REGULATION OF COLLAGEN TRAFFICKING PATHWAYS IN HOMEOSTASIS AND DISEASE

By

Gokhan Unlu

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
in
Cell and Developmental Biology
May 10, 2019
Nashville, Tennessee

Approved:

Kathleen L. Gould, Ph.D.
Chin Chiang, Ph.D.
Anne K. Kenworthy, Ph.D.
Florent Elefteriou, Ph.D.
Ela W. Knapik, M.D.

DEDICATION

For my family

ACKNOWLEDGEMENTS

I would firstly like to present my deepest gratitude to my advisor Ela Knapik, without whom this thesis study would neither be started nor completed. I owe her a debt of thanks for her endless moral support, guidance, encouragement, tolerance, patience and understanding. I cannot feel more fortunate to have such a mentor who has encouraged me to expand my horizon, get out of my comfort zone, try out new experiments and test new ideas. I am thankful to her for being such a friendly and supportive mentor, creating a productive lab environment and investing such a great deal in my academic growth.

I also want to thank my thesis committee members, Kathy Gould, Chin Chiang, Anne Kenworthy and Florent Elefteriou. I am grateful to them for sharing their expertise to better my research, helping me improve my scientific communication skills and their endless support throughout this thesis work. I feel very fortunate to have committee members who invested in me greatly by both providing constructive feedback on my thesis project and helping me shape my academic career.

I am thankful to Vanderbilt International Scholar Program (VISP), which recruited me to Vanderbilt and provided me financial and social support. I thank VISP director Kathy Gould, founding steering committee members, Anne Kenworthy, Roger Colbran, Mark Denison, Karoly Mirnics and VISP manager Amanda Connolly, as well as other VISP members, for putting together such a unique program that is tailor-made for the needs of international PhD students. I cannot emphasize enough how important the support I received from VISP program throughout my graduate school experience, from beginning to the end.

I want to thank collaborators, including Antonis Hatzopolous, Nancy J. Cox, Eric R. Gamazon, Fowzan S. Alkuraya, Rui Chen, Bingshan Li, Mike Boyce and Clare Baker for the opportunity of participating in their exciting projects and their great contributions to this thesis project. I have learned a great deal from these collaborations. I especially would like to extend my gratitude to Nancy J. Cox, who has been a great advocate for graduate students, for providing me continuous support and mentorship throughout our collaborative projects.

I am thankful to my fellow lab members, Lauryn Luderman, Dharmendra Choudhary, Dylan Ritter, Amy Rushing, Cory Guthrie, Amanda Goodrich, Sabeen Rehman; as well as past members of the Knapik Lab, Daniel Levic, David Melville, Kinsey Qi and Joseph Breeyear. I appreciate them for their help and making Knapik Lab a stimulating, productive and fun environment I have always enjoyed working in.

I like to thank my Nashville friends and VISP community for making my Vanderbilt experience a unique and enjoyable one. Especially, I am grateful to have such great and supportive friends as Mehmet Takar and Nergis Kara for always being there for me.

Last but not the least, I would like to thank my parents, Nuran & Fikret Ünlü; and my sisters Eda Ersöz and Funda Mercimek for their endless love and support throughout graduate school, and being my very first teachers.

TABLE OF CONTENTS

Page

Chapters

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

Overview

Extracellular matrix (ECM) facilitates organ morphogenesis, growth and homeostasis by providing structural support to tissues and by harboring soluble ligands that regulate growth factor signaling (Hynes, 2009). Development and maintenance of skeletal tissues like bone and cartilage rely heavily on ECM deposition. Mutations affecting synthesis, post-translational modification and secretion of ECM components lead to debilitating skeletal disorders such as osteogenesis imperfecta (OI, also known as brittle bone disease), scoliosis and osteoporosis (Cabral et al., 2005; Dalgleish, 1997; Francomano et al., 1987; Luderman et al., 2017; Montanaro et al., 2006; Willing et al., 1994).

Composition of ECM is specific to each tissue to fulfill their distinct mechano-pyhsical demands. While bone matrix is mineralized and rich in collagen-I, cartilage ECM is composed mainly of collagen-II fibers and proteoglycans (Gentili and Cancedda, 2009). Major ECM components, collagens constitute over 30% of dry body weight across vertebrates (Ishikawa and Bachinger, 2013). Twenty-eight different types of vertebrate collagens have been identified, each having distinct expression patterns and biochemical properties (Shoulders and Raines, 2009). A characteristic structural feature of collagens is a triple helix structure, in which every third amino acid of each polypeptide is glycine (Gly). The first and the second amino acids could vary; however, they are often proline and 4-hydroxyproline, respectively. Collagen triple helices (this trimeric structure is also known as tropocollagen) assemble into higher order oligomers, consequently forming fibers (i.e. fibrillar collagen types), such as those in cartilage and bone, or networks (i.e. network-forming collagen types), as in the case of basement membrane.

Despite their abundance and critical physiological functions, fibrillar collagens, such as collagen type-I and –II, present a challenge for the secretory pathway since they already form triple helix structures in the endoplasmic reticulum (ER) and span ~300 nm in width, which is larger than average size secretory cargo (Ishikawa and Bachinger, 2013). In the last ten years, it has been reported that tropocollagen molecules require specialized machinery to accommodate their unusual size and oligomeric structure (Gorur et al., 2017; Malhotra et al., 2015; Unlu et al., 2014). Recent biochemical and genetic studies have begun to reveal components of trafficking machinery regulating ER-exit and packaging of larger-than-average size collagen cargos into COPII (coat protein II complex) carriers (Lang et al., 2006; Saito et al., 2009b; Saito et al., 2011; Sarmah et al., 2010; Venditti et al., 2012). However, factors facilitating intra/post-Golgi transport of bulky collagen trimers still remain elusive.

The goal of this dissertation is to identify the factors and pathways regulating transport of ECM components during tissue homeostasis and disease conditions. This introduction describes components of the molecular trafficking machinery components and how they specifically regulate ECM secretion during vertebrate development and cause genetic disorders caused by their malfunction. Chapter II describes an RNA-seq based study comprehensively identifying ECM components and trafficking machinery in a highly proficient matrix producing cell type, the zebrafish chondrocyte. Chapter III examines transcriptional regulation of COPII-dependent collagen transport during skeletal morphogenesis in zebrafish. Chapter IV describes a novel trafficking pathway regulating post-Golgi transport of collagen and a previously undiagnosed human syndrome caused by mutations in this pathway. Chapter V describes construction of ‘PredixVU gene-by-medical phenome catalog’ and its application to discover novel disease mechanisms, specifically *GRIK5*-associated eye and vascular diseases. Chapter VI concludes with a discussion of contributions and future implications of this dissertation research to developmental genetics studies, and the pathophysiology of skeletal disorders. Portions of the introduction chapter are adapted from the following review articles:

Unlu, G., Levic, D.S., Melville, D.B., Knapik, E.W., 2014. Trafficking mechanisms of extracellular matrix macromolecules: insights from vertebrate development and human diseases. Int J Biochem Cell Biol 47, 57-67.

Vacaru, A.M., **Unlu, G.**, Spitzner, M., Mione, M., Knapik, E.W., Sadler, K.C., 2014. In vivo cell biology in zebrafish - providing insights into vertebrate development and disease. J Cell Sci 127, 485-495.

Luderman, L.N.*, **Unlu, G.***, Knapik, E.W., 2017. Zebrafish Developmental Models of Skeletal Diseases. Curr Top Dev Biol 124, 81-124.

*) co-first author.

Significance of the trafficking mechanisms of extracellular matrix macromolecules in vertebrate development and human diseases

Cellular life depends on protein transport and membrane traffic. In multicellular organisms, membrane traffic is required for extracellular matrix deposition, cell adhesion, growth factor release, and receptor signaling, which are collectively required to integrate the development and physiology of tissues and organs. Understanding the regulatory mechanisms that govern cargo and membrane flow presents a prime challenge in cell biology. Extracellular matrix (ECM) secretion remains poorly understood, although given its essential roles in the regulation of cell migration, differentiation, and survival, ECM secretion mechanisms are likely to be tightly controlled.

Recent studies in vertebrate model systems, from fishes to mammals and in human patients, have revealed complex and diverse loss-of-function phenotypes associated with mutations in components of the secretory machinery. A broad spectrum of diseases from skeletal and cardiovascular to neurological deficits have been linked to ECM trafficking. These discoveries have directly challenged the prevailing view of secretion as an essential, but uniform process. Here, I will discuss the latest findings on mechanisms of ECM trafficking in vertebrates.

Extracellular matrix (ECM) is a complex non-cellular structure synthesized by all tissues and is composed of water, proteins, and polysaccharides, as well as mineral deposits in skeletal tissues (Bosman and Stamenkovic, 2003). ECM composition is unique to each tissue and is deposited by fibroblasts or other specialized secretory cells. For example, epithelial cells secrete basement membrane proteins, such as collagens, fibronectin and laminin, whereas chondrocytes and osteoblasts secrete type II and type I collagens that are characteristic of mature cartilage and bone, respectively (Gay et al., 1976; Li et al., 1995; Reddi et al., 1977). Large structural ECM proteins are typically fibrillar, including collagens, fibronectin, and laminins. The most abundant of these, collagens constitute over 30% of the total protein mass in multicellular organisms (Ishikawa and Bachinger, 2013) and are rapidly secreted during development or in response to pathological conditions such as wound healing after tissue injury (e.g. skin damage, myocardial infarction, liver cirrhosis) (Cleutjens et al., 1995; Clore et al., 1979; Gay et al., 1975; Pinzani et al., 2011).

The rapid secretion of large cargos such as collagens requires specialized transport machinery (Melville et al., 2011; Saito et al., 2009a). Procollagen has been extensively used as a model-cargo in secretory pathway studies (Arnold and Fertala, 2013; Bonfanti et al., 1998; Ishikawa and Bachinger, 2013; Stephens and Pepperkok, 2002). As with all ECM proteins, procollagen is synthesized and initially post-translationally modified in the Endoplasmic Reticulum (ER), from which it is transported in a COPII (coat protein II complex)-dependent manner to the ER-to-Golgi intermediate compartment (ERGIC) *en route* to the Golgi complex, where further post-translational processing occurs (Canty and Kadler, 2005). Procollagen is, then, transported in tubular carriers to be secreted to the extracellular space, where it is cleaved and assembled into higher-order structures (Arnold and Fertala, 2013; Ishikawa and Bachinger, 2013; Polishchuk et al., 2003; Polishchuk et al., 2009).

