen I are non-lethal, whereas there are two regions (helix positions 691–823 and 910–964), in which substitutions can cause lethality because they align with main ligand-binding sites for integrins, matrix metalloproteinases, fibronectin and cartilage oligomeric matrix protein (Marini et al., 2007). In Alport syndrome, the collagen IV scaffold is mutated, which leads to progressive organ failure in kidney, ear and eye (Williamson, 1961; Hudson, 2004; Hudson et al., 2003; Cosgrove

et al., 2007; Cosgrove and Liu, 2017; Kashtan, 1993; Chew and Lennon, 2018). Over 1700 mutations have been found to occur in the collagen IV scaffold that is composed of  $\alpha 3$ ,  $\alpha 4$  and  $\alpha 5$  chains. Of these, ~85% are glycine substitutions that are located in the triple helix (Hertz et al., 2012) (Table 1).

In summary, the many collagen diseases that involve glycine mutations directly demonstrate the essentiality of the triple helix for tissue architecture and function. Its essentiality is further supported experimentally by collagen knockout studies in mice (Table 1). The respective mouse phenotypes include developmental lethality upon knockout of collagen types I, II, III, IV, V, VII, XII, XVII, XIX, XXVI and XXVII, muscle deformities in that of types VI, XII, XIII, XV and XXV, and bone and cartilage deformities in that of types IX and XI (see references in Table 1).

**Conclusions and perspectives**

The triple helix is unique among all other protein structures – globular or fibrous – in its capacity to encode vast amounts of information that is available on its surface for utilization on the outside of cells (Fig. 5A). The triple helix, arranged in various patterns forming diverse supramolecular scaffolds, tethers and



**Fig. 5. The fundamental importance of a triple helix in enabling animal multicellularity and tissue evolution.** (A) The unique structure and vast encoding properties of the collagen triple helix outside the cell (left) evokes an analogy to the DNA double helix inside the cell (right). (B) The triple helix protein structure was present in unicellular organisms and was co-opted in the form of collagen IV, enabling the transition to multicellular animals. The triple helix was also adapted as a key feature of all members of the diverse collagen superfamily in the ECM, enabling tissue evolution. LCA, last common ancestor.



spatially organizes macromolecules, thus providing tensile strength to tissues and influencing cell behavior. This unique structure of the triple helix with its encoded degree of information evokes an analogy to the DNA double helix (Fig. 5A).

The biological importance of the triple helix is also evident from the almost 40 genetic diseases and its ubiquitous presence in animals. The triple helix was co-opted in the form of collagen IV to enable the evolutionary transition from unicellular organisms to multicellular animals, and the triple helix was also adapted to give rise to all the other members of the diverse collagen superfamily, thereby enabling the evolution of tissues and organs (Fig. 5B). Thus, the triple helix represents a fundamental protein structure that nature adapted for building an extracellular matrix.

There are many provocative and unanswered questions regarding the function of triple helices in suprastructures and their dysfunction in diseases. These include: (i) What are the unknown sites encoded in the triple helix for binding partners as exemplified by those in collagen I fibrils and collagen IV networks (Di Lullo et al., 2002; Parkin et al., 2011)? (ii) How is information in the triple helix used to assemble suprastructures (Orgel et al., 2011)? (iii) What are the mechanisms of how triple helices within suprastructures influence cell function? (iv) What is the impact of PTMs on the structure and function of the triple helix? (v) How do genetic mutations in the triple helix cause tissue dysfunction? (vi) How do genetic backgrounds affect phenotype variations? (vii) What are the mechanisms for the function of the triple helix in the transition from unicellular organisms to multicellular animals.

To answer such fundamental questions, collagen IV is an ideal archetype because it is the most ancient of the collagens, and it is present in unicellular organisms and non-bilaterian animals (Fig. 5B). Furthermore, recent studies have revealed that, in some organisms, collagen IV also occurs in the absence of a BM, such as in certain ctenophores and sponges, placozoans and the unicellular filasterean *Ministeria vibrans* (Fidler et al., 2017; Grau-Bove et al., 2017; Schierwater et al., 2009). Moreover, a recent study of *Drosophila* development provided evidence that, in the absence of BM, collagen IV has a role in intercellular adhesion and pro-growth signaling (Dai et al., 2017; Zajac and Horne-Badovinac, 2017). Together, these recent studies clearly indicate that collagen IV can have a direct role in influencing cell behavior, outside of the BM; thus, the core functions of the collagen triple helix, as well as its role in the transition from unicellular organisms to multicellular animals, can be addressed by comparative studies in these organisms. Such knowledge may provide insights into unknown roles of collagens in cell biology, disease pathogenesis and evolution of animals.

#### Acknowledgements

This article is the culmination of a ten-year Aspinaut expedition to the dawn of the animal kingdom in search of the evolutionary origin of collagen IV. The expedition involved over 100 middle school, high school, undergraduate and graduate students from disadvantaged backgrounds, and was championed by Aaron Fidler, Julie Hudson and Billy Hudson. We are grateful for the contributions of all of these students in the discovery that collagen IV occurred in all animals, and the fundamental importance of the triple helical protein structure in animal evolution as presented in this capstone article.

#### Competing interests

The authors declare no competing or financial interests.

#### Funding

This work was supported by National Institutes of Health DK18381 to B.G.H., National Science Foundation DEB-1442113 to Antonis Rokas, March of Dimes Foundation March of Dimes Prematurity Research Center Ohio Collaborative to Antonis Rokas, and the Aspinaut Program to Julie K. Hudson and B.G.H. Deposited in PMC for release after 12 months.

#### Supplementary information

Supplementary information available online at <http://jcs.biologists.org/lookup/doi/10.1242/jcs.203950.supplemental>

#### References

- Abedin, M. and King, N.** (2010). Diverse evolutionary paths to cell adhesion. *Trends Cell Biol.* **20**, 734-742.
- Adzhubei, A. A., Sternberg, M. J. E. and Makarov, A. A.** (2013). Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.* **425**, 2100-2132.
- Ala-Kokko, L., Baldwin, C. T., Moskowitz, R. W. and Prockop, D. J.** (1990). Single base mutation in the type II procollagen gene (COL2A1) as a cause of primary osteoarthritis associated with a mild chondrodysplasia. *Proc. Natl. Acad. Sci. USA* **87**, 6565-6568.
- Aldahmesh, M. A., Khan, A. O., Mohamed, J. Y., Levin, A. V., Wuthisiri, W., Lynch, S., Mccreery, K. and Alkuraya, F. S.** (2013). No evidence for locus heterogeneity in Knobloch syndrome. *J. Med. Genet.* **50**, 565-566.
- Añazco, C., López-Jiménez, A. J., Rafi, M., Vega-Montoto, L., Zhang, M.-Z., Hudson, B. G. and Vanacore, R. M.** (2016). Lysyl oxidase-like-2 cross-links collagen IV of glomerular basement membrane. *J. Biol. Chem.* **291**, 25999-26012.
- Annunen, S., Körkkö, J., Czarny, M., Warman, M. L., Brunner, H. G., Kaariainen, H., Mulliken, J. B., Tranebjaerg, L., Brooks, D. G., Cox, G. F. et al.** (1999a). Splicing mutations of 54-bp exons in the COL11A1 gene cause Marshall syndrome, but other mutations cause overlapping Marshall/Stickler phenotypes. *Am. J. Hum. Genet.* **65**, 974-983.
- Annunen, S., Paassilta, P., Lohiniva, J., Perala, M., Pihlajamaa, T., Karppinen, J., Tervonen, O., Kroger, H., Lahde, S., Vanharanta, H. et al.** (1999b). An allele of COL9A2 associated with intervertebral disc disease. *Science* **285**, 409-412.
- Ansoorge, H. L., Meng, X., Zhang, G., Veit, G., Sun, M., Klement, J. F., Beason, D. P., Soslowsky, L. J., Koch, M. and Birk, D. E.** (2009). Type XIV collagen regulates fibrillogenesis: PREMATURE COLLAGEN FIBRIL GROWTH AND TISSUE DYSFUNCTION IN NULL MICE. *J. Biol. Chem.* **284**, 8427-8438.
- Aouacheria, A., Geourjon, C., Aghajari, N., Navratil, V., Deleage, G., Lethias, C. and Exposito, J.-Y.** (2006). Insights into early extracellular matrix evolution: spongin short chain collagen-related proteins are homologous to basement membrane type IV collagens and form a novel family widely distributed in invertebrates. *Mol. Biol. Evol.* **23**, 2288-2302.
- Asamura, K., Abe, S., Imamura, Y., Aszodi, A., Suzuki, N., Hashimoto, S., Takumi, Y., Hayashi, T., Fässler, R., Nakamura, Y. et al.** (2005). Type IX collagen is crucial for normal hearing. *Neuroscience* **132**, 493-500.
- Badenas, C., Praga, M., Tazon, B., Heidet, L., Arrondel, C., Armengol, A., Andres, A., Morales, E., Camacho, J. A., Lens, X. et al.** (2002). Mutations in the COL4A4 and COL4A3 genes cause familial benign hematuria. *J. Am. Soc. Nephrol.* **13**, 1248-1254.
- Baker, N. L., Morgelin, M., Peat, R., Goemans, N., North, K. N., Bateman, J. F. and Lamande, S. R.** (2005). Dominant collagen VI mutations are a common cause of Ullrich congenital muscular dystrophy. *Hum. Mol. Genet.* **14**, 279-293.
- Baker, N. L., Mörgelin, M., Pace, R. A., Peat, R. A., Adams, N. E., Gardner, R. J. M. K., Rowland, L. P., Miller, G., De Jonghe, P., Ceulemans, B. et al.** (2007). Molecular consequences of dominant Bethlem myopathy collagen VI mutations. *Ann. Neurol.* **62**, 390-405.
- Baker, S., Booth, C., Fillman, C., Shapiro, M., Blair, M. P., Hyland, J. C. and Ala-Kokko, L.** (2011). A loss of function mutation in the COL9A2 gene causes autosomal recessive Stickler syndrome. *Am. J. Med. Genet. A* **155A**, 1668-1672.
- Baldwin, C. T., Constantinou, C. D., Dumars, K. W. and Prockop, D. J.** (1989). A single base mutation that converts glycine 907 of the alpha 2(I) chain of type I procollagen to aspartate in a lethal variant of osteogenesis imperfecta. The single amino acid substitution near the carboxyl terminus destabilizes the whole triple helix. *J. Biol. Chem.* **264**, 3002-3006.
- Barat-Houari, M., Sarabay, G., Gatinois, V., Fabre, A., Dumont, B., Genevieve, D. and Toutou, I.** (2016). Mutation update for COL2A1 gene variants associated with type II collagenopathies. *Hum. Mutat.* **37**, 7-15.
- Barker, D. F., Hostikka, S. L., Zhou, J., Chow, L. T., Oliphant, A. R., Gerken, S. C., Gregory, M. C., Skolnick, M. H., Atkin, C. L. and Tryggvason, K.** (1990). Identification of mutations in the COL4A5 collagen gene in Alport syndrome. *Science* **248**, 1224-1227.
- Bateman, J. F., Wilson, R., Freddi, S., Lamandé, S. R. and Savarirayan, R.** (2005). Mutations of COL10A1 in Schmid metaphyseal chondrodysplasia. *Hum. Mutat.* **25**, 525-534.
- Bella, J., Eaton, M., Brodsky, B. and Berman, H. M.** (1994). Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**, 75-81.
- Bhave, G., Cummings, C. F., Vanacore, R. M., Kumagai-Cresse, C., Ero-Tolliver, I. A., Rafi, M., Kang, J.-S., Pedchenko, V., Fessler, L. I., Fessler, J. H. et al.** (2012). Peroxidase forms sulfilimine chemical bonds using hypohalous acids in tissue genes. *Nat. Chem. Biol.* **8**, 784-790.
- Bishop, J. R., Passos-Bueno, M. R., Fong, L., Stanford, K. I., Gonzales, J. C., Yeh, E., Young, S. G., Bensadoun, A., Witztum, J. L., Esko, J. D. et al.** (2010). Deletion of the basement membrane heparan sulfate proteoglycan type XVIII collagen causes hypertriglyceridemia in mice and humans. *PLoS ONE* **5**, e13919.