The first leg of this journey is export from the ER, which is mediated by the COPII complex (Fig. 1.1). Pioneering work using yeast genetics led to the identification of 23 genes whose products are required for secretory activity (Kaiser and Schekman, 1990; Novick et al., 1980). Among them were components of the COPII complex (Barlowe et al., 1994). COPII formation is initiated when the cytoplasmic GTPase Sar1 undergoes a conformational change upon GTP binding and associates with the ER membrane (Barlowe et al., 1993; Kuge et al., 1994; Nakano and Muramatsu, 1989). Sar1 then recruits Sec23/Sec24 heterodimers to form the "inner coat" complex (Bi et al., 2002; Matsuoka et al., 1998).

Two additional ER associated proteins, Sec12 that acts as a GEF (guanine nucleotide exchange factor) for Sar1 (Barlowe and Schekman, 1993) and Sec16 that is a large scaffold protein shown to associate with ER Exit Sites (ERES) (Connerly et al., 2005; Espenshade et al., 1995; Watson et al., 2006) and contribute to initiation of vesicle formation. While Sec23 serves as GAP (GTPase activating protein) for Sar1, resulting in coat dissociation from the vesicle membrane (Yoshihisa et al., 1993), Sec24 acts as a cargo adaptor by selecting distinct proteins for ER exit (Miller et al., 2002). Assembly of the inner coat is followed by recruitment of the Sec13-Sec31 heterotetramer of the "outer coat" complex, which is thought to stabilize the coat (Bhattacharya et al., 2012; Bi et al., 2007; Copic et al., 2012; Stagg et al., 2006; Tang et al., 2000). The molecular mechanisms of these steps have been reviewed elsewhere (Brandizzi and Barlowe, 2013; Szul and Sztul, 2011).

Unlike the baker's yeast (*Saccharomyces cerevisiae*) genome that harbors single copies of essential COPII genes, vertebrate genomes have an expanded repertoire of COPII genes, including *SAR1A* and *SAR1B* (Jones et al., 2003; Loftus et al., 2012), *SEC23A* and *SEC23B* (Paccaud et al., 1996; Wadhwa et al., 1993), *SEC24A*, *SEC24B*, *SEC24C* and *SEC24D* (Tang et al., 1999), *SEC13* (Swaroop et al., 1994), *SEC31A* and *SEC31B* (Stankewich et al., 2006; Tang et al., 2000), (Fig. 1.1). Gene multiplication of the coat components might have been evolutionarily driven by expansion of the genomes to encode novel extracellular matrix proteins and support more complex body plans.

Many COPII paralogs are specific to vertebrates and therefore they might be involved in unique aspects of vertebrate development, such as formation of diverse basement membranes surrounding internal organs and building internal skeleton, which is primarily composed of cartilage and bone matrix (Braasch and Postlethwait, 2012; Forster et al., 2010; Norum et al., 2010). Thus, it is not surprising that most loss-of-function mutations in trafficking machinery components result in skeletal dysmorphology.

Here I will discuss cargo- and tissue-specific functions of the COPII machinery, post-translational modifications, phosphorylation and ubiquitylation of COPII proteins, and the effects they have on vesicle biogenesis. In addition, I will discuss auxiliary proteins such as cargo receptors and guide proteins that were shown to assist in loading of ECM macromolecules into vesicular carriers. Finally, one of the most intriguing unanswered questions in the regulation of secretion is transcriptional control of the secretory machinery. Most coat genes are ubiquitously expressed but are enriched in specific tissues at defined developmental time points or in pathological conditions. However, little is known about transcriptional control of secretion with only a single factor of the OASIS family, Creb3L2, implicated in the process so far.

Here, I will focus on recently discovered trafficking mechanisms in the initial leg of the secretory pathway, from the ER to Golgi, by highlighting studies of vertebrate model organisms and human genetic mutations.

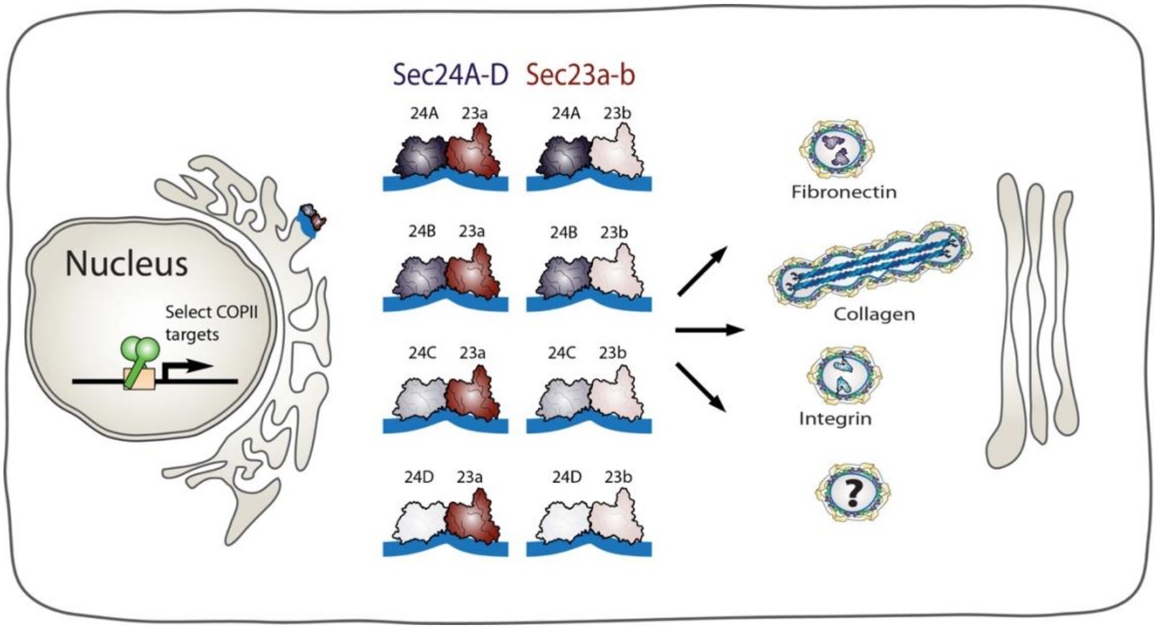


Figure 1.1. The secretory module, Transcription Factor–COPII Adaptor–ECM Cargo, operates in a spatio-temporal manner.

Recent discovery of a "*secretory module*" consisting of Creb3L2, a transcription factor that regulates the expression of Sec23A-Sec24D, which then facilitate procollagen cargo traffic during embryonic skeletal development, provided the first evidence for the "*secretory code*". The existence of such a *secretory code* is supported by studies with mutant animal models and human patient samples discussed in this review. To date only the secretory modules for type I and type II collagens were tested, which were conducted primarily in the zebrafish system using *feelgood-crusher-bulldog* mutations (*creb3L2-sec23A-sec24D*, respectively). It is hypothesized that unknown transcription factors regulate the expression of distinct COPII cargo adaptors (Sec24A-D and Sec23A-B), leading to preferential availability of the various coat components that are required for transport of distinct ECM cargos, such as fibronectin and thrombospondins, at any given time point. Evidence suggests that a diverse array of COPII coats containing specific combinations of core components and associated proteins is required for transport of structurally divergent cargos, such as globular, fibrillar, or transmembrane proteins. Further studies will be required to unravel the complexity of the secretory code to understand how the system integrates cellular operations at regulatory and structural levels in a spatio-temporal manner in a living organism. Adapted from Unlu et al., 2014.

ER-to-Golgi transport is facilitated by Coat Protein II (COPII) vesicular carriers

Cargo selection by Sec24 components of the inner coat

Tandem genome duplication expanded the ancestral *SEC24* gene to two syntenic groups, one of *SEC24A* and *SEC24B* and the second of *SEC24C* and *SEC24D* (Tang et al., 1999). The four genes are highly divergent in sequence between the two groups (20% similarity) and approximately 50% similar between each pair, but each paralog is highly conserved within vertebrates (up to 90% sequence similarity between fish and human) (Sarmah et al., 2010). Pioneering work on cargo recognition by the Sec24 paralogs using model cargo in *SEC24*-depleted HeLa cells revealed for the first time selectivity and redundancy in the cargo selection process (Wendeler et al., 2007). Recent *in vivo* evidence obtained from phenotype-driven genetic screens in zebrafish, medaka and mouse have begun to uncover the complexity of Sec24-based cargo selection. To date, only *SEC24D* has been directly implicated in ECM secretion. The zebrafish mutant *bulldog/sec24d* fails to secrete type II collagen and matrilin from chondrocytes, fibroblasts and notochord sheath cells, leading to severe craniofacial dysmorphology, short body length and kinked pectoral fins (Sarmah et al., 2010). This phenotype is largely recapitulated by the *vbi/sec24d* medaka mutant carrying a nonsense mutation predicted to remove C-terminal portion of Sec24d protein (Ohisa et al., 2010). In both zebrafish and medaka, Sec24d-deficient chondrocytes accumulate type II collagen in distended rough endoplasmic reticulum (rER). However, other ECM and transmembrane proteins including fibronectin, cadherin, and β1-integrin appear to be trafficked normally to the extracellular space and plasma membrane (Table 1.1). In mouse, however, gene-trap mediated knockout of the *Sec24d* gene leads to pre-implantation lethality with no discernible phenotype in a haploinsufficient condition

(Baines et al., 2013). Though the mouse data are consistent with the idea that cargos that are required as early as the 8-cell stage are transported in a Sec24D-dependent manner, the inability to study *Sec24d*-dependent ECM transport during organogenesis in global mouse knockouts highlight the importance and usefulness of studies in teleost fish such as zebrafish and medaka, which receive maternal Sec24d protein and mRNA that allow for normal gastrulation. Indeed, fish models are ideally suited to study cargo adaptor functions at later developmental stages (Melville and Knapik, 2011; Ohisa et al., 2010; Sarmah et al., 2010).

Although Sec24d and Sec24c cargo adaptors recognize similar cargo binding motifs in *in vitro* assays (Wendeler et al., 2007), they appear to transport unique cargos in cell culture and *in vivo* conditions. For example, *Sec24c* has been shown to be essential for secretion of neurotransmitter transporters (Sucic et al., 2011) and the fusion of pre-chylomicron (large lipoprotein complex containing triglycerides) transport vesicles with Golgi membranes (Siddiqi et al., 2010). Furthermore, although Sec24d depletion in zebrafish results in both craniofacial skeleton deficits as well as impaired notochord extension (axial skeleton), Sec24c depletion only affects notochord extension and not the head skeleton. Notably, combined depletion of Sec24c and Sec24d results in a significantly more severe phenotype. These findings suggest that Sec24c and Sec24d are exclusively required for secretion of select notochord basement membrane matrix proteins, whereas other matrix proteins are secreted in a redundant fashion by the two paralogs (Melville and Knapik, 2011; Sarmah et al., 2010). So far, only a few ECM cargos have been matched with specific adaptors. Future in depth studies will be needed to establish what the combinatorial network of cargos and their respective adaptors is, and to understand

Table 1.1. The effects of pre-Golgi trafficking machinery component depletion on cargo transport.
Adapted from Unlu et al., 2014

Trafficking machinery component	Cargo	Effect of depletion on cargo transport	Tissue/Cell Type	Organism	Reference
Sec23a	Collagen I	Accumulation in the ER	Fibroblasts	Human	Boyadjiev et al., 2006
	Collagen II	Accumulation in the ER	Chondrocytes	Zebrafish	Lang et al., 2006
Sec24B	Vangl2	Failure of localization to plasma membrane	Neural Tube	Mouse	Merte at al., 2010
					Wansleeben et al., 2010
Sec24D	Collagen II	Accumulation in the ER	Chondrocytes	Zebrafish	Sarmah et al., 2010
	Matrilin	Accumulation in the ER	Chondrocytes		
	Fibronectin	Unaffected	Chondrocytes		
	Integrin β1	Unaffected	Chondrocytes		
	Pan-cadherin	Unaffected	Chondrocytes		
	Collagen II	Intracellular accumulation	Chondrocytes	Medaka	Ohisa et al., 2010
	Collagen II	Intracellular accumulation	Notochord		
	Collagen II	Intracellular accumulation	Myoseptum		
	GABA transporter 1	Accumulation in the ER	HEK293 cells	Human	Farhan et al., 2007
Sec24C	Serotonin transporter	Reduced surface targeting	HeLa and HEK293 cells	Human	Sucic et al., 2011
Sar1b	Chylomicron	Retention in membrane-bound compartments	Enterocytes	Human	Dannoura et al., 1999
			Enterocytes & chondrocytes	Zebrafish	Levic et al., 2015
Sec13	Collagen I	Defective secretion and deposition	Fibroblasts	Human	Townley et al, 2008
	Opsin	Intracellular accumulation	Photoreceptor cells	Zebrafish	Schmidt et al, 2013
	Syntaxin-3A	Unaffected	Photoreceptor cells		
TANGO1	Collagen VII	Accumulation in the ER	Fibroblasts	Human	Saito et al., 2009b
	Collagen I	Intracellular accumulation	Chondrocytes	Mouse	Wilson et al., 2011
	Collagen II	Intracellular accumulation	Chondrocytes		
	Collagen III	Intracellular accumulation	Chondrocytes		
	Collagen IV	Intracellular accumulation	Endothelial cells		
	Collagen VII	Intracellular accumulation	Embryonic fibroblasts		
	Collagen IX	Intracellular accumulation	Embryonic fibroblasts		
	COMP	Intracellular accumulation	Chondrocytes		
	Fibronectin	Unaffected	Chondrocytes		
	Aggrecan	Unaffected	Chondrocytes		
cTAGE5	Collagen VII	Accumulation in the ER	A431 cells	Human	Saito et al., 2011
Sedlin	Collagen I	Accumulation in the ER	Fibroblasts	Human	Venditti et al., 2012
	Collagen II	Accumulation in the ER	Chondrocytes		

ER: Endoplasmic Reticulum, Vangl2: Van Gogh-like protein 2, TANGO1: Transport and Golgi organization 1, COMP: Cartilage oligomeric matrix protein, cTAGE5: cutaneous T-cell lymphoma-associated antigen 5.

how this network is modulated to meet the secretory demand of different tissues during embryonic development as well as in pathological and homeostasis conditions.

Zebrafish and medaka models of *sec24d* mutants have been particularly helpful guides for the identification of *SEC24D* mutations in a syndromic form of osteogenesis imperfecta (OI). Garber and colleagues conducted a whole-exome sequencing on a 7-year old boy with severe ossification defects and two fetuses with impaired skull formation and fractures throughout the skeleton (Garbes et al., 2015). The study yielded 5 candidate genes including *SEC24D*. The published reports of zebrafish and medaka *sec24d* mutants leading to skeletal defects similar to OI patients, (Ohisa et al., 2010; Sarmah et al., 2010) reporting skeletal defects similar to OI patients led them to prioritize *SEC24D* for Sanger sequencing. Consequently, they identified three functional mutations leading to OI; one nonsense (Q205X) and two missense mutations (Q978P and S1015F) likely to disrupt preotein function. Three individuals included in this study were compound heterozygotes for the reported *SEC24D* mutations. The 7-year old boy presented with short stature, craniofacial dysmorphology, micrognathia and non-fusing sutures in the skull, highly reminiscent of zebrafish *sec24d* (*bulldog*) mutant phenotypes. Furthermore, fibroblasts isolated from the OI patient displayed intracellular procollagen accumulation and distended ER, as determined by immunofluorescence electron microscopy, respectively. Both of these molecular phenotypes were observed in zebrafish *sec24d* mutant chondrocytes (Sarmah et al., 2010) and reported similar results, confirming the conservation of ECM secretion machinery between human and zebrafish, and underscoring the power of zebrafish to model human skeletal disorders.

Genetic findings suggest that *SEC24* genes likely diverged to assume distinct cargo selection properties. In particular, *SEC24B* mutations were recently identified in patients carrying severe neural tube defects (Yang et al., 2013), while *SEC24D* mutations are linked a form of *osteogenesis imperfecta*, presenting with high fragility in long bones (Garbes et al., 2015). Prior work in mice helped to explain why *Sec24b* is required for neural tube closure. Analyses of mice obtained from a phenotype-driven chemical screen showed that *Sec24b* is essential for the secretion of Vangl2 (Van Gogh-like 2), a protein known to act in planar cell polarity (PCP) and be involved in the gastrulation movements of convergence and extension during early embryonic development. Defects in these processes result in craniorachischisis and neural tube closure defects in *Sec24b* mouse mutants (Merte et al., 2010; Wansleeben et al., 2010). Although Vangl2 is not a matrix protein *per se*, its function has been shown to regulate MMP-14 (*membrane type-1 matrix metalloproteinase*), which is involved in fibronectin remodeling (Williams et al., 2012a; Williams et al., 2012b) during gastrulation (Latimer and Jessen, 2010). It will be interesting to determine whether a single or multiple cargo adaptors execute ER egress of multiple components of a single pathway such as PCP.

Sec23A and Sec23B paralogs perform tissue-specific functions

The discovery of tissue-specific phenotypes in carriers of *SEC23* mutations (Boyadjiev et al., 2006; Schwarz et al., 2009) challenged the prevailing view that Sec24 adaptors are solely responsible for cargo specificity of COPII carriers. Vertebrate genomes harbor two paralogs of the ancestral Sec23 gene, Sec23a and Sec23b (Paccaud et al., 1996). Unbiased, phenotype-driven genetic screens in vertebrate model organisms and disease genotyping in human patients provided the first evidence that *SEC23A* and *SEC23B* are not only acting in tissue specific manner, but also are differentially used in ECM macromolecule traffic.

crusher, a zebrafish *sec23a* mutant, was isolated in a genetic screen designed to identify defects in embryonic patterning and organogenesis (Driever et al., 1996; Knapik, 2000; Neuhauss et al., 1996). *crusher* carries a nonsense mutation at amino acid 402 in the Sec23a gene, resulting in a predicted stop codon and truncation of almost half of the protein (Lang et al., 2006). Craniofacial dysmorphology and shorter body length are the predominant phenotypes of *sec23a* mutants, indicating deficits in skeletal development. At the subcellular level, *crusher* chondrocytes exhibit distended rough endoplasmic reticulum (rER) in electron micrographs (TEM) and accumulate type-II collagen deposits as shown by immunofluorescence (IF). This intracellular backlog consequently leads to reduced Collagen and Matrilin content in cartilage ECM.

In parallel with discoveries of *sec23a* function in zebrafish, mutations in the human *SEC23A* gene were reported to cause cranio-lenticulo-sutural-dysplasia (CLSD), an autosomal recessive disorder characterized by facial dysmorphism and axial skeleton defects (Boyadjiev et al., 2006). Electron microscopy and IF studies in fibroblasts isolated from CLSD patients revealed dilated ER structures and accumulation of procollagen in rER cisternae. To date, two missense mutations located near the Sec31 binding site of the SEC23A folded protein were identified in patients. The F382L mutation was shown to hinder SEC23A's ability to recruit the SEC13-SEC31 outer coat and to ultimately prevent vesicle budding (Fromme et al., 2007), whereas the M702V appears to activate SAR1B more efficiently than the wild type allele, resulting in premature dissociation of the COPII coat from ER membranes (Boyadjiev et al., 2011; Kim et al., 2012). Although other cargo molecules are packaged into COPII vesicles normally, procollagen accumulates in the ER of M702V mutant fibroblasts. These results suggest that COPII vesicles with a longer occupancy on the ER membrane may be required to form sufficiently large carriers to transport procollagen (Kim et al., 2012). The zebrafish *crusher* and two CLSD mutations affect approximately the same region of the Sec23A protein and result in remarkably similar phenotypes. This highlights the high degree of conservation of COPII-mediated collagen transport and further establishes zebrafish as a strong *in vivo* system to model human mutations (Vacaru et al., 2014).