- Biswas, S., Munier, F. L., Yardley, J., Hart-Holden, N., Perveen, R., Cousin, P., Sutphin, J. E., Noble, B., Batterbury, M., Kilty, C. et al. (2001). Missense mutations in COL8A2, the gene encoding the alpha2 chain of type VIII collagen, cause two forms of corneal endothelial dystrophy. *Hum. Mol. Genet.* **10**, 2415-2423.
- Bonaldo, P., Braghetta, P., Zanetti, M., Piccolo, S., Volpin, D. and Bressan, G. M. (1998). Collagen VI deficiency induces early onset myopathy in the mouse: an animal model for Bethlem myopathy. *Hum. Mol. Genet.* **7**, 2135-2140.
- Bonnans, C., Chou, J. and Werb, Z. (2014). Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **15**, 786-801.
- Boot-Handford, R. P. and Tuckwell, D. S. (2003). Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest. *BioEssays* **25**, 142-151.
- Boute, N., Exposito, J.-Y., Boury-Esnault, N., Vacelet, J., Noro, N., Miyazaki, K., Yoshizato, K. and Garrone, R. (1996). Type IV collagen in sponges, the missing link in basement membrane ubiquity. *Biol. Cell* **88**, 37-44.
- Boye, E., Mollet, G., Forestier, L., Cohen-Solal, L., Heidet, L., Cochat, P., Grünfeld, J.-P., Palcoux, J.-B., Gubler, M.-C. and Antignac, C. (1998). Determination of the genomic structure of the COL4A4 gene and of novel mutations causing autosomal recessive Alport syndrome. *Am. J. Hum. Genet.* **63**, 1329-1340.
- Breedveld, G., De Coo, I. F., Lequin, M. H., Arts, W. F., Heutink, P., Gould, D. B., John, S. W., Oostra, B. and Mancini, G. M. (2006). Novel mutations in three families confirm a major role of COL4A1 in hereditary porencephaly. *J. Med. Genet.* **43**, 490-495.
- Brown, K. L., Darris, C., Rose, K. L., Sanchez, O. A., Madu, H., Avance, J., Brooks, N., Zhang, M.-Z., Fogo, A., Harris, R. et al. (2015). Hypohalous acids contribute to renal extracellular matrix damage in experimental diabetes. *Diabetes* **64**, 2242-2253.
- Brown, K. L., Cummings, C. F., Vanacore, R. and Hudson, B. (2017). Building collagen IV smart scaffolds on the outside of cells. *Protein Sci.* **26**, 2151-2161.
- Buzza, M., Dagher, H., Wang, Y. Y., Wilson, D., Babon, J. J., Cotton, R. G. and Savage, J. (2003). Mutations in the COL4A4 gene in thin basement membrane disease. *Kidney Int.* **63**, 447-453.
- Chakchouk, I., Grati, M., Bademci, G., Bensaid, M., Ma, Q., Chakroun, A., Foster, J., II, Yan, D., Duman, D., Diaz-Horta, O. et al. (2015). Novel mutations confirm that COL11A2 is responsible for autosomal recessive non-syndromic hearing loss DFNB53. *Mol. Genet. Genomics* **290**, 1327-1334.
- Chan, D., Cole, W. G., Chow, C. W., Mundlos, S. and Bateman, J. F. (1995). A COL2A1 mutation in achondrogenesis type II results in the replacement of type II collagen by type I and III collagens in cartilage. *J. Biol. Chem.* **270**, 1747-1753.
- Chan, D., Taylor, T. K. and Cole, W. G. (1993). Characterization of an arginine 789 to cysteine substitution in alpha 1 (II) collagen chains of a patient with spondyloepiphyseal dysplasia. *J. Biol. Chem.* **268**, 15238-15245.
- Chen, W., Kahrizi, K., Meyer, N. C., Riazalhosseini, Y., Van Camp, G., Najmabadi, H. and Smith, R. J. (2005). Mutation of COL11A2 causes autosomal recessive non-syndromic hearing loss at the DFNB53 locus. *J. Med. Genet.* **42**, e61.
- Chew, C. and Lennon, R. (2018). Basement membrane defects in genetic kidney diseases. *Front. Pediatr.* **6**, 11.
- Christiano, A. M., Greenspan, D. S., Hoffman, G. G., Zhang, X., Tamai, Y., Lin, A. N., Dietz, H. C., Hovnanian, A. and Uitto, J. (1993). A missense mutation in type VII collagen in two affected siblings with recessive dystrophic epidermolysis bullosa. *Nat. Genet.* **4**, 62-66.
- Christiano, A. M., Rynnman, M. and Uitto, J. (1994). Dominant dystrophic epidermolysis bullosa: identification of a Gly->Ser substitution in the triple-helical domain of type VII collagen. *Proc. Natl. Acad. Sci. USA* **91**, 3549-3553.
- Christiano, A. M., Morricone, A., Paradisi, M., Angelo, C., Mazzanti, C., Cavalieri, R. and Uitto, J. (1995). A glycine-to-arginine substitution in the triple-helical domain of type VII collagen in a family with dominant dystrophic epidermolysis bullosa. *J. Invest. Dermatol.* **104**, 438-440.
- Christiano, A. M., McGrath, J. A. and Uitto, J. (1996). Influence of the second COL7A1 mutation in determining the phenotypic severity of recessive dystrophic epidermolysis bullosa. *J. Invest. Dermatol.* **106**, 766-770.
- Cohn, D. H., Byers, P. H., Steinmann, B. and Gelinus, R. E. (1986). Lethal osteogenesis imperfecta resulting from a single nucleotide change in one human pro alpha 1(I) collagen allele. *Proc. Natl. Acad. Sci. USA* **83**, 6045-6047.
- Cosgrove, D. and Liu, S. (2017). Collagen IV diseases: a focus on the glomerular basement membrane in Alport syndrome. *Matrix Biol.* **57-58**, 45-54.
- Cosgrove, D., Meehan, D. T., Grunkemeyer, J. A., Kornak, J. M., Sayers, R., Hunter, W. J. and Samuelson, G. C. (1996). Collagen COL4A3 knockout: a mouse model for autosomal Alport syndrome. *Genes Dev.* **10**, 2981-2992.
- Cosgrove, D., Samuelson, G., Meehan, D. T., Miller, C., Mcgee, J., Walsh, E. J. and Siegel, M. (1998). Ultrastructural, physiological, and molecular defects in the inner ear of a gene-knockout mouse model for autosomal Alport syndrome. *Hear. Res.* **121**, 84-98.
- Cosgrove, D., Kalluri, R., Miner, J.-H., Segal, Y. and Borza, D.-B. (2007). Choosing a mouse model to study the molecular pathobiology of Alport glomerulonephritis. *Kidney Int.* **71**, 615-618.
- Cummings, C. F., Pedchenko, V., Brown, K. L., Colon, S., Rafi, M., Jones-Paris, C., Pokydesha, E., Liu, M., Pastor-Pareja, J. C., Stothers, C. et al. (2016). Extracellular chloride signals collagen IV network assembly during basement membrane formation. *J. Cell Biol.* **213**, 479-494.
- Czarny-Ratajczak, M., Lohiniva, J., Rogala, P., Kozłowski, K., Perälä, M., Carter, L., Spector, T. D., Kolodziej, L., Seppänen, U., Glazar, R. et al. (2001). A mutation in COL9A1 causes multiple epiphyseal dysplasia: further evidence for locus heterogeneity. *Am. J. Hum. Genet.* **69**, 969-980.
- Dai, J., Ma, M., Feng, Z. and Pastor-Pareja, J. C. (2017). Inter-adipocyte adhesion and signaling by collagen IV intercellular concentrations in *Drosophila*. *Curr. Biol.* **27**, 2729-2740.e4.
- Daley, W. P. and Yamada, K. M. (2013). ECM-modulated cellular dynamics as a driving force for tissue morphogenesis. *Curr. Opin. Genet. Dev.* **23**, 408-414.
- De Paepe, A., Nuytinck, L., Hausser, I., Anton-Lamprecht, I. and Naeyaert, J. M. (1997). Mutations in the COL5A1 gene are causal in the Ehlers-Danlos syndromes I and II. *Am. J. Hum. Genet.* **60**, 547-554.
- De Vries, L. S., Koopman, C., Groenendaal, F., Van Schooneveld, M., Verheijen, F. W., Verbeek, E., Witkamp, T. D., Van Der Worp, H. B. and Mancini, G. (2009). COL4A1 mutation in two preterm siblings with antenatal onset of parenchymal hemorrhage. *Ann. Neurol.* **65**, 12-18.
- Di Lullo, G. A., Sweeney, S. M., Körkkö, J., Ala-Kokko, L. and San Antonio, J. D. (2002). Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J. Biol. Chem.* **277**, 4223-4231.
- Drake, M. P., Davison, P. F., Bump, S. and Schmitt, F. O. (1966). Action of proteolytic enzymes on tropocollagen and insoluble collagen. *Biochemistry* **5**, 301-312.
- Eble, J. A., Golbik, R., Mann, K. and Kuhn, K. (1993). The alpha 1 beta 1 integrin recognition site of the basement membrane collagen molecule [alpha 1(IV)]2 alpha 2(IV). *EMBO J.* **12**, 4795-4802.
- Eklund, L., Piihola, J., Komulainen, J., Sormunen, R., Ongvarrasopone, C., Fassler, R., Muona, A., Ilves, M., Ruskoaho, H., Takala, T. E. S. et al. (2001). Lack of type XV collagen causes a skeletal myopathy and cardiovascular defects in mice. *Proc. Natl. Acad. Sci. USA* **98**, 1194-1199.
- Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J. and Liddington, R. C. (2000). Structural basis of collagen recognition by integrin alpha2beta1. *Cell* **101**, 47-56.
- Exposito, J.-Y., Le Guellec, D., Lu, Q. and Garrone, R. (1991). Short chain collagens in sponges are encoded by a family of closely related genes. *J. Biol. Chem.* **266**, 21923-21928.
- Exposito, J.-Y., Valcourt, U., Cluzel, C. and Lethias, C. (2010). The fibrillar collagen family. *Int. J. Mol. Sci.* **11**, 407-426.
- Eyre, D. R., Weis, M., Hudson, D. M., Wu, J.-J. and Kim, L. (2011). A novel 3-hydroxyproline (3Hyp)-rich motif marks the triple-helical C terminus of tendon type I collagen. *J. Biol. Chem.* **286**, 7732-7736.
- Fahey, B. and Degnan, B. M. (2010). Origin of animal epithelia: insights from the sponge genome. *Evol. Dev.* **12**, 601-617.
- Fahey, B. and Degnan, B. M. (2012). Origin and evolution of laminin gene family diversity. *Mol. Biol. Evol.* **29**, 1823-1836.
- Fassler, R., Schnegelsberg, P. N., Dausman, J., Shinya, T., Muragaki, Y., McCarthy, M. T., Olsen, B. R. and Jaenisch, R. (1994). Mice lacking alpha 1 (IX) collagen develop noninflammatory degenerative joint disease. *Proc. Natl. Acad. Sci. USA* **91**, 5070-5074.
- Feuda, R., Dohrmann, M., Pett, W., Philippe, H., Rota-Stabelli, O., Lartillot, N., Worheide, G. and Pisani, D. (2017). Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. *Curr. Biol.* **27**, 3864-3870.e4.
- Fidler, A. L., Vanacore, R. M., Chetyrkin, S. V., Pedchenko, V. K., Bhawe, G., Yin, V. P., Stothers, C. L., Rose, K. L., McDonald, W. H., Clark, T. A. et al. (2014). A unique covalent bond in basement membrane is a primordial innovation for tissue evolution. *Proc. Natl. Acad. Sci. USA* **111**, 331-336.
- Fidler, A. L., Darris, C. E., Chetyrkin, S. V., Pedchenko, V. K., Boudko, S. P., Brown, K. L., Gray Jerome, W., Hudson, J. K., Rokas, A. and Hudson, B. G. (2017). Collagen IV and basement membrane at the evolutionary dawn of metazoan tissues. *Elife* **6**, e24176.
- Fine, J. D., Johnson, L. and Wright, T. (1989). Epidermolysis bullosa simplex superficialis. A new variant of epidermolysis bullosa characterized by subcorneal skin cleavage mimicking peeling skin syndrome. *Arch. Dermatol.* **125**, 633-638.
- Forlino, A. and Marini, J. C. (2016). Osteogenesis imperfecta. *Lancet* **387**, 1657-1671.
- Forlino, A., Cabral, W. A., Barnes, A. M. and Marini, J. C. (2011). New perspectives on osteogenesis imperfecta. *Nat. Rev. Endocrinol.* **7**, 540-557.
- Fox, M. A., Sanes, J. R., Borza, D.-B., Eswarakumar, V. P., Fassler, R., Hudson, B. G., John, S. W. M., Ninomiya, Y., Pedchenko, V., Pfaff, S. L. et al. (2007). Distinct target-derived signals organize formation, maturation, and maintenance of motor nerve terminals. *Cell* **129**, 179-193.
- Fu, H.-L., Valiathan, R. R., Arkwright, R., Sohail, A., Mihai, C., Kumarasiri, M., Mahasanen, K. V., Mobashery, S., Huang, P., Agarwal, G. et al. (2013). Discoidin domain receptors: unique receptor tyrosine kinases in collagen-mediated signaling. *J. Biol. Chem.* **288**, 7430-7437.



- Fukai, N., Eklund, L., Marneros, A. G., Oh, S. P., Keene, D. R., Tamarkin, L., Niemela, M., Ilves, M., Li, E., Pihlajaniemi, T. et al. (2002). Lack of collagen XVIII/endostatin results in eye abnormalities. *EMBO J.* **21**, 1535-1544.
- Gensure, R. C., Mäkitie, O., Barclay, C., Chan, C., Depalma, S. R., Bastepe, M., Abuzahra, H., Couper, R., Mundlos, S., Silience, D. et al. (2005). A novel COL1A1 mutation in infantile cortical hyperostosis (Caffey disease) expands the spectrum of collagen-related disorders. *J. Clin. Invest.* **115**, 1250-1257.
- Giunta, C. and Steinmann, B. (2000). Compound heterozygosity for a disease-causing G1489E [corrected] and disease-modifying G530S substitution in COL5A1 of a patient with the classical type of Ehlers-Danlos syndrome: an explanation of intrafamilial variability? *Am. J. Med. Genet.* **90**, 72-79.
- Giunta, C., Nuytinck, L., Raghunath, M., Hausser, I., De Paepe, A. and Steinmann, B. (2002). Homozygous Gly530Ser substitution in COL5A1 causes mild classical Ehlers-Danlos syndrome. *Am. J. Med. Genet.* **109**, 284-290.
- Giusti, B., Lucarini, L., Pietroni, V., Lucicoli, B., Bandinelli, B., Sabatelli, P., Squarzone, S., Petrini, S., Gartoux, C., Talim, B. et al. (2005). Dominant and recessive COL6A1 mutations in Ullrich scleroatonic muscular dystrophy. *Ann. Neurol.* **58**, 400-410.
- Gonzaga-Jauregui, C., Gamble, C. N., Yuan, B., Penney, S., Jhangiani, S., Muzny, D. M., Gibbs, R. A., Lupski, J. R. and Hecht, J. T. (2015). Mutations in COL27A1 cause Steel syndrome and suggest a founder mutation effect in the Puerto Rican population. *Eur. J. Hum. Genet.* **23**, 342-346.
- Gould, D. B., Phalan, F. C., Breedveld, G. J., Van Mil, S. E., Smith, R. S., Schimenti, J. C., Aguglia, U., Van Der Knaap, M. S., Heutink, P. and John, S. W. (2005). Mutations in Col4a1 cause perinatal cerebral hemorrhage and porencephaly. *Science* **308**, 1167-1171.
- Gould, D. B., Phalan, F. C., Van Mil, S. E., Sundberg, J. P., Vahedi, K., Massin, P., Bousser, M. G., Heutink, P., Miner, J. H., Tournier-Lasserre, E. et al. (2006). Role of COL4A1 in small-vessel disease and hemorrhagic stroke. *N. Engl. J. Med.* **354**, 1489-1496.
- Grant, N. H. and Albun, H. E. (1960). Collagen solubilization by mammalian proteinases. *Arch. Biochem. Biophys.* **89**, 262-270.
- Grau-Bove, X., Torruella, G., Donachie, S., Suga, H., Leonard, G., Richards, T. A. and Ruiz-Trillo, I. (2017). Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife* **6**, e26036.
- Haftner, R. and Höermann, H. (1963). [The effect of pepsin on the structure and fiber-forming properties of collagen]. *Hoppe Seylers Z. Physiol. Chem.* **330**, 169-181.
- Harbers, K., Kuehn, M., Delius, H. and Jaenisch, R. (1984). Insertion of retrovirus into the first intron of alpha 1(I) collagen gene to embryonic lethal mutation in mice. *Proc. Natl. Acad. Sci. USA* **81**, 1504-1508.
- Heidet, L., Arrondel, C., Forestier, L., Cohen-Solal, L., Mollet, G., Gutierrez, B., Stavrou, C., Gubler, M. C. and Antignac, C. (2001). Structure of the human type IV collagen gene COL4A3 and mutations in autosomal Alport syndrome. *J. Am. Soc. Nephrol.* **12**, 97-106.
- Heinonen, S., Mannikko, M., Klement, J. F., Whitaker-Menezes, D., Murphy, G. F. and Uitto, J. (1999). Targeted inactivation of the type VII collagen gene (Col7a1) in mice results in severe blistering phenotype: a model for recessive dystrophic epidermolysis bullosa. *J. Cell Sci.* **112**, 3641-3648.
- Henry, M. D. and Campbell, K. P. (1998). A role for dystroglycan in basement membrane assembly. *Cell* **95**, 859-870.
- Hertz, J. M., Thomassen, M., Storey, H. and Flintner, F. (2012). Clinical utility gene card for: Alport syndrome. *Eur. J. Hum. Genet.* **20**, 20.
- Hicks, D., Farsani, G. T., Laval, S., Collins, J., Sarkozy, A., Martoni, E., Shah, A., Zou, Y., Koch, M., Bonnemann, C. G. et al. (2014). Mutations in the collagen XII gene define a new form of extracellular matrix-related myopathy. *Hum. Mol. Genet.* **23**, 2353-2363.
- Holden, P., Canty, E. G., Mortier, G. R., Zabel, B., Spranger, J., Carr, A., Grant, M. E., Loughlin, J. A. and Briggs, M. D. (1999). Identification of novel pro-alpha2(I) collagen gene mutations in two families with distinctive oligo-epiphyseal forms of multiple epiphyseal dysplasia. *Am. J. Hum. Genet.* **65**, 31-38.
- Holstein, T. W., Benoit, M., Herder, G. V., David, C. N., Wanner, G. and Gaub, H. E. (1994). Fibrous mini-collagens in hydra nematocysts. *Science* **265**, 402-404.
- Hopfer, U., Fukai, N., Hopfer, H., Wolf, G., Joyce, N., Li, E. and Olsen, B. R. (2005). Targeted disruption of Col8a1 and Col8a2 genes in mice leads to anterior segment abnormalities in the eye. *FASEB J.* **19**, 1232-1244.
- Hudson, B. G. (2004). The molecular basis of Goodpasture and Alport syndromes: beacons for the discovery of the collagen IV family. *J. Am. Soc. Nephrol.* **15**, 2514-2527.
- Hudson, B. G., Tryggvason, K., Sundaramoorthy, M. and Neilson, E. G. (2003). Alport's syndrome, Goodpasture's syndrome, and type IV collagen. *N. Engl. J. Med.* **348**, 2543-2556.
- Hudson, D. M., Weis, M. A., Rai, J., Joeng, K. S., Dimori, M., Lee, B. H., Morello, R. and Eyre, D. R. (2017). P3h3-null and Sc65-null mice phenocopy the collagen lysine under-hydroxylation and cross-linking abnormality of Ehlers-Danlos syndrome type VIA. *J. Biol. Chem.* **292**, 3877-3887.
- Hurskainen, T., Moilanen, J., Sormunen, R., Franzke, C.-W., Soininen, R., Loeffek, S., Huilaja, L., Nuutinen, M., Bruckner-Tuderman, L., Autio-Harmainen, H. et al. (2012). Transmembrane collagen XVII is a novel component of the glomerular filtration barrier. *Cell Tissue Res.* **348**, 579-588.
- Hynes, R. O. (2009). The extracellular matrix: not just pretty fibrils. *Science* **326**, 1216-1219.
- Hynes, R. O. (2012). The evolution of metazoan extracellular matrix. *J. Cell Biol.* **196**, 671-679.
- Inman, J. L., Robertson, C., Mott, J. D. and Bissell, M. J. (2015). Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development* **142**, 1028-1042.
- Izu, Y., Sun, M., Zvolanek, D., Veit, G., Williams, V., Cha, B., Jepsen, K. J., Koch, M. and Birk, D. E. (2011). Type XII collagen regulates osteoblast polarity and communication during bone formation. *J. Cell Biol.* **193**, 1115-1130.
- Jayadev, R. and Sherwood, D. R. (2017). Basement membranes. *Curr. Biol.* **27**, R207-R211.
- Jeanne, M., Labelle-Dumais, C., Jorgensen, J., Kauffman, W. B., Mancini, G. M., Favor, J., Valant, V., Greenberg, S. M., Rosand, J. and Gould, D. B. (2012). COL4A2 mutations impair COL4A1 and COL4A2 secretion and cause hemorrhagic stroke. *Am. J. Hum. Genet.* **90**, 91-101.
- Jeong, C., Lee, J. Y., Kim, J., Chae, H., Park, H. I., Kim, M., Kim, O.-H., Kim, P., Lee, Y. K. and Jung, J. (2014). Novel COL9A3 mutation in a family diagnosed with multiple epiphyseal dysplasia: a case report. *BMC Musculoskelet. Disord.* **15**, 371.
- Jöbsis, G. J., Keizers, H., Vreijling, J. P., De Visser, M., Speer, M. C., Wolterman, R. A., Baas, F. and Bolhuis, P. A. (1996). Type VI collagen mutations in Bethlem myopathy, an autosomal dominant myopathy with contractures. *Nat. Genet.* **14**, 113-115.
- Jones-Paris, C. R., Paria, S., Berg, T., Saus, J., Bhavé, G., Paria, B. C. and Hudson, B. G. (2016). Embryo implantation triggers dynamic spatiotemporal expression of the basement membrane toolkit during uterine reprogramming. *Matrix Biol.* **57-58**, 347-365.
- Jonsson, F., Byström, B., Davidson, A. E., Backman, L. J., Kellgren, T. G., Tuft, S. J., Koskela, T., Rydén, P., Sandgren, O., Danielson, P. et al. (2015). Mutations in collagen, type XVII, alpha 1 (COL17A1) cause epithelial recurrent erosion dystrophy (ERED). *Hum. Mutat.* **36**, 463-473.
- Jurgens, J., Sobreira, N., Modaff, P., Reiser, C. A., Seo, S. H., Seong, M.-W., Park, S. S., Kim, O. H., Cho, T.-J. and Pauli, R. M. (2015). Novel COL2A1 variant (c.619G>A, p.Gly207Arg) manifesting as a phenotype similar to progressive pseudorheumatoid dysplasia and spondyloepiphyseal dysplasia, Stanescu type. *Hum. Mutat.* **36**, 1004-1008.
- Kadler, K. E., Baldock, C., Bella, J. and Boot-Handford, R. P. (2007). Collagens at a glance. *J. Cell Sci.* **120**, 1955-1958.
- Kannu, P., O'rielly, D. D., Hyland, J. C. and Kokko, L. A. (2011). Avascular necrosis of the femoral head due to a novel C propeptide mutation in COL2A1. *Am. J. Med. Genet. A* **155A**, 1759-1762.
- Kashtan, C. E. (1993). Alport syndrome and thin basement membrane nephropathy. In *GeneReviews((R))* (ed. M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens and A. Amemiya). Seattle, WA.
- Kashtan, C. E. (2004). Familial hematuria due to type IV collagen mutations: Alport syndrome and thin basement membrane nephropathy. *Curr. Opin. Pediatr.* **16**, 177-181.
- Kashtan, C. (2017). Alport syndrome: facts and opinions. *F1000Res* **6**, 50.
- Kawahara, G., Okada, M., Morone, N., Ibarra, C. A., Nonaka, I., Noguchi, S., Hayashi, Y. K. and Nishino, I. (2007). Reduced cell anchorage may cause sarcolemma-specific collagen VI deficiency in Ullrich disease. *Neurology* **69**, 1043-1049.
- Khoshnoodi, J., Cartailier, J.-P., Alvares, K., Veis, A. and Hudson, B. G. (2006). Molecular recognition in the assembly of collagens: terminal noncollagenous domains are key recognition modules in the formation of triple helical promoters. *J. Biol. Chem.* **281**, 38117-38121.
- Khoshnoodi, J., Pedchenko, V. and Hudson, B. G. (2008). Mammalian collagen IV. *Microsc. Res. Tech.* **71**, 357-370.
- King, N. and Rokas, A. (2017). Embracing uncertainty in reconstructing early animal evolution. *Curr. Biol.* **27**, R1081-R1088.
- King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I. et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783-788.
- Knebelmann, B., Deschenes, G., Gros, F., Hors, M. C., Grunfeld, J. P., Zhou, J., Tryggvason, K., Gubler, M. C. and Antignac, C. (1992). Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments. *Am. J. Hum. Genet.* **51**, 135-142.
- Körkkö, J., Ritvanen, P., Haataja, L., Kaariainen, H., Kivirikko, K. I., Prockop, D. J. and Ala-Kokko, L. (1993). Mutation in type II procollagen (COL2A1) that substitutes aspartate for glycine alpha 1-67 and that causes cataracts and retinal detachment: evidence for molecular heterogeneity in the Wagner syndrome and the Stickler syndrome (arthro-ophthalmopathy). *Am. J. Hum. Genet.* **53**, 55-61.
- Körkkö, J., Cohn, D. H., Ala-Kokko, L., Krakow, D. and Prockop, D. J. (2000). Widely distributed mutations in the COL2A1 gene produce achondrogenesis type II/hypochondrogenesis. *Am. J. Med. Genet.* **92**, 95-100.
- Kvist, A.-P., Latvanlehto, A., Sund, M., Eklund, L., Väisänen, T., Hägg, P., Sormunen, R., Komulainen, J., Fässler, R. and Pihlajaniemi, T. (2001). Lack of