In contrast to those of *SEC23A,* human *SEC23B* mutations in humans, in contrast to those of *SEC23A*, are linked to Congenital Dyserythropoietic Anemia Type II (CDAII), an autosomal recessive disease characterized by ineffective erythropoiesis, hemolysis, and the presence of multinucleated erythroblasts in bone marrow. The precise molecular mechanisms leading to these phenotypes are not understood, but transmission electron micrographs (TEM) of erythrocytes in peripheral blood showed double plasma membranes, and SDS-PAGE experiments revealed hypoglycosylation of membrane proteins (Bianchi et al., 2009), suggesting a requirement of SEC23B in the trafficking of components of glycosylation pathway components. Over 50 variants in *SEC23B* have been identified throughout the length of the coding region, and most CDAII patients are homozygous for missense mutations or compound heterozygotes presenting primarily with an anemia phenotype (Iolascon et al., 2010; Khoriaty et al., 2012; Russo et al., 2010; Schwarz et al., 2009). No homozygotes of nonsense mutations have been identified in approximately 370 reported cases (Iolascon et al., 2012). Interestingly, three independent gene-trap insertion lines in mouse, each of which results in predicted null alleles, do not present an anemia phenotype and die at birth with profound developmental and exocrine organ (pancreas, salivary and intestinal glands) secretion deficits (Tao et al., 2012). However, zebrafish larvae depleted in *sec23b* present an anemia phenotype and hemolysis similar to CDAII patients (Fig. 1.2A) (Schwarz et al., 2009), and in addition they lack the entire neural crest-derived craniofacial skeleton (Lang et al., 2006; Schwarz et al., 2009). At present, it is not clear whether the range of phenotypes represents species-specific functional differences (human-mouse) or variations between hypomorph and null alleles. One potential explanation for the tissue-specific phenotypes observed with *sec23a* and *sec23b* mutations is that the two paralogs are required to transport distinct cargos. This notion, however, contrasts with the prevailing view that Sec24, but not Sec23, participates in COPII cargo selection. Whether Sec23b is required (directly or indirectly) to transport unique cargos from those of Sec23a remains unknown, and future studies will be needed to address these questions to better explain tissue- and species-specific phenotypes.

Although Sec23 is not known to possess intrinsic cargo-binding activity, Sec23a and Sec23b could indirectly convey cargo specificity to the COPII coat by selective interaction with Sec24 proteins, which are known to directly interact with cargos (Barlowe, 2003; Mancias and Goldberg, 2008; Miller et al., 2002; Miller et al., 2003). In such a scenario, Sec23a may be viewed as a critical partner for a collagen-specific Sec24 paralog. Alternatively, Sec23 could participate in direct cargo/receptor/adapter binding through undiscovered mechanisms. The complexity of ECM cargo selection is just being uncovered in an *in vivo* setting of a vertebrate body, and it has become clear that this is a complex problem that will require extensive investigation.

Sar1A and Sar1B may differentially contribute to carrier size

Sar1 is a small GTP-binding protein that initiates COPII coat assembly on ER membranes (Aridor et al., 2001; Bielli et al., 2005; Kuge et al., 1994). Similar to other COPII components, vertebrate genomes harbor two Sar1 genes, *SAR1A* and *SAR1B* (He et al., 2002; Jones et al., 2003; Levic et al., 2015). The two paralogs are highly related and the human genes vary by only 20 residues, whereas the fish and mammalian homologs share over 90% sequence identity. Despite this remarkable similarity in sequence the two paralogs function in a distinct manner.

The distinct functions might be explained by differential interactions of Sar1a and Sar1b with the outer COPII coat. Under Sec31-stimulated conditions, Sar1b hydrolyzes GTP more slowly than Sar1a, possibly due to higher binding affinity of Sec13-Sec31 to Sar1a than to Sar1b containing coats (Fromme et al., 2007). Conceivably, this high-affinity binding of Sar1a to the outer coat could lead to tightly packaged small vesicles, whereas loosely packed Sar1b-outer coat complex may allow for larger COPII carriers (Fromme et al., 2008). Although low-affinity binding of Sar1a to the outer coat could potentially help to explain how enlarged COPII carriers are formed, it remains elusive whether different binding affinities of Sar1a and Sar1b for the COPII outer coat observed *in vitro* would translate into functional differences in a physiological context. Alternatively, for example, Sar1b may accommodate the formation of larger sized carriers primarily due to its slower

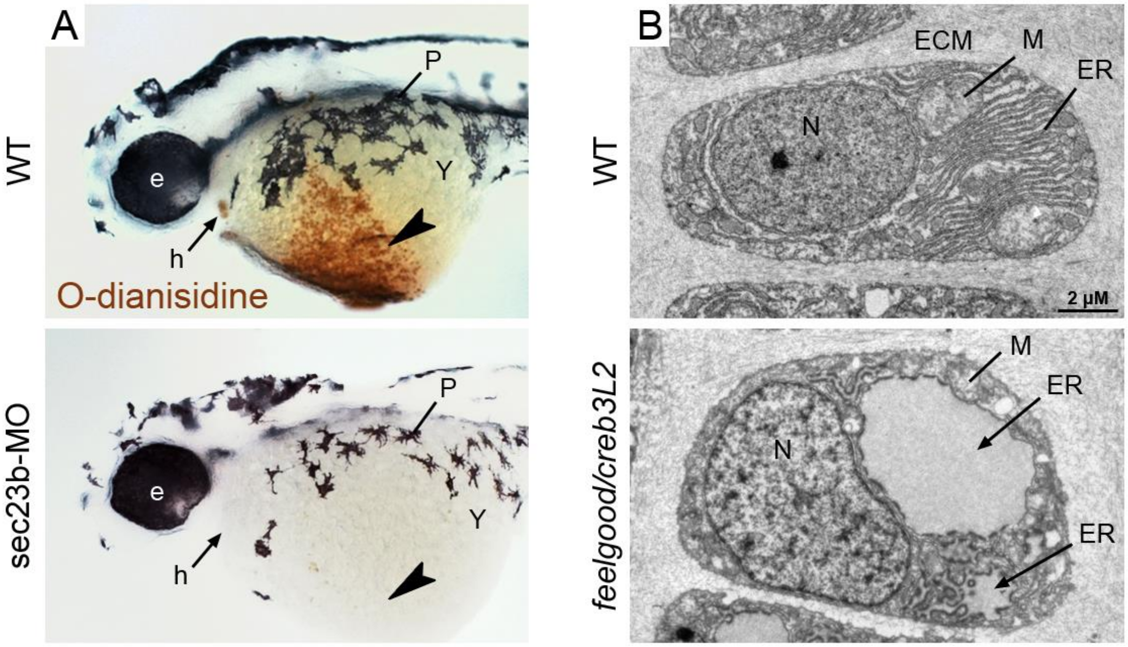


Figure 1.2. Deficits in secretory machinery lead to tissue-specific phenotypes.
A. Wild-type 2-day zebrafish embryos stained with o-dianisidine, which is oxidized by heme in the presence of peroxide and colored brown, display abundant hemoglobin within the ducts of Cuvier (arrowhead) and the heart (arrow). *Sec23b* morphants, which have defects in erythrocyte development, are deficient of hemoglobin at 2 days. **B.** Transmission Electron Micrographs (TEM) images of wild-type zebrafish chondrocytes (3-day old) show characteristics of highly secretory cells, including abundant rough ER (arrow), mitochondria (line), and Golgi. *feelgood/creb3L2* mutant chondrocytes contain highly distended rER (arrow) due to collagen backlog. Symbols: e, eye; h, heart; P, pigment; Y, yolk; ECM, extracellular matrix; ER, endoplasmic reticulum; M, mitochondria; N, nucleus. Adapted from Unlu et al., 2014.

rate of GTP hydrolysis, which slows down the kinetics of vesicle budding to facilitate larger vesicles to form.

Consistent with the carrier size hypothesis, chylomicrons (75-600 nm in size) are significantly larger than typical 90 nm COPII vesicles, and are observed to be preferentially secreted in a Sar1b-dependent manner (Shoulders et al., 2004). Mutations in human *SAR1B* are cause chylomicron retention disease (CMRD), a lipid absorption disorder characterized by deficits in intestinal lipid uptake and hypocholesterolemia (Jones et al., 2003). Besides lipid malabsorption, some CMRD patients are diagnosed with other symptoms, including exocrine pancreatic insufficiency, decreased bone mineral density, and cerebellar ataxia. The pathophysiology of the disease is not understood; however, *sar1b* loss-of-function experiments in zebrafish embryos showed not only intestinal lipid absorption deficits but also craniofacial dysmorphology due to failure of type II collagen secretion to extracellular cartilage matrix (Levic et al., 2015). These findings suggest that Sar1b is needed not only for secretion not only of lipids in large chylomicrons, but also large extracellular matrix proteins such as collagen.

Interestingly, no mutations in the *SAR1A* coding region have been identified in patients (Charcosset et al., 2008; Kumkhaek et al., 2008) and depletion of Sar1a in zebrafish did not result in a gross dysmorphology (Levic et al., 2015). Recent experiments in mammalian cell culture setting examined the roles of Sar1a and Sar1b and revealed that cells depleted in both paralogs were still capable of secreting small globular proteins (VSV-G) in a COPII-independent manner using a previously uncharacterized, atypical COPI-dependent secretory mechanism. Furthermore, procollagen type I was retained in the ER and did not sort to ERES (Cutrona et al., 2013). These findings open a new, intriguing possible mechanism of ECM transport, particularly because COPI genetic mutations in zebrafish have deficits in secretion of notochord basement membrane proteins (Coutinho et al., 2004). For example, *copa* depletion results in a highly similar phenotype in the craniofacial skeleton compared to *sec23a* mutants alone; however, combined *copa/sec23a* mutants had a significantly shorter axial skeleton than either condition alone (Lang et al., 2006). These results suggest that distinct trafficking pathways are required for morphogenesis of the two tissues, which are composed of distinct ECM components. Further investigations are needed to better understand the function of the COPII inner coat in ECM secretion, as well as the regulatory mechanisms that provide the proper combinations and stoichiometry of different inner coat paralogs to meet the secretory demand of various tissues and organs – the so-called *secretory code*, which remains an elusive riddle in cell biology (Fig. 1.1).

Sec13-Sec31of the outer coat contributes to cargo specificity

Sec13-Sec31 heterotetramers form the outer shell of the COPII coat (Stagg et al., 2006) and are thought to provide rigidity to its structure (Copic et al., 2012). Most vertebrate genomes contain at least two Sec31 genes, but only one Sec13 (Stankewich et al., 2006; Swaroop et al., 1994; Tang et al., 2000). To date, no human patients have been reported to carry mutations in the outer coat genes. The absence of Sec13 and Sec31 syndromes might be a result of *in utero* lethality, or alternatively might reflect redundant and non-essential functions of these genes in human development. Further characterization of loss-of-function phenotypes in vertebrate model organisms will help narrow down the spectrum of human syndromes that could be tested for potential mutations in Sec13 and Sec31.

Sec13 has been suggested to play critical roles in development of craniofacial structures. In zebrafish, knockdown of *Sec13* leads to morphological abnormalities in craniofacial cartilage elements, short pectoral fins, small eyes and cardiac edema, likely resulting from defects in proteoglycan and collagen secretion (Townley et al., 2008). At the molecular level, Sec13 depletion in cultured human cells led to loss of Sec31 from the outer coat of COPII vesicles, although budding and curvature of the vesicles were unaffected as determined by TEM analyses (Niu et al., 2012; Townley et al., 2008). Furthermore, Sec13 depleted cells failed to export collagen from distended rough ER (RER), whereas tsO45-VSV-G-YFP, a secretory cargo marker, and other small soluble cargos were normally secreted (Cutrona et al., 2013).

Sec13 is also implicated in the development of digestive system organs. The zebrafish s*ec13* genetic mutation leads to hypoplastic intestine, exocrine pancreas, and liver. The gene appears to be dispensable for initial specification and patterning stages of development, but is essential for tissue growth and cell proliferation (Niu et al., 2012). Similar morphological deficits in intestinal epithelium morphology were observed in a morphant model, further emphasizing the requirement of Sec13-Sec31-driven secretion in the development of the digestive system (Townley et al., 2012). A small eye phenotype in Sec13-depleted zebrafish embryos was linked to degeneration of photoreceptor cells, in addition to impaired trafficking of collagen in retinal pigment epithelium, supporting the model that efficient assembly of the COPII outer coat is required for trafficking of large ECM cargos in skeletal tissues, tubular organs and the retina (Schmidt et al., 2013).

Collectively, these studies underscore the importance of a fully functional Sec13-Sec31 outer coat for the development of various, highly secretory organs such as cartilage, exocrine pancreas, liver, and retina. The outer coat appears to be required for the transport of bulky collagen cargos as opposed to small soluble proteins. Further studies in vertebrate model organisms and patient samples will be required to identify other ECM cargos secreted in a Sec13-31-dependent manner and to identify determinants of the secretory code.

Transcriptional regulation of ECM secretion

An unsolved problem in cell biology is how regulatory mechanisms directing development and adult homeostasis communicate with secretory pathway programs to assure sufficient and timely availability of cargo-specific coats. In a living multicellular organism, the secretory pathway responds to constant demands for cargo delivery during development, physiological changes and tissue repair. The transcriptional mechanisms and signaling pathways that govern this coordination are just being illuminated. The first identified *bona fide* transcriptional regulators of the secretory machinery have recently been identified as CREB3L2 (*cAMP responsive element binding protein-3 like 2*) in mice and zebrafish; and CrebA (Creb3L2 ortholog) in flies (Abrams and Andrew, 2005; Fox et al., 2010; Melville et al., 2011; Saito et al., 2009a; Tanegashima et al., 2009).

Creb3L2, also known as BBF2H7, has been linked to collagen secretion in vertebrates. Creb3L2 is a transmembrane transcription factor, highly expressed in cartilage, that is synthesized in the ER and traffics to the cis-Golgi. After COPII-dependent ER export, Creb3L2 is cleaved in the cis-Golgi to an active, soluble peptide that dimerizes and then is translocated to the nucleus to activate the expression of target genes (Lui et al., 2008; Panagopoulos et al., 2007). *Creb3L2* knockout mice display chondrodysplasia and die shortly after birth (Saito et al., 2009a). Proliferating chondrocytes in E18.5 stage mice have distended ER and intracellular collagen II accumulation. In cultured ATDC5 cells, luciferase assays revealed that *Sec23a* promoter activity is regulated by CREB3L2, while chromatin immunoprecipitation showed binding of CREB3L2 to the *Sec23a* promoter region. Taken together, this study linked CREB3L2-mediated transcriptional regulation of *Sec23a* expression to collagen trafficking.

Concurrent study of the zebrafish *feelgood* mutation identified a missense variant in the *creb3L2* coding region resulting in a hypomorphic allele (Melville et al., 2011). Luciferase assays of the *feelgood* variant revealed an approximately 50% reduced transcriptional activity to and consequently, a significant decrease in *sec23a* expression by quantitative PCR experiments. Further analyses showed that Creb3L2 is required for the transcription of specific subsets of the COPII machinery, such as Sec24d but not Sec24c. In addition, trafficking of collagen II in chondrocytes and collagen IV in notochord sheath cells were disrupted, whereas other cargos including laminins were not affected. These data suggest that spatio-temporal expression of COPII components is differentially regulated by Creb3L2 and possibly by other, yet unknown transcription factors. TEM analyses in *feelgood* chondrocytes revealed a backlog of protein secretion and a distended ER structure (Fig. 1.2B) as well as progressive loss of cartilage matrix. One potential interpretation of the *feelgood* phenotype is that in the initial phase of collagen secretion, the cargo load is modest and available coats are sufficient to initiate the process. As matrix secretion increases over the course of development, available coats are not able to meet the demands of the cell and ECM deposition decreases, resulting in progressive reduction of the collagen fibrils in growing skeletal tissues. Creb3L2 is, so far, the only transcription factor identified as a regulator of COPII-dependent collagen secretion. Future studies are needed to discover the full range of transcriptional regulatory mechanisms governing the expression of specific COPII components and mediators of post-translational modifications mediating ECM secretion.

Post-translational modifications of the trafficking machinery

Recent findings have highlighted the importance of an additional level of regulation in ECM secretion through post-translational modifications of COPII components that have regulate COPII vesicle architecture and assembly. For example, phosphorylation and ubiquitination of COPII proteins can change the affinity of individual coat components for each other and can impact vesicle size.

Monoubiquitination

Monoubiquitination (addition of a single ubiquitin molecule) has emerged recently as a novel mechanism to modulate protein function (Lauwers et al., 2009), unlike polyubiquitination, which initially characterized as a signal for proteasomal degradation in addition to its recently reported functions (Angers et al., 2006; Sadowski et al., 2012). The Cullin3 (*Cul3, cullin-based E3 ligase*) and KLHL12 (*kelch-like family member 12*) complex is a ubiquitin ligase that monoubiquitinates Sec31 (Jin et al., 2012; Stephens, 2012). This modification is required to drive procollagen secretion. The concerted action of the cytoplasmic Cul3-KLHL12 complex, under overexpression conditions in a mammalian cell culture system, leads to production of enlarged COPII-coated carriers that are up to 500 nm in diameter – large enough to accommodate procollagen molecules. However, this modification is not required for trafficking of smaller or more flexible cargos such as fibronectin, EGF receptors, or integrin β1 (Jin et al., 2012). Despite this important discovery, the precise mechanism of how Sec31 monoubiquitination regulates the size of COPII coated carriers in collagen traffic is unknown. Particularly, it remains to be established how the cytoplasmic Cul3-KLHL12 complex recognizes secretory cargo within the ER lumen, which is separated from the complex by ER membranes. Conceivably, extra time required to package bulky cargo like collagen could prolong the time Sec31 spends on a budding vesicle, which may provide Cul3-KLHL12 complex enough time to target Sec31. In any case, Sec31 appears to be a key target for post-translational regulation of COPII carrier's size as it is both monoubiquitinated and phosphorylated.

Phosphorylation

Although Sec31 was initially detected as a phosphoprotein many years ago (Salama et al., 1997; Shugrue et al., 1999), casein kinase II (CK2) has only recently been identified as a kinase responsible for Sec31 phosphorylation. This modification was suggested to regulate Sec31 association with ER membranes through interaction with the COPII inner coat (Koreishi et al., 2013). Ultracentrifugation assays revealed that phosphorylated Sec31 has reduced membrane association, whereas a non-phosphorylatable Sec31 mutant remains at the ERES longer and binds more strongly to Sec23. A model where Sec31 phosphorylation interferes with its binding to the COPII inner coat, which ultimately delays vesicle budding, has been proposed (Koreishi et al., 2013). Though preliminary, this model could explain a regulatory mechanism that controls COPII vesicle size. Notably, the function of other kinases in phosphorylation of COPII outer coat components has not been investigated *in vivo* (Dephoure et al., 2008; Franz-Wachtel et al., 2012; Olsen et al., 2006). Identification and characterization of the phosphatase(s) responsible for dephosphorylation of COPII coat subunits will help in further understanding the molecular mechanisms underlying the regulation of collagen secretion.

Akt (*Protein kinase B*) phosphorylates recombinant human SEC24C and SEC24D. SEC24 proteins phosphorylated by Akt show greater affinity for SEC23 as detected by co-IP from CHO-7 cells (Sharpe et al., 2011). Although specific roles of SEC24D phosphorylation in collagen trafficking has not been investigated, differential phosphorylation of individual SEC24 paralogs could explain COPII coat diversity in accommodating ECM cargo.

Collagen secretion is critical to homeostasis of the arterial wall, and malfunction can result in blood vessel rupture. *PKCδ* (*Protein kinase C-, member of the family of serine/threonine kinases*) has been implicated in collagen -I secretion in smooth muscle cells. PKCδ knockout mice display reduced collagen-I deposition to the arterial wall and intracellular accumulation in surrounding smooth muscle cells. Backlogged collagen-I was primarily found in the TGN (*trans*-Golgi network) of the Golgi complex. PKCδ-null smooth muscle cells exhibited reduced levels of the Rho GTPase Cdc42, and restoration of Cdc42 rescued collagen I secretion defects (Lengfeld et al., 2012). This study supports a model wherein PKCδ phosphorylates yet unidentified collagen I secretion factors in a

Cdc42-dependent manner. This study has identified a novel phenotype in the collagen secretory pathway and future studies might uncover its specific functions in ECM cargo transit through the Golgi complex (Lengfeld et al., 2012).

Auxiliary proteins supporting collagen secretion

Recent phenotype-driven genetic screens have identified two novel proteins that are associated with the COPII coat machinery and are essential for efficient packaging of large ECM cargos. Tango1/Mia3 and cTage5 are transmembrane proteins of the ER (Saito et al., 2009b; Saito et al., 2011). They bind to Sec23-Sec24 subunits of the COPII coat on their cytoplasmic side and to procollagen on their ER luminal side. Tango1 and cTage5 are postulated to aid cargo selection and concentration into carriers, as well as to delay vesicle scission to allow extra time for cargo loading.