- cytosolic and transmembrane domains of type XIII collagen results in progressive myopathy. *Am. J. Pathol.* **159**, 1581-1592.
- Labhard, M. E., Wirtz, M. K., Pope, F. M., Nicholls, A. C. and Hollister, D. W.** (1988). A cysteine for glycine substitution at position 1017 in an alpha 1(I) chain of type I collagen in a patient with mild dominantly inherited osteogenesis imperfecta. *Mol. Biol. Med.* **5**, 197-207.
- Lamande, S. R., Dahl, H. H., Cole, W. G. and Bateman, J. F.** (1989). Characterization of point mutations in the collagen COL1A1 and COL1A2 genes causing lethal perinatal osteogenesis imperfecta. *J. Biol. Chem.* **264**, 15809-15812.
- Lampe, A. K., Dunn, D. M., Von Niederhausern, A. C., Hamil, C., Aoyagi, A., Laval, S. H., Marie, S. K., Chu, M. L., Swoboda, K., Muntoni, F. et al.** (2005). Automated genomic sequence analysis of the three collagen VI genes: applications to Ullrich congenital muscular dystrophy and Bethlem myopathy. *J. Med. Genet.* **42**, 108-120.
- Lampert, D. T.** (1974). The role of hydroxyproline-rich proteins in the extracellular matrix of plants. *Symp. Soc. Dev. Biol.* **30**, 113-130.
- Lebleu, V., Sund, M., Sugimoto, H., Birrane, G., Kanasaki, K., Finan, E., Miller, C. A., Gattone, V. H., II, McLaughlin, H., Shield, C. F. III et al.** (2010). 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.* **285**, 41874-41885.
- Lee, B., Vissing, H., Ramirez, F., Rogers, D. and Rimoin, D.** (1989). Identification of the molecular defect in a family with spondyloepiphyseal dysplasia. *Science* **244**, 978-980.
- Lemmink, H. H., Nillesen, W. N., Mochizuki, T., Schröder, C. H., Brunner, H. G., Van Oost, B. A., Monnens, L. A. and Smeets, H. J.** (1996). Benign familial hematuria due to mutation of the type IV collagen alpha4 gene. *J. Clin. Invest.* **98**, 1114-1118.
- Leung, A. W. L., Wong, S. Y. Y., Chan, D., Tam, P. P. L. and Cheah, K. S. E.** (2010). Loss of procollagen IIA from the anterior mesendoderm disrupts the development of mouse embryonic forebrain. *Dev. Dyn.* **239**, 2319-2329.
- Leys, S. P., Nichols, S. A. and Adams, E. D. M.** (2009). Epithelia and integration in sponges. *Integr. Comp. Biol.* **49**, 167-177.
- Li, S. W., Prockop, D. J., Helminen, H., Fassler, R., Lapveteläinen, T., Kiraly, K., Peltari, A., Arokoski, J., Lui, H., Arita, M. et al.** (1995a). Transgenic mice with targeted inactivation of the Col2 alpha 1 gene for collagen II develop a skeleton with membranous and periosteal bone but no endochondral bone. *Genes Dev.* **9**, 2821-2830.
- Li, Y., Lacerda, D. A., Warman, M. L., Beier, D. R., Yoshioka, H., Ninomiya, Y., Oxford, J. T., Morris, N. P., Andrikopoulos, K., Ramirez, F. et al.** (1995b). A fibrillar collagen gene, Col11a1, is essential for skeletal morphogenesis. *Cell* **80**, 423-430.
- Liu, Y.-F., Chen, W.-M., Lin, Y.-F., Yang, R.-C., Lin, M.-W., Li, L.-H., Chang, Y.-H., Jou, Y.-S., Lin, P.-Y., Su, J.-S. et al.** (2005). Type II collagen gene variants and inherited osteonecrosis of the femoral head. *N. Engl. J. Med.* **352**, 2294-2301.
- Logan, C. V., Cossins, J., Rodriguez Cruz, P. M., Parry, D. A., Maxwell, S., Martínez-Martínez, P., Riepsaame, J., Abdelhamed, Z. A., Lake, A. V. R., Moran, M. et al.** (2015). Congenital myasthenic syndrome type 19 is caused by mutations in COL13A1, encoding the atypical non-fibrillar collagen type XIII alpha1 chain. *Am. J. Hum. Genet.* **97**, 878-885.
- Lucioli, S., Giusti, B., Mercuri, E., Vanegas, O. C., Lucarini, L., Pietroni, V., Urtizbera, A., Ben Yaou, R., De Visser, M., Van Der Kooi, A. J. et al.** (2005). Detection of common and private mutations in the COL6A1 gene of patients with Bethlem myopathy. *Neurology* **64**, 1931-1937.
- Lukashev, M. E. and Werb, Z.** (1998). ECM signalling: orchestrating cell behaviour and misbehaviour. *Trends Cell Biol.* **8**, 437-441.
- Malfait, F., Coucke, P., Symoens, S., Loey, B., Nuytinck, L. and De Paepe, A.** (2005). The molecular basis of classic Ehlers-Danlos syndrome: a comprehensive study of biochemical and molecular findings in 48 unrelated patients. *Hum. Mutat.* **25**, 28-37.
- Malfait, F., Symoens, S., Coucke, P., Nunes, L., De Almeida, S. and De Paepe, A.** (2006). Total absence of the alpha2(I) chain of collagen type I causes a rare form of Ehlers-Danlos syndrome with hypermobility and propensity to cardiac valvular problems. *J. Med. Genet.* **43**, e36.
- Marini, J. C., Forlino, A., Cabral, W. A., Barnes, A. M., San Antonio, J. D., Milgrom, S., Hyland, J. C., Körrkö, J., Prockop, D. J., De Paepe, A. et al.** (2007). Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum. Mutat.* **28**, 209-221.
- Matsui, Y., Michigami, T., Tachikawa, K., Yamazaki, M., Kawabata, H. and Nishimura, G.** (2009). Czech dysplasia occurring in a Japanese family. *Am. J. Med. Genet. A* **149A**, 2285-2289.
- Mcalinden, A., Majava, M., Bishop, P. N., Perveen, R., Black, G. C. M., Pierpont, M. E., Ala-Kokko, L. and Männikkö, M.** (2008). Missense and nonsense mutations in the alternatively-spliced exon 2 of COL2A1 cause the ocular variant of Stickler syndrome. *Hum. Mutat.* **29**, 83-90.
- Mccall, A. S., Cummings, C. F., Bhawe, G., Vanacore, R., Page-Mccaw, A. and Hudson, B. G.** (2014). Bromine is an essential trace element for assembly of collagen IV scaffolds in tissue development and architecture. *Cell* **157**, 1380-1392.