Tango1 / Mia3

Tango1 (*transport and Golgi organization 1*) was identified in a genome-wide screen for genes required for constitutive protein secretion using *Drosophila* S2 cells (Bard et al., 2006). Tango1 resides at ERES and potentially acts as a guide protein for large cargo loading (Fig. 1.3) (Saito et al., 2009b). The protein contains two well-characterized domains: a C-terminal proline rich domain (PRD) required for its localization to the ERES, and a luminal SH3 (SRC homology 3) domain interacting with procollagen, as shown by immunoprecipitation assays using transfected COS7 cells. Tango1 is essential for collagen VII transport, and its depletion in cultured skin fibroblasts results in a backlog of collagen VII in the ER (Saito et al., 2009b). The same study showed that Tango1 does not get packaged into COPII vesicles, but remains at ERES after collagen loading is completed. Tango1 is suggested to work as a guide by binding to collagen via its SH3 domain to help bulky collagen molecules to be packaged into vesicles. Aside from its interaction with procollagen, Tango1 can also bind the Sec23-Sec24 inner coat through its proline-rich domain (Bi et al., 2007; Shaywitz et al., 1997; Shugrue et al., 1999). Tango1 is proposed to antagonize Sec31 during budding to slow down vesicle biogenesis while procollagen molecules are being loaded; once the loading process is completed, Sec31 can be recruited to a procollagen-filled vesicle upon dissociation of Tango1 from both procollagen and Sec23-Sec24 complex through a conformational change (Fig. 1.3).

Developmental roles of Tango1 were investigated in mouse models. Tango1 knockout mice exhibit global secretion defects in collagen I, II, III, IV, VII and IX from various cell types such as chondrocytes, fibroblasts and endothelial cells. The overall development of Tango1 knockout mice is severely compromised, resulting in dwarfism and edema (Wilson et al., 2011). In Tango1-null mice, ECM deposition is reduced, leading to deficits in cartilage and bone development. These analyses in knockout mouse have revealed critical and widespread roles for Tango1 in vertebrate development, especially for global collagen secretion and skeletogenesis.

cTAGE5

cTAGE5 (*cutaneous T-cell lymphoma-associated antigen 5*) was originally found to be overexpressed in various tumors and considered a tumor-specific antigen (Heckel et al., 1997), and has recently been associated with collagen trafficking in mammalian cells.

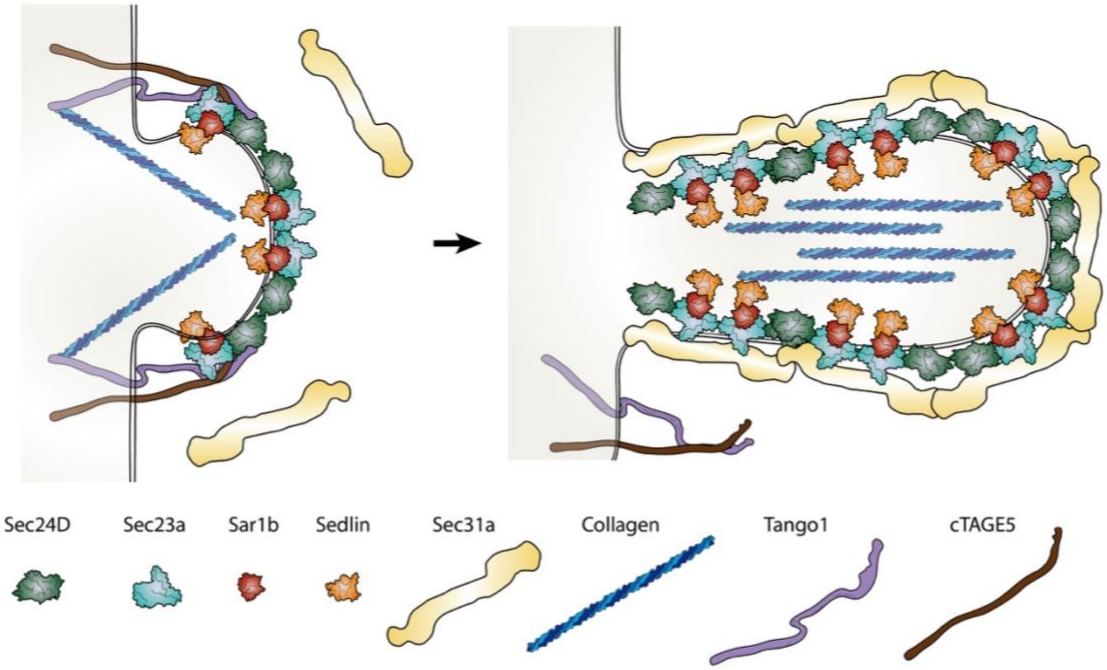


Figure 1.3. Packaging of procollagen fibrils into large COPII carriers.
Procollagen is a rigid, fibrillar protein aggregate that is significantly larger than the typical size COPII-coated vesicles. Recent work has uncovered auxiliary proteins that aid in the packaging and transport of procollagen into mega-sized COPII carriers. Procollagen is initially loaded into budding vesicles through the concerted action of transmembrane proteins TANGO1 and cTAGE5, which both bind to the Sec23-Sec24 inner coat complex on the cytoplasmic side and collagen on the luminal side. TANGO1/cTAGE5 interaction with the inner coat is thought to inhibit the association of the COPII outer coat complex with the inner coat, which delays the fission of vesicles from the ER exit sites and results in the formation of large-size carriers. TANGO1 is also essential for recruiting Sedlin, which interacts with Sar1 and provides efficient cycling of Sar1-GTP hydrolysis, further delaying coat dissociation from the membranes. After procollagen loading is completed, TANGO1 undergoes a conformational change and dissociates from both procollagen and the inner coat complex but is left behind in the ER membrane after COPII carrier fission. Recruitment of Sec13-Sec31 outer coat is the final step of coat formation before fission. Adapted from Unlu et al., 2014.

cTAGE5 is an integral membrane protein that localizes to ERES. It contains a transmembrane domain, a proline rich-domain and coiled-coil domains.

Immunoprecipitation experiments showed that cTAGE5, via its coiled coil motifs, interacts with Tango1 in mammalian cells. Yeast-two-hybrid assays revealed an interaction between the proline-rich domain of cTAGE5 and the Sec23-Sec24 inner COPII coat complex. Moreover, knockdown of cTAGE5 in mammalian cells resulted in accumulation of collagen VII within the ER (Saito et al., 2011). These results suggest that cTAGE5 may serve as an essential co-receptor for Tango1 to facilitate packaging of procollagen into COPII carriers. This notion is supported by the following observations: (1) cTAGE5 can directly interact with Tango1 and (2) knockdown of cTAGE5 leads to collagen export defects regardless of the presence of proper localization and expression of Tango1 (Saito et al., 2011).

Sedlin

Although Tango1 and cTAGE5 are known to facilitate loading of procollagen into COPII vesicles, their action can only partially explain the mechanisms governing the growth of large COPII carriers. Sedlin, a TRAPP (*TRAfficking Protein Particle*) component that was reported to be defective in spondyloepiphyseal dysplasia tarda (SEDT) patients (Davis et al., 2013; Gedeon et al., 1999; Gedeon et al., 2001; Matsui et al., 2001; Mumm et al., 2000; Mumm et al., 2001) has been implicated in procollagen export from the ER. The *SEDLIN* gene is mutated in SEDT patients with chondrogenesis defects. In *SEDLIN*-depleted chondrocytes, procollagen accumulates in the ER while small cargos are trafficked properly. Further analysis showed that Sedlin is recruited to ERES in a Tango1-

dependent manner. In the absence of Sedlin, Sar1-GTPase cycle is hyperactive, which results in premature membrane constrictions as detected by electron tomography and 3D reconstruction analysis in fibroblasts from SEDT patients (Venditti et al., 2012). A current model proposes that Tango1-mediated Sedlin recruitment to ER exit sites facilitates efficient Sar1-GTPase cycles to stabilize inner COPII coat and prevent premature membrane constrictions.

Approaches to model ECM-based skeletal disorders in zebrafish

Vertebrate skeletogenesis begins with chondrogenesis, a process by which cartilage is formed and ossified into bone, and involves two highly conserved processes: endochondral ossification and intramembranous ossification (Berendsen and Olsen, 2015; Goldring et al., 2006). To begin, mesenchymal stem cells migrate to locations of future bones and condense before differentiating into either chondrocytes or osteoblasts (Berendsen and Olsen, 2015). Endochondral ossification involves newly differentiated chondrocytes secreting cartilage ECM that will ultimately turn into bone in areas including the skull posterior and base, axial skeleton, and appendicular skeleton (Berendsen and Olsen, 2015). Mesenchymal stem cells differentiated into osteoblasts will directly produce bone as part of the membranous neuro- and viscerocranium and clavicle through intramembranous ossification (Berendsen and Olsen, 2015). Secretion of ECM products will ultimately be used to signal growth of cartilage, bone and muscle of the skeletal system.

Surrounding the mesenchymal stem cells are epithelial cells, which contribute to the process of chondrogenesis through signaling for stem cell differentiation. Chondrocyte maturation markers, such as Sox9, act as transcriptional activators required for the production and secretion of ECM components such as collagen type-II and aggrecan. As chondrocytes mature and grow, they intercalate and flatten to form cartilage elements like the pharyngeal apparatus (Kimmel et al., 1998). Newly differentiated chondrocytes secrete a cartilage matrix into the extracellular space. These cartilage components include collagens II, IX, XI, and proteoglycans cartilage. Proteoglycans produced and secreted by chondrocytes include both heparan sulfate (HSPG) and chondroitin sulfate (CSPG) modifications (Holmborn et al., 2012).

Zebrafish loss-of-function models of cartilage ECM synthesis genes allow efficient analysis of cartilage development and mutant phenotypes and can be assessed through a variety of techniques. Zebrafish, as a model organism, offers multiple gene depletion strategies including classical chemical mutagenesis with N-ethyl-N-nitrosourea (ENU), insertional mutagenesis, gene knockdown via morpholino-oligonucleotide injection, and newer techniques such as TALEN, zinc-finger nuclease (ZFN) and CRISPR/Cas9 genome editing approaches (Vacaru et al., 2014).

ENU-based forward genetics screens have been instrumental in identifying skeletal development models in zebrafish (Andreeva et al., 2011; Driever et al., 1996; Knapik, 2000; Neuhauss et al., 1996). These unbiased screens revealed previously unidentified factors essential for craniofacial and vertebral development. With forward genetics approaches, mutations in genes encoding for cartilage and bone matrix components have been identified, and these zebrafish models have since been used to understand hereditary skeletal disorders. Identification of mutated genes in zebrafish has been fast and reliable using positional cloning strategies due to availability of genetic linkage maps and microsatellite markers (Bradley et al., 2007; Fornzler et al., 1998; Knapik et al., 1996; Knapik et al., 1998).

Remaining questions and challenges in ECM secretion pathways

Despite sustained interest and advances in understanding, numerous questions remain about fundamental cargo selection mechanisms of the COPII inner coat, and. For example, are Sec24 paralogs the only inner coat subunits that function in cargo selection, or are the Sec23 paralogs also capable of influencing cargo selection/traffic? Are all Sec23 and Sec24 paralogs present at stoichiometric levels within individual cells during embryonic development, tissue repair, and at homeostatic conditions? Alternatively, do transcriptional and post-translational regulations dynamically balance relative levels of inner coat paralogs to meet secretory demand? What types of cargos use specific adaptors? Are the processes of cargo and adapter expression regulated synchronously or dynamically during cellular differentiation? Is there an overarching "secretory code" that could predict cargo-adaptor relationships, and could such understanding be used therapeutically as "druggable targets" to promote or suppress secretion of factors involved in processes such as cancer metastasis and cellular differentiation? What are the functions of conserved and divergent domains in cargo adaptors, and do these domains confer cargo specificity directly or indirectly?

These unanswered questions have critical human health implications since coordinated function of secretory pathway is essential for organ and cell physiology. Particularly, establishment of basement membranes and extracellular matrices, as well as cell polarity, migration, adhesion, and secretion of growth factors rely very heavily on efficient function of secretory machinery. All of these events are tightly regulated, and the precise availability of COPII components appears to be required to deliver these molecules in a coordinated, spatiotemporal fashion. In summary, recent discoveries of secretory machinery functions in ECM transport using *in vivo* vertebrate model systems, as well as elegant *in vitro* and culture models, have created new paradigms for understanding secretory biology. Continued research will help to translate these principles into a therapeutic framework.

This thesis study contributed to better understanding of mechanisms regulating ECM trafficking during homeostasis and disease. During this thesis work, I identified novel factors and regulatory axes modulating collagen trafficking. Taking a comparative transcriptomics approach to identify matrix genes and secretory pathway components significantly enriched in live zebrafish chondrocytes, I identified a regulatory axis wherein Creb3L2 transcriptionally activates *sec24d* and *sec23a* to promote collagen transport in chondrocytes. Genetic and biochemical assays functionally validated that Creb3L2-Sec23a-Sec24d collagen secretory axis functions to ensure efficient collagen secretion during craniofacial cartilage development in zebrafish.

This thesis provides an answer to long debated topic of post-COPII fate of collagen in the secretory pathway. Using zebrafish chondrocytes, notochord sheath cells and human dermal fibroblasts, I discovered that Rab6a activation by Ric1-Rgp1 GEF complex is required in trans-Golgi network for efficient collagen secretion. Disruption of this axis leads to craniofacial malformations in zebrafish larvae. Furthermre, we identified a novel *RIC1*-linked human syndrome, which we named CATIFA for common symptoms: Cataract, cleft lip, Tooth abnormality, Intellectual disability, Facial dysmorphism and ADHD.

Lastly, this thesis study describes, for the first time, application and functional validation of PredixVU, a gene-based PheWAS catalog developed using PrediXcan and BioVU biobank. Taking advantage of PredixVU catalog, we built a three-prong approach, based on animal models, Mendelian conditions and biobanks, to study disease mechanisms. Application of this approached yielded novel RIC1-associated phenotypes, which were consistent in CATIFA patients, and also validated in zebrafish models. Independently, we conducted a genome-wide analysis using gene-based PheWAS method and identified phenome associated with reduced expression of *GRIK5*. Performing functional studies in *grik5*-depleted zebrafish models and reiterating the findings in biobanks, we identified increased joint risk of eye and vascular diseases in reduced *GRIK5*-expression conditions.

CHAPTER II

THE CHONDROCYTE REFERENCE TRANSCRIPTOME REVEALS NOVEL CARTILAGE MARKERS AND DISEASE ORTHOLOGY FROM ZEBRAFISH TO HUMANS

Gokhan Unlu[1,2,4], Rui Chen[2,3,5], Bingshan Li[2,3,5], Ela W. Knapik[1,2,4]

[1]Division of Genetic Medicine, Department of Medicine; [2]Vanderbilt Genetics Institute, Vanderbilt University Medical Center; [3]Center for Quantitative Sciences, [4]Department of Cell & Developmental Biology, [5]Department of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN 37232, USA

Abstract

Chondrocytes produce extracellular matrix (ECM) and pattern tissues such as skeletal joints, rings of trachea and nostrils. Despite critical physiological roles in skeletal biology, the transcriptome of chondrocytes has been difficult to obtain and consequently sparsely sampled. Here, we developed a robust protocol to isolate chondrocytes embedded in dense matrix from live zebrafish cartilage, allowing us to perform a deep transcriptome analysis. FACS isolated chondrocytes expressing EGFP from the *col2a1a* promoter together with remaining GFP-negative cells of the head were used for library construction and next generation RNA-seq analysis. By comparing the chondrocyte to non-chondrocyte transcriptome, we identified 4,033 genes enriched in wild type chondrocytes. Gene ontology and pathway analyses showed high expression of ECM enzymes, core components of the collagen secretory machinery (e.g. *sec23a*, *sec24d*), and developmental genes (e.g. *creb3l2*) in chondrocytes. Defects in chondrocyte function lead to developmental disorders and skeletal deformities. We used orthology analysis and comparative transcriptomics to identify 211 common genes highly expressed in zebrafish chondrocytes and human fetal cartilage. Disease association analysis identified 110 human orthologs as associated with skeletal dysmorphology, dysplasia or joint diseases. Remaining ones were previously unrecognized for their role in cartilage biology. The chondrocyte isolation protocol and the transcriptome datasets will help advance discovery of basic cellular processes in cartilage biology and the underlying causes of undiagnosed skeletal disorders.

Introduction

Chondrocytes are highly secretory cells that produce, process and secrete extracellular matrix (ECM) proteins to build the cartilage matrix (Unlu et al., 2014). Highly enriched macromolecules in cartilage include collagen type II and proteoglycans such as aggrecan, matrilin and tenascin (Chevalier et al., 1994; Hardingham and Bayliss, 1990; Mendler et al., 1989). Synthesis and processing of cartilage matrix proteins require the activities of the protein folding machinery, glycosylation enzymes and specialized secretory machinery for large ECM molecules such as collagen (Lang et al., 2006; Sarmah et al., 2010; Unlu et al., 2014). Identification of genes expressed in chondrocytes is critical to understand how these activities are regulated at the genetic level. Yet, many cartilage-expressed genes remain unrecognized because of difficulty in obtaining relatively pure populations of viable chondrocytes from poorly accessible, dense cartilage tissue. Discovering the complete cohort of genes expressed specifically in chondrocytes in an unbiased manner would inform on essential cartilage functions, namely, production and secretion of ECM components, cartilage growth and chondrocyte maturation; and may reveal novel aspects of chondrocyte biology.

Zebrafish has been extensively used to study chondrocyte differentiation and cartilage matrix formation (Barrallo-Gimeno et al., 2004; Eames et al., 2010; Eames et al., 2011b; Holmborn et al., 2012; Kessels et al., 2014; Kimmel et al., 1998; Kimmel et al., 2001; Lang et al., 2006; LeClair et al., 2009; Melville et al., 2011; Montero-Balaguer et al., 2006; Sachdev et al., 2001; Sarmah et al., 2010; Sisson et al., 2015; Yan et al., 2002). Recently, ECM protein-encoding genes in zebrafish have been cataloged *in silico* by the matrisome project (Nauroy et al., 2018). Zebrafish forward genetic screens have characterized several genes essential for cartilage homeostasis and function (Driever et al., 1996; Knapik, 2000; Neuhauss et al., 1996; Piotrowski et al., 1996; Schilling et al., 1996). These genes are highly conserved from fish to humans, and a number of them have been linked to genetic syndromes, such as chondrodysplasia, Ehlers-Danlos syndrome and osteogenesis imperfecta (OI) (Garbes et al., 2015; Mumm et al., 2001; Schalkwijk et al., 2001). Therefore, identification of transcripts enriched in zebrafish chondrocytes will likely inform on human cartilage biology and facilitate understanding of disease conditions.

Chondrocytes are exclusive producers of all ECM content in cartilage. Transcriptome profiling of human chondrocytes has been done previously utilizing cultured primary chondrocytes and patient samples by microarray approaches (Bowen et al., 2014; Funari et al., 2007; Schibler et al., 2009; Wu et al., 2013). Recently, human chondrocyte transcriptome analysis of RNA-seq experiments was reported from cadaver biopsies, fetal samples and post-operative tissue (Lewallen et al., 2016; Li et al., 2017; Soul et al., 2018) as well as chondrocytes differentiated from mesenchymal stem cells (MSC) (Jaager et al., 2012). However, obtaining the chondrocyte-specific transcriptome representing live chondrocytes using next-generation sequencing presents a formidable challenges including difficulty in dissociating live cells from dense ECM, recovering sufficient number of cells, and lack of optimized cell sorting protocols, and, thus, live chondrocyte transcriptome is largely unknown. Animal models, such as zebrafish, are highly amenable to genetic manipulation and provide a powerful discovery tool to advance cartilage biology field. Although zebrafish have been used in gene discovery (Driever et al., 1994; Kettleborough et al., 2013; Mitchell et al., 2013), high throughput transcriptional analyses of chondrocytes have, until recently, been hindered by the limited availability of transgenic reporter lines and robust single-cell isolation techniques for fluorescence-associated cell sorting (FACS).