- Mcgrath, J. A., Gatalica, B., Li, K., Dunnill, M. G., McMillan, J. R., Christiano, A. M., Eady, R. A. and Uitto, J.** (1996). Compound heterozygosity for a dominant glycine substitution and a recessive internal duplication mutation in the type XVII collagen gene results in junctional epidermolysis bullosa and abnormal dentition. *Am. J. Pathol.* **148**, 1787-1796.
- Mcguire, W. T., Prasad, S. D., Griffith, A. J., Kunst, H. P. M., Green, G. E., Shpargel, K. B., Runge, C., Huybrechts, C., Mueller, R. F., Lynch, E. et al.** (1999). Mutations in COL11A2 cause non-syndromic hearing loss (DFNA13). *Nat. Genet.* **23**, 413-419.
- Mcintosh, I., Abbott, M. H., Warman, M. L., Olsen, B. R. and Francomano, C. A.** (1994). Additional mutations of type X collagen confirm COL10A1 as the Schmid metaphyseal chondrodysplasia locus. *Hum. Mol. Genet.* **3**, 303-307.
- Mehta, L. and Jim, B.** (2017). Hereditary renal diseases. *Semin. Nephrol.* **37**, 354-361.
- Melkoniemi, M., Brunner, H. G., Manouvrier, S., Hennekam, R., Superti-Furga, A., Kääriäinen, H., Pauli, R. M., Van Essen, T., Warman, M. L., Bonaventure, J. et al.** (2000). Autosomal recessive disorder otospondyloomegaepiphyseal dysplasia is associated with loss-of-function mutations in the COL11A2 gene. *Am. J. Hum. Genet.* **66**, 368-377.
- Meyer, F. and Moussian, B.** (2009). Drosophila multiplexin (Dmp) modulates motor axon pathfinding accuracy. *Dev. Growth Differ.* **51**, 483-498.
- Michalickova, K., Susic, M., Willing, M. C., Wenstrup, R. J. and Cole, W. G.** (1998). Mutations of the alpha2(V) chain of type V collagen impair matrix assembly and produce ehlers-danlos syndrome type I. *Hum. Mol. Genet.* **7**, 249-255.
- Miyamoto, Y., Matsuda, T., Kitoh, H., Haga, N., Ohashi, H., Nishimura, G. and Ikegawa, S.** (2007). A recurrent mutation in type II collagen gene causes Legg-Calvé-Perthes disease in a Japanese family. *Hum. Genet.* **121**, 625-629.
- Mochizuki, T., Lemmink, H. H., Mariyama, M., Signac, C., Gubler, M.-C., Pirson, Y., Verellen-Dumoulin, C., Chan, B., Schröder, C. H., Smeets, H. J. et al.** (1994). Identification of mutations in the alpha 3(IV) and alpha 4(IV) collagen genes in autosomal recessive Alport syndrome. *Nat. Genet.* **8**, 77-81.
- Moroz, L. L., Kocot, K. M., Citarella, M. R., Dosung, S., Norekian, T. P., Povolotskaya, I. S., Grigorenko, A. P., Dailey, C., Berezikov, E., Buckley, K. M. et al.** (2014). The tenebrionid genome and the evolutionary origins of neural systems. *Nature* **510**, 109-114.
- Mortier, G. R., Wilkin, D. J., Wilcox, W. R., Rimoin, D. L., Lachman, R. S., Eyre, D. R. and Cohn, D. H.** (1995). A radiographic, morphologic, biochemical and molecular analysis of a case of achondrogenesis type II resulting from substitution for a glycine residue (Gly<sup>691</sup>→Arg) in the type II collagen trimer. *Hum. Mol. Genet.* **4**, 285-288.
- Mortier, G. R., Weis, M., Nuytinck, L., King, L. M., Wilkin, D. J., De Paepe, A., Lachman, R. S., Rimoin, D. L., Eyre, D. R. and Cohn, D. H.** (2000). Report of five novel and one recurrent COL2A1 mutations with analysis of genotype-phenotype correlation in patients with a lethal type II collagen disorder. *J. Med. Genet.* **37**, 263-271.
- Nagel, M., Nagorka, S. and Gross, O.** (2005). Novel COL4A5, COL4A4, and COL4A3 mutations in Alport syndrome. *Hum. Mutat.* **26**, 60.
- Narcisi, P., Richards, A. J., Ferguson, S. D. and Pope, F. M.** (1994). A family with Ehlers-Danlos syndrome type III/articular hypermobility syndrome has a glycine 637 to serine substitution in type III collagen. *Hum. Mol. Genet.* **3**, 1617-1620.
- Nelson, C. M. and Bissell, M. J.** (2006). Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer. *Annu. Rev. Cell Dev. Biol.* **22**, 287-309.
- Nishie, W., Sawamura, D., Goto, M., Ito, K., Shibaki, A., McMillan, J. R., Sakai, K., Nakamura, H., Olasz, E., Yancey, K. B. et al.** (2007). Humanization of autoantigen. *Nat. Med.* **13**, 378-383.
- Nishimura, G., Nakashima, E., Mabuchi, A., Shimamoto, K., Shimamoto, T., Shima, Y., Nagai, T., Yamaguchi, T., Kosaki, R., Ohashi, H. et al.** (2004). Identification of COL2A1 mutations in platyspondylic skeletal dysplasia, Torrance type. *J. Med. Genet.* **41**, 75-79.
- Nuytinck, L., Freund, M., Lagae, L., Pierard, G. E., Hermanns-Le, T. and De Paepe, A.** (2000). Classical Ehlers-Danlos syndrome caused by a mutation in type I collagen. *Am. J. Hum. Genet.* **66**, 1398-1402.
- Okuyama, K., Okuyama, K., Arnott, S., Takayanagi, M. and Kakudo, M.** (1981). Crystal and molecular structure of a collagen-like polypeptide (Pro-Pro-Gly)<sub>10</sub>. *J. Mol. Biol.* **152**, 427-443.
- Orgel, J. P. R. O., San Antonio, J. D. and Antipova, O.** (2011). Molecular and structural mapping of collagen fibril interactions. *Connect. Tissue Res.* **52**, 2-17.
- Ozbek, S., Balasubramanian, P. G., Chiquet-Ehrismann, R., Tucker, R. P. and Adams, J. C.** (2010). The evolution of extracellular matrix. *Mol. Biol. Cell* **21**, 4300-4305.
- Paasilta, P., Lohiniva, J., Annunen, S., Bonaventure, J., Le Merrer, M., Pai, L. and Ala-Kokko, L.** (1999). COL9A3: a third locus for multiple epiphyseal dysplasia. *Am. J. Hum. Genet.* **64**, 1036-1044.
- Paasilta, P., Lohiniva, J., Goring, H. H., Perala, M., Raina, S. S., Karppinen, J., Hakala, M., Palm, T., Kroger, H., Kaitila, I. et al.** (2001). Identification of a novel common genetic risk factor for lumbar disk disease. *JAMA* **285**, 1843-1849.
- Pan, T. C., Zhang, R. Z., Pericak-Vance, M. A., Tandan, R., Fries, T., Stajich, J. M., Viles, K., Vance, J. M., Chu, M. L. and Speer, M. C.** (1998). Missense



- mutation in a von Willebrand factor type A domain of the alpha 3(VI) collagen gene (COL6A3) in a family with Bethlem myopathy. *Hum. Mol. Genet.* **7**, 807-812.
- Park, A. C., Phillips, C. L., Pfeiffer, F. M., Roenneburg, D. A., Kernien, J. F., Adams, S. M., Davidson, J. M., Birk, D. E. and Greenspan, D. S.** (2015). Homozygosity and heterozygosity for null Col5a2 alleles produce embryonic lethality and a novel classic Ehlers-Danlos syndrome-related phenotype. *Am. J. Pathol.* **185**, 2000-2011.
- Parkin, J. D., San Antonio, J. D., Pedchenko, V., Hudson, B., Jensen, S. T. and Savage, J.** (2011). Mapping structural landmarks, ligand binding sites, and missense mutations to the collagen IV heterotrimers predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes. *Hum. Mutat.* **32**, 127-143.
- Pedchenko, V., Bondar, O., Fogo, A. B., Vanacore, R., Voziyan, P., Kitching, A. R., Wieslander, J., Kashtan, C., Borza, D.-B., Neilson, E. G. et al.** (2010). Molecular architecture of the Goodpasture autoantigen in anti-GBM nephritis. *N. Engl. J. Med.* **363**, 343-354.
- Peltonen, L., Palotie, A. and Prockop, D. J.** (1980). A defect in the structure of type I procollagen in a patient who had osteogenesis imperfecta: excess mannose in the COOH-terminal propeptide. *Proc. Natl. Acad. Sci. USA* **77**, 6179-6183.
- Penades, J. R., Bernal, D., Revert, F., Johansson, C., Fresquet, V. J., Cervera, J., Wieslander, J., Quinones, S. and Saus, J.** (1995). Characterization and expression of multiple alternatively spliced transcripts of the Goodpasture antigen gene region. Goodpasture antibodies recognize recombinant proteins representing the autoantigen and one of its alternative forms. *Eur. J. Biochem.* **229**, 754-760.
- Pepe, G., Bertini, E., Giusti, B., Brunelli, T., Comeglio, P., Saitta, B., Merlini, L., Chu, M. L., Federici, G. and Abbate, R.** (1999). A novel de novo mutation in the triple helix of the COL6A3 gene in a two-generation Italian family affected by Bethlem myopathy. A diagnostic approach in the mutations' screening of type VI collagen. *Neuromuscul. Disord.* **9**, 264-271.
- Pepin, M. G., Murray, M. L. and Byers, P. H.** (1993). Vascular Ehlers-Danlos syndrome. In *GeneReviews*(R) (ed. M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens and A. Amemiya). Seattle, WA.
- Petersen, O. W., Ronnov-Jessen, L., Howlett, A. R. and Bissell, M. J.** (1992). Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells. *Proc. Natl. Acad. Sci. USA* **89**, 9064-9068.
- Pihlajamaa, T., Prockop, D. J., Faber, J., Winterpacht, A., Zabel, B., Giedion, A., Wiesbauer, P., Spranger, J. and Ala-Kokko, L.** (1998). Heterozygous glycine substitution in the COL11A2 gene in the original patient with the Weissenbacher-Zweymüller syndrome demonstrates its identity with heterozygous OSMED (nonocular Stickler syndrome). *Am. J. Med. Genet.* **80**, 115-120.
- Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., Lartillot, N. and Worheide, G.** (2015). Genomic data do not support comb jellies as the sister group to all other animals. *Proc. Natl. Acad. Sci. USA* **112**, 15402-15407.
- Plaisier, E., Gribouval, O., Alamowitch, S., Mougnot, B., Prost, C., Verpont, M. C., Marro, B., Desmettre, T., Cohen, S. Y., Rouillet, E. et al.** (2007). COL4A1 mutations and hereditary angiopathy, nephropathy, aneurysms, and muscle cramps. *N. Engl. J. Med.* **357**, 2687-2695.
- Plaisier, E., Chen, Z., Gekeler, F., Benhassine, S., Dahan, K., Marro, B., Alamowitch, S., Paques, M. and Ronco, P.** (2010). Novel COL4A1 mutations associated with HANAC syndrome: a role for the triple helical CB3[IV] domain. *Am. J. Med. Genet. A* **152A**, 2550-2555.
- Plumb, D. A., Ferrara, L., Torbica, T., Knowles, L., Mironov, A., Jr., Kadler, K. E., Briggs, M. D. and Boot-Handford, R. P.** (2011). Collagen XXVII organizes the pericellular matrix in the growth plate. *PLoS ONE* **6**, e29422.
- Pokidysheva, E., Zientek, K. D., Ishikawa, Y., Mizuno, K., Vranka, J. A., Montgomery, N. T., Keene, D. R., Kawaguchi, T., Okuyama, K. and Bachinger, H. P.** (2013). Posttranslational modifications in type I collagen from different tissues extracted from wild type and prolyl 3-hydroxylase 1 null mice. *J. Biol. Chem.* **288**, 24742-24752.
- Poschl, E., Schlotzer-Schrehardt, U., Brachvogel, B., Saito, K., Ninomiya, Y. and Mayer, U.** (2004). Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development. *Development* **131**, 1619-1628.
- Ramachandran, G. N. and Kartha, G.** (1954). Structure of collagen. *Nature* **174**, 269-270.
- Raya, A., Revert, F., Navarro, S. and Saus, J.** (1999). Characterization of a novel type of serine/threonine kinase that specifically phosphorylates the human goodpasture antigen. *J. Biol. Chem.* **274**, 12642-12649.
- Renieri, A., Seri, M., Myers, J. C., Pihlajaniemi, T., Massella, L., Rizzoni, G. and De Marchi, M.** (1992). De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome. *Hum. Mol. Genet.* **1**, 127-129.
- Revert, F., Penades, J. R., Plana, M., Bernal, D., Johansson, C., Itarte, E., Cervera, J., Wieslander, J., Quinones, S. and Saus, J.** (1995). Phosphorylation of the Goodpasture antigen by type A protein kinases. *J. Biol. Chem.* **270**, 13254-13261.
- Rheault, M. N., Kren, S. M., Thielen, B. K., Mesa, H. A., Crosson, J. T., Thomas, W., Sado, Y., Kashtan, C. E. and Segal, Y.** (2004). Mouse model of X-linked Alport syndrome. *J. Am. Soc. Nephrol.* **15**, 1466-1474.
- Ricard-Blum, S.** (2011). The collagen family. *Cold Spring Harb. Perspect. Biol.* **3**, a004978.
- Ricard-Blum, S. and Ruggiero, F.** (2005). The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathol. Biol. (Paris)* **53**, 430-442.
- Rich, A. and Crick, F. H.** (1955). The structure of collagen. *Nature* **176**, 915-916.
- Richards, A. J., Yates, J. R., Williams, R., Payne, S. J., Pope, F. M., Scott, J. D. and Snead, M. P.** (1996). A family with Stickler syndrome type 2 has a mutation in the COL11A1 gene resulting in the substitution of glycine 97 by valine in alpha 1 (X1) collagen. *Hum. Mol. Genet.* **5**, 1339-1343.
- Richards, A. J., Martin, S., Nicholls, A. C., Harrison, J. B., Pope, F. M. and Burrows, N. P.** (1998). A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type II. *J. Med. Genet.* **35**, 846-848.
- Richards, A. J., Baguley, D. M., Yates, J. R., Lane, C., Nicol, M., Harper, P. S., Scott, J. D. and Snead, M. P.** (2000). Variation in the vitreous phenotype of Stickler syndrome can be caused by different amino acid substitutions in the X position of the type II collagen Gly-X-Y triple helix. *Am. J. Hum. Genet.* **67**, 1083-1094.
- Richards, A. J., Mcninch, A., Martin, H., Oakhill, K., Rai, H., Waller, S., Treacy, B., Whittaker, J., Meredith, S., Poulson, A. et al.** (2010). Stickler syndrome and the vitreous phenotype: mutations in COL2A1 and COL11A1. *Hum. Mutat.* **31**, E1461-E1471.
- Richter, D. J. and King, N.** (2013). The genomic and cellular foundations of animal origins. *Annu. Rev. Genet.* **47**, 509-537.
- Rosati, R., Horan, G. S. B., Pinero, G. J., Garofalo, S., Keene, D. R., Horton, W. A., Vuorio, E., De Crombrughe, B. and Behringer, R. R.** (1994). Normal long bone growth and development in type X collagen-null mice. *Nat. Genet.* **8**, 129-135.
- Rost, S., Bach, E., Neuner, C., Nanda, I., Dysek, S., Bittner, R. E., Keller, A., Bartsch, O., Mlynski, R., Haaf, T. et al.** (2014). Novel form of X-linked nonsyndromic hearing loss with cochlear malformation caused by a mutation in the type IV collagen gene COL4A6. *Eur. J. Hum. Genet.* **22**, 208-215.
- Ryan, J. F., Pang, K., Schnitzler, C. E., Nguyen, A.-D., Moreland, R. T., Simmons, D. K., Koch, B. J., Francis, W. R., Havlak, P., Program, N. C. S. et al.** (2013). The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution. *Science* **342**, 1242592.
- Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y. and Prockop, D. J.** (1973). Synthesis of (Pro-Hyp-Gly) n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. *Biochim. Biophys. Acta* **303**, 198-202.
- Scacheri, P. C., Gillanders, E. M., Subramony, S. H., Vedanarayanan, V., Crowe, C. A., Thakore, N., Bingler, M. and Hoffman, E. P.** (2002). Novel mutations in collagen VI genes: expansion of the Bethlem myopathy phenotype. *Neurology* **58**, 593-602.
- Schierwater, B., De Jong, D. and Desalle, R.** (2009). Placozoa and the evolution of Metazoa and intrasomatic cell differentiation. *Int. J. Biochem. Cell Biol.* **41**, 370-379.
- Schnieke, A., Harbers, K. and Jaenisch, R.** (1983). Embryonic lethal mutation in mice induced by retrovirus insertion into the alpha 1(I) collagen gene. *Nature* **304**, 315-320.
- Schumann, H., Hammami-Hauasli, N., Pulkkinen, L., Mauviel, A., Küster, W., Lüthi, U., Owaribe, K., Uitto, J. and Bruckner-Tuderman, L.** (1997). Three novel homozygous point mutations and a new polymorphism in the COL17A1 gene: relation to biological and clinical phenotypes of junctional epidermolysis bullosa. *Am. J. Hum. Genet.* **60**, 1344-1353.
- Sertie, A. L., Sossi, V., Camargo, A. A., Zatz, M., Brahe, C. and Passos-Bueno, M. R.** (2000). Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). *Hum. Mol. Genet.* **9**, 2051-2058.
- Shah, S., Ellard, S., Kneen, R., Lim, M., Osborne, N., Rankin, J., Stoodley, N., Van Der Knaap, M., Whitney, A. and Jardine, P.** (2012). Childhood presentation of COL4A1 mutations. *Dev. Med. Child Neurol.* **54**, 569-574.
- Sherwood, D. R.** (2015). A developmental biologist's "outside-the-cell" thinking. *J. Cell Biol.* **210**, 369-372.
- Shinkuma, S.** (2015). Dystrophic epidermolysis bullosa: a review. *Clin. Cosmet. Investig. Dermatol.* **8**, 275-284.
- Shinwari, J. M. A., Khan, A., Awad, S., Shinwari, Z., Alaiya, A., Alanazi, M., Tahir, A., Poizat, C. and Al Tassan, N.** (2015). Recessive mutations in COL25A1 are a cause of congenital cranial dysinnervation disorder. *Am. J. Hum. Genet.* **96**, 147-152.
- Shoulders, M. D. and Raines, R. T.** (2009). Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929-958.
- Sibon, I., Coupy, I., Menegon, P., Bouchet, J.-P., Gorry, P., Burgelin, I., Calvas, P., Orignac, I., Dousset, V., Lacombe, D. et al.** (2007). COL4A1 mutation in Axenfeld-Rieger anomaly with leukoencephalopathy and stroke. *Ann. Neurol.* **62**, 177-184.
- Sirko-Osadsa, D. A., Murray, M. A., Scott, J. A., Lavery, M. A., Warman, M. L. and Robin, N. H.** (1998). Stickler syndrome without eye involvement is caused by

- mutations in COL11A2, the gene encoding the alpha2(XI) chain of type XI collagen. *J. Pediatr.* **132**, 368-371.
- Smith, L. B., Hadoke, P. W. F., Dyer, E., Denvir, M. A., Brownstein, D., Miller, E., Nelson, N., Wells, S., Cheeseman, M. and Greenfield, A. (2011). Haploinsufficiency of the murine Col3a1 locus causes aortic dissection: a novel model of the vascular type of Ehlers-Danlos syndrome. *Cardiovasc. Res.* **90**, 182-190.
- Steven, F. S. (1965). The cleavage of tyrosyl peptides by pepsin from collagen solubilised by the Nishihara technique. *Biochim. Biophys. Acta* **97**, 465-471.
- Sulko, J., Czarny-Ratajczak, M., Wozniak, A., Latos-Bielenska, A. and Kozlowski, K. (2005). Novel amino acid substitution in the Y-position of collagen type II causes spondyloepimetaphyseal dysplasia congenita. *Am. J. Med. Genet. A* **137A**, 292-297.
- Sumiyoshi, H., Mor, N., Lee, S. Y., Doty, S., Henderson, S., Tanaka, S., Yoshioka, H., Rattan, S. and Ramirez, F. (2004). Esophageal muscle physiology and morphogenesis require assembly of a collagen XIX-rich basement membrane zone. *J. Cell Biol.* **166**, 591-600.
- Symoens, S., Malfait, F., Renard, M., André, J., Hausser, I., Loeys, B., Coucke, P. and De Paepe, A. (2009). COL5A1 signal peptide mutations interfere with protein secretion and cause classic Ehlers-Danlos syndrome. *Hum. Mutat.* **30**, E395-E403.
- Tanaka, T., Wakabayashi, T., Ozumi, H., Nishio, S., Sato, T., Harada, A., Fujii, D., Matsuo, Y., Hashimoto, T. and Iwatsubo, T. (2014). CLAC-P/collagen type XXV is required for the intramuscular innervation of motoneurons during neuromuscular development. *J. Neurosci.* **34**, 1370-1379.
- Tasanen, K., Eble, J. A., Aumailley, M., Schumann, H., Baetge, J., Tu, H., Bruckner, P. and Bruckner-Tuderman, L. (2000a). Collagen XVII is destabilized by a glycine substitution mutation in the cell adhesion domain Col15. *J. Biol. Chem.* **275**, 3093-3099.
- Tasanen, K., Floeth, M., Schumann, H. and Bruckner-Tuderman, L. (2000b). Hemizygoty for a glycine substitution in collagen XVII: unfolding and degradation of the ectodomain. *J. Invest. Dermatol.* **115**, 207-212.
- Telford, M. J., Moroz, L. L. and Halanych, K. M. (2016). Evolution: a sisterly dispute. *Nature* **529**, 286-287.
- Tiller, G. E., Rimoin, D. L., Murray, L. W. and Cohn, D. H. (1990). Tandem duplication within a type II collagen gene (COL2A1) exon in an individual with spondyloepiphyseal dysplasia. *Proc. Natl. Acad. Sci. USA* **87**, 3889-3893.
- Tiller, G. E., Polumbo, P. A., Weis, M. A., Bogaert, R., Lachman, R. S., Cohn, D. H., Rimoin, D. L. and Eyre, D. R. (1995). Dominant mutations in the type II collagen gene, COL2A1, produce spondyloepimetaphyseal dysplasia, Strudwick type. *Nat. Genet.* **11**, 87-89.
- Ting, K., Ramachandran, H., Chung, K. S., Shah-Hosseini, N., Olsen, B. R. and Nishimura, I. (1999). A short isoform of Col9a1 supports alveolar bone repair. *Am. J. Pathol.* **155**, 1993-1999.
- Tompson, S. W., Bacino, C. A., Safina, N. P., Bober, M. B., Proud, V. K., Funari, T., Wangler, M. F., Nevarez, L., Ala-Kokko, L., Wilcox, W. R. et al. (2010). Fibrochondrogenesis results from mutations in the COL11A1 type XI collagen gene. *Am. J. Hum. Genet.* **87**, 708-712.
- Tompson, S. W., Faqeih, E. A., Ala-Kokko, L., Hecht, J. T., Miki, R., Funari, T., Funari, V. A., Nevarez, L., Krakow, D. and Cohn, D. H. (2012). Dominant and recessive forms of fibrochondrogenesis resulting from mutations at a second locus, COL11A2. *Am. J. Med. Genet. A* **158A**, 309-314.
- Tromp, G., Kuivaniemi, H., Shikata, H. and Prockop, D. J. (1989a). A single base mutation that substitutes serine for glycine 790 of the alpha 1 (III) chain of type III procollagen exposes an arginine and causes Ehlers-Danlos syndrome IV. *J. Biol. Chem.* **264**, 1349-1352.
- Tromp, G., Kuivaniemi, H., Stolle, C., Pope, F. M. and Prockop, D. J. (1989b). Single base mutation in the type III procollagen gene that converts the codon for glycine 883 to aspartate in a mild variant of Ehlers-Danlos syndrome IV. *J. Biol. Chem.* **264**, 19313-19317.
- Tzschach, A., Tinschert, S., Kaminsky, E., Lusga, E., Mundlos, S. and Graul-Neumann, L. M. (2008). Czech dysplasia: report of a large family and further delineation of the phenotype. *Am. J. Med. Genet. A* **146A**, 1859-1864.
- Utriainen, A., Sormunen, R., Kettunen, M., Carvalhaes, L. S., Sajanti, E., Eklund, L., Kauppinen, R., Kitten, G. T. and Pihlajaniemi, T. (2004). Structurally altered basement membranes and hydrocephalus in a type XVIII collagen deficient mouse line. *Hum. Mol. Genet.* **13**, 2089-2099.
- Uzawa, K., Marshall, M. K., Katz, E. P., Tanzawa, H., Yeowell, H. N. and Yamauchi, M. (1998). Altered posttranslational modifications of collagen in keloid. *Biochem. Biophys. Res. Commun.* **249**, 652-655.
- Valiathan, R. R., Marco, M., Leitinger, B., Kleer, C. G. and Fridman, R. (2012). Discoidin domain receptor tyrosine kinases: new players in cancer progression. *Cancer Metastasis Rev.* **31**, 295-321.
- Van Camp, G., Snoeckx, R. L., Hilgert, N., Van Den Ende, J., Fukuoka, H., Wagatsuma, M., Suzuki, H., Smets, R. M., Vanhoenacker, F., Declau, F. et al. (2006). A new autosomal recessive form of Stickler syndrome is caused by a mutation in the COL9A1 gene. *Am. J. Hum. Genet.* **79**, 449-457.
- Van Der Loop, F. T. L., Heidet, L., Timmer, E. D. J., Van Den Bosch, B. J. C., Leinonen, A., Antignac, C., Jefferson, J. A., Maxwell, A. P., Monnens, L. A. H., Schröder, C. H. et al. (2000). Autosomal dominant Alport syndrome caused by a COL4A3 splice site mutation. *Kidney Int.* **58**, 1870-1875.
- Vanacore, R., Ham, A.-J. L., Voehler, M., Sanders, C. R., Conrads, T. P., Veenstra, T. D., Sharpless, K. B., Dawson, P. E. and Hudson, B. G. (2009). A sulfilimine bond identified in collagen IV. *Science* **325**, 1230-1234.
- Vikkula, M., Marimán, E. C. M., Lui, V. C. H., Zhidkova, N. I., Tiller, G. E., Goldring, M. B., Van Beersum, S. E. C., De Waal Malefijt, M. C., Van Den Hoogen, F. H., Ropers, H. H. et al. (1995). Autosomal dominant and recessive osteochondrodysplasias associated with the COL11A2 locus. *Cell* **80**, 431-437.
- Vissing, H., D'alesio, M., Lee, B., Ramirez, F., Godfrey, M. and Hollister, D. W. (1989). Glycine to serine substitution in the triple helical domain of pro-alpha 1 (II) collagen results in a lethal perinatal form of short-limbed dwarfism. *J. Biol. Chem.* **264**, 18265-18267.
- Vogel, B. E., Minor, R. R., Freund, M. and Prockop, D. J. (1987). A point mutation in a type I procollagen gene converts glycine 748 of the alpha 1 chain to cysteine and destabilizes the triple helix in a lethal variant of osteogenesis imperfecta. *J. Biol. Chem.* **262**, 14737-14744.
- Wallis, G. A., Rash, B., Sweetman, W. A., Thomas, J. T., Super, M., Evans, G., Grant, M. E. and Boot-Handford, R. P. (1994). Amino acid substitutions of conserved residues in the carboxyl-terminal domain of the alpha 1(X) chain of type X collagen occur in two unrelated families with metaphyseal chondrodysplasia type Schmid. *Am. J. Hum. Genet.* **54**, 169-178.
- Wang, X., Harris, R. E., Bayston, L. J. and Ashe, H. L. (2008). Type IV collagens regulate BMP signalling in *Drosophila*. *Nature* **455**, 72-77.
- Watanabe, M., Nakagawa, R., Naruto, T., Kohmoto, T., Suga, K., Goji, A., Kagami, S., Masuda, K. and Imoto, I. (2016). A novel missense mutation of COL5A2 in a patient with Ehlers-Danlos syndrome. *Hum. Genome Var.* **3**, 16030.
- Weng, Y. C., Sonni, A., Labelle-Dumais, C., De Leau, M., Kauffman, W. B., Jeanne, M., Biffi, A., Greenberg, S. M., Rosand, J. and Gould, D. B. (2012). COL4A1 mutations in patients with sporadic late-onset intracerebral hemorrhage. *Ann. Neurol.* **71**, 470-477.
- Wenstrup, R. J., Cohn, D. H., Cohen, T. and Byers, P. H. (1988). Arginine for glycine substitution in the triple-helical domain of the products of one alpha 2(I) collagen allele (COL1A2) produces the osteogenesis imperfecta type IV phenotype. *J. Biol. Chem.* **263**, 7734-7740.
- Wenstrup, R. J., Florer, J. B., Brunskill, E. W., Bell, S. M., Chervoneva, I. and Birk, D. E. (2004). Type V collagen controls the initiation of collagen fibril assembly. *J. Biol. Chem.* **279**, 53331-53337.
- Whelan, N. V., Kocot, K. M., Moroz, L. L. and Halanych, K. M. (2015). Error, signal, and the placement of *Ctenophora* sister to all other animals. *Proc. Natl. Acad. Sci. USA* **112**, 5773-5778.
- Whittaker, C. A., Bergeron, K.-F., Whittle, J., Brandhorst, B. P., Burke, R. D. and Hynes, R. O. (2006). The echinoderm adhesome. *Dev. Biol.* **300**, 252-266.
- Wilkin, D. J., Bogaert, R., Lachman, R. S., Rimoin, D. L., Eyre, D. R. and Cohn, D. H. (1994). A single amino acid substitution (G103D) in the type II collagen triple helix produces Kniest dysplasia. *Hum. Mol. Genet.* **3**, 1999-2003.
- Williams, C. J., Cuxidine, E. L., Knowlton, R. G., Reginato, A., Neumann, G., Harrison, D., Buxton, P., Jimenez, S. and Prockop, D. J. (1993). Spondyloepiphyseal dysplasia and precocious osteoarthritis in a family with an Arg75→Cys mutation in the procollagen type II gene (COL2A1). *Hum. Genet.* **92**, 499-505.
- Williamson, D. A. J. (1961). Alport's syndrome of hereditary nephritis with deafness. *Lancet* **2**, 1321-1323.
- Winterpacht, A., Superti-Furga, A., Schwarze, U., Stoss, H., Steinmann, B., Spranger, J. and Zabel, B. (1996). The deletion of six amino acids at the C-terminus of the alpha 1 (II) chain causes overmodification of type II and type XI collagen: further evidence for the association between small deletions in COL2A1 and Kniest dysplasia. *J. Med. Genet.* **33**, 649-654.
- Wirtz, M. K., Glanville, R. W., Steinmann, B., Rao, V. H. and Hollister, D. W. (1987). Ehlers-Danlos syndrome type VIIB. Deletion of 18 amino acids comprising the N-telopeptide region of a pro-alpha 2(I) chain. *J. Biol. Chem.* **262**, 16376-16385.
- Woodley, D. T., Keene, D. R., Atha, T., Huang, Y., Lipman, K., Li, W. and Chen, M. (2004). Injection of recombinant human type VII collagen restores collagen function in dystrophic epidermolysis bullosa. *Nat. Med.* **10**, 693-695.
- Wu, Y., Li, G. and Zhu, X. (2002). A novel homozygous point mutation in the COL17A1 gene in a Chinese family with generalized atrophic benign epidermolysis bullosa. *J. Dermatol. Sci.* **28**, 181-186.
- Yalak, G. and Olsen, B. R. (2015). Proteomic database mining opens up avenues utilizing extracellular protein phosphorylation for novel therapeutic applications. *J. Transl. Med.* **13**, 125.
- Yamada, Y., Avvedimento, V. E., Mudryj, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. and De Crombrughe, B. (1980). The collagen gene: evidence for its evolutionary assembly by amplification of a DNA segment containing an exon of 54 bp. *Cell* **22**, 887-892.
- Yamauchi, M. and Shiiba, M. (2008). Lysine hydroxylation and cross-linking of collagen. *Methods Mol. Biol.* **446**, 95-108.
- Yoneda, Y., Haginoya, K., Arai, H., Yamaoka, S., Tsurusaki, Y., Doi, H., Miyake, N., Yokochi, K., Osaka, H., Kato, M. et al. (2012). De novo and inherited

- mutations in COL4A2, encoding the type IV collagen alpha2 chain cause porencephaly. *Am. J. Hum. Genet.* **90**, 86-90.
- Yoneda, Y., Haginoya, K., Kato, M., Osaka, H., Yokochi, K., Arai, H., Kakita, A., Yamamoto, T., Otsuki, Y., Shimizu, S. et al. (2013). Phenotypic spectrum of COL4A1 mutations: porencephaly to schizencephaly. *Ann. Neurol.* **73**, 48-57.
- Yurchenco, P. D. (2011). Basement membranes: cell scaffoldings and signaling platforms. *Cold Spring Harb. Perspect. Biol.* **3**, a004911.
- Zajac, A. L. and Horne-Badovinac, S. (2017). Tissue structure: a CIVICs lesson for adipocytes. *Curr. Biol.* **27**, R1013-R1015.
- Zankl, A., Neumann, L., Ignatius, J., Nikkels, P., Schrandner-Stumpel, C., Mortier, G., Omran, H., Wright, M., Hilbert, K., Bonafe, L. et al. (2005). Dominant negative mutations in the C-propeptide of COL2A1 cause platyspondylic lethal skeletal dysplasia, torrance type, and define a novel subfamily within the type 2 collagenopathies. *Am. J. Med. Genet. A* **133A**, 61-67.
- Zech, M., Lam, D. D., Francescatto, L., Schormair, B., Salminen, A. V., Jochim, A., Wieland, T., Lichtner, P., Peters, A., Gieger, C. et al. (2015). Recessive mutations in the alpha3 (VI) collagen gene COL6A3 cause early-onset isolated dystonia. *Am. J. Hum. Genet.* **96**, 883-893.
- Zenteno, J. C., Crespi, J., Buentello-Volante, B., Buil, J. A., Bassaganyas, F., Vela-Segarra, J. I., Diaz-Casajosa, J. and Marieges, M. T. (2014). Next generation sequencing uncovers a missense mutation in COL4A1 as the cause of familial retinal arteriolar tortuosity. *Graefes Arch. Clin. Exp. Ophthalmol.* **252**, 1789-1794.
- Zhou, J., Barker, D. F., Hostikka, S. L., Gregory, M. C., Atkin, C. L. and Tryggvason, K. (1991). Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome. *Genomics* **9**, 10-18.
- Zhou, J., Hertz, J. M., Leinonen, A. and Tryggvason, K. (1992). Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient. *J. Biol. Chem.* **267**, 12475-12481.
- Zou, Y., Zwolanek, D., Izu, Y., Gandhy, S., Schreiber, G., Brockmann, K., Devoto, M., Tian, Z., Hu, Y., Veit, G. et al. (2014). Recessive and dominant mutations in COL12A1 cause a novel EDS/myopathy overlap syndrome in humans and mice. *Hum. Mol. Genet.* **23**, 2339-2352.