Here, we present a FACS-based method to isolate zebrafish chondrocytes from larval craniofacial cartilage and to extract high-quality RNA from FACS-sorted chondrocytes suitable for next-generation RNA-seq. Comparing the chondrocyte transcriptome to that of concurrently isolated non-chondrocytic cells of the four-day old zebrafish head (non-chondrocyte transcriptome containing, brain, eyes and other collagen II-negative cell types), we identified 4,033 cartilage-enriched genes that exhibit more than 2-fold higher expression levels in chondrocytes. These genes clustered mainly in skeletal system development, cartilage development and extracellular matrix/structure organization pathways. Novel genes in cartilage-enriched dataset can provide entry points to study pathways of chondrocyte function and maturation. Orthology analysis between zebrafish and human gene expression revealed enrichment of skeletal disorder-related genes and confirmed high conservation of the cartilage transcriptome. Our dataset presents a rich discovery platform for novel factors regulating cartilage development and function, as well as a set of candidate genes, mutations in which may cause human skeletal disorders.

Methods

Zebrafish maintenance and breeding

The transgenic Tg[Col2a1a:caax-EGFP] zebrafish line used in this study was generated and shared by the laboratory of Dr. Jacek Topczewski (Dale and Topczewski, 2011). Briefly, the *collagen 2a1a* gene promoter element drives expression of plasma membrane tethered EGFP (by the aid of CAAX motif (Hancock et al., 1990)) as a reporter in craniofacial cartilage, ear, notochord, floor plate, hypochord and fins. The line was maintained and raised at 28.5°C under standard laboratory conditions following the guidelines established by the IACUC at Vanderbilt University Medical Center. Embryos were collected upon natural mating and grown until 4 days post-fertilization (dpf) (Kimmel et al., 1995).

Using the same *col2a1a* promoter element, we have engineered a parallel line that drives expression of the histone H2a-mCherry fusion protein in the same cell types, including craniofacial chondrocytes, thereby marking their nuclei (Luderman et al., 2017).

Tissue preparation and dissociation

Embryos at the 4 dpf stage were screened for EGFP fluorescence using a GFP3 filter equipped Leica MZ16F stereomicroscope. Approximately 100 transgenic embryos per experiment were anesthetized in Tricaine (Sigma). Head tissues were dissected from the end of the otic vesicle to exclude EGFP+ notochord, floor plate and hypochord cells from the analysis and kept in ice-cold Ringer's solution (16 mM NaCl, 2.9 mM KCl, 5mM HEPES, 1 mM EDTA) until dissociation. Dissected heads were washed three times with Dulbecco's PBS without magnesium and calcium (GIBCO). Then, tissues were treated with 3 µg/µl Collagenase IA (Sigma), 2.5 u/µl Dispase (BD Biosciences) and 50,000 u of DNase I diluted in DMEM (GIBCO-11995) at 37°C for 45 minutes. To enhance dissociation, the tissue-protease mix was pipetted up and down every 10 minutes.

Fluorescence Microscopy

Dissociated cells were wet-mounted on a microscope slide and imaged with Zeiss AxioImager.Z1 under a Cy2/GFP filter for caax-eGFP and under a Cy3.5/mCherry filter for H2a-mCherry, in addition to differential interference contrast (DIC) settings.

Electron Microscopy

4 dpf AB zebrafish embryos (AB is a standard laboratory wild-type strain) were fixed in 2.5% glutaraldehyde in 0.1 M sodium cacodylate, for 1 hour at room temperature, then at 4 ºC overnight. Fixative was removed and samples were rinsed with 0.1 M sodium cacodylate at room temperature. Embryos were post-fixed in 1% osmium tetroxide in 0.1 M sodium cacodylate for 1 hour. After an additional rinse, embryos were dehydrated in ethanol, then propylene oxide, infiltrated with resin stepwise, and consequently embedded in resin for 48 hours at 60°C. 70-nm sections were collected on a Leica Ultracut Microtome and imaged on a Phillips CM-12 Transmission Electron Microscope housed in the Vanderbilt Cell Imaging Shared Resource.

Fluorescence-associated cell sorting

Samples were prepared for FACS as previously described by Khuansuwan and Gamse (Khuansuwan and Gamse, 2014). Dissociated tissue mix was diluted 1:2 with DPBS with 30,000 u of DNase I and filtered through a 40 µm cell strainer placed in a 50-ml conical tube in ice. Filtered tissue was centrifuged at 1,000 rpm for 10 minutes at 4°C to collect cells. Supernatant was removed and pelleted cells were resuspended in DPBS containing 15,000 u of DNase I. Propidium Iodide was added in order to label and exclude dead cells during sorting. Non-transgenic, AB embryos were processed likewise as compensation controls for FACS. EGFP+ and EGFP- cell populations from the Tg[Col2a1a:caax-EGFP] line were sorted into separate tubes containing TRIzol-LS reagent (ThermoFisher Scientific).

RNA isolation

Total RNA from sorted cells was isolated using TRIzol-LS (ThermoFisher Scientific) as described by manufacturer. Quality and integrity of RNA were analyzed by RNA pico chip (Agilent Technologies). RNA quantity was measured with Qubit fluorometer (Life Technologies). RNA integrity (RIN) values between 8.0-10.0 were considered good quality for proceeding with library preparation for next generation RNA-sequencing.

cDNA library preparation and next generation sequencing

Two independent RNA samples for EGFP+ and EGFP– obtained from separate clutches of 100 zebrafish embryos each were made to increase statistical power and note potential discrepancies. cDNA libraries were prepared and processed for RNA-seq by Vanderbilt Technologies for Advanced Genomics (VANTAGE) core facility using the Illumina TruSeq mRNA library prep kit. High-throughput next-generation sequencing was performed using Illumina HiSeq 2500 technology. For each EGFP+ or EGFP- group, 30 million 50-bp single end reads were collected.

Sequence alignment and differential expression analysis methods

The reads were pre-processed by removing adapters and ambiguous leading and tailing nucleotides. Then, overall quality was examined using FastQC (S, 2010). Next, reads were aligned to the zebrafish transcriptome (based on GRCz10 release 80) and genome (Genome assembly: GRCz10) using Tophat2 (Kim et al., 2013) with default parameters. Tophat2 aligns reads in three tiers: first, it maps all the reads to the known transcriptome; second, it maps the unmapped reads to the genome; and, third, it maps multi-exon spanning reads to spliced regions. This workflow ensures that as many reads as possible are mapped. Mapping quality was assessed by QPLOT (Li et al., 2013) and potential outliers with obviously abnormal quality were excluded from following analyses. Differentially expressed genes were called by DESeq2 (Love et al., 2014) in Bioconductor (Huber et al., 2015). Data are deposited to a raw data public repository.

Protein association network analysis of human orthologous genes

List of human orthologous genes included in the specific GO category was uploaded to STRING protein association database v10.5 (string-db.org) (Szklarczyk et al., 2015) and ‘Homo sapiens’ was selected as ‘Organism’. Under default settings, STRING protein association network was generated. Using STRING app installed into Cytoscape (Kohl et al., 2011) (version 3.6.0), node coloring, label size and network map were adjusted for better visibility.

Gene Ontology & Associated Disease Enrichment Analysis

Gene Ontology enrichment analysis was performed with the WebGestalt tool (Wang et al., 2013; Zhang et al., 2005). Enrichment analysis was performed for biological processes using the hypergeometric test with 'Top10 significance level' and 'minimum 2 genes per category' criteria. Using the same default settings, WebGestalt tool was used to run disease enrichment analysis on human orthologous genes. Associated disease groups were based on the PharmGKB database.

Reverse Transcription PCR

RNA extracted from sorted cells was treated with DNase I (Ambion). 10 ng DNase I treated RNA was used as a template in a one-step RT-PCR reaction using Superscript II RT/Platinum Taq DNA polymerase (ThermoFisher Scientific). Gene-specific primers used were as follows:

β-actin forward primer 5'-GACTCAGGATGCGGAAACTG-3',

β-actin reverse primer 5'-AAGTCCTGCAAGATCTTCAC-3';

creb3l2 forward primer 5'-CACAGAACCACCACCATGAG-3',

creb3l2 reverse primer 5'-ACAGGAGAGTCGCAGGAAAA-3';

col2a1a forward primer 5'-ACCTGAAGAAGGCCATTCTG-3',

col2a1a reverse primer 5'-TTACAAGAAGCAGACTGCGC-3'.

Cryosectioning and Immunohistochemistry

Zebrafish embryos were fixed in 4% paraformaldehyde at the 4 dpf stage. Fixative was removed with PBS rinses, then replaced with 30% sucrose/PBS. Embryos were incubated at 4ºC overnight prior to embedding in Cryomatrix (Thermo Scientific), then moved to -20ºC for freezing. Frozen sections were collected with a Leica #CM1900 cryotome. Collagen type-II (Rockland, #600-401-104) antibody and Alexa Fluor-488 conjugated wheat germ agglutinin (WGA, Life Technologies) were used for histological analysis. Collagen-II recognizing antibodies were detected with Alexa Fluor-555 labeled secondary antibody (ThermoFisher Scientific). Mounted slides were imaged with Zeiss AxioImager Z1 equipped with an Apotome and EC Plan-Neofluar 40x/0.75 & EC Plan Neofluar 100x/1.30 Oil objectives.

Analysis of temporal expression profiles

Temporal expression profiles of COPII-related zebrafish genes across developmental stages were analyzed using reference dataset generated by White et al. (White et al., 2017). Pre-computed count profiles (TPM values) were browsed and obtained through Expression Atlas (http://www.ebi.ac.uk/gxa/experiments/E-ERAD-475). Expression data were analyzed through Broad Institute's Morpheus web-tool (https://software.broadinstitute.org/morpheus/) to generate heatmap profiles under default settings. Hierarchical clustering was conducted using the following settings built into Morpheus: one minus Pearson correlation metric and average linkage method.

Comparative analysis of human orthologous genes to zebrafish chondrocyte-enriched genes and human fetal cartilage-selective genes

Human orthologous genes were obtained as described above (fold change>5, p<0.05). Human fetal cartilage-selective gene dataset was obtained from Li et al. study (Li et al., 2017). To identify a comparable dataset to human orthologous gene set, cartilage expression level of each transcript was averaged and compared to its average expression level across non-cartilage tissues. Genes with one or more transcripts to reach >5-fold expression in cartilage were filtered as fetal cartilage-selective dataset. This dataset contained 1205 genes. Comparative analysis between human orthologous genes of zebrafish chondrocyte-enriched genes and human fetal cartilage-selective genes was carried out to identify common and exclusive genes.

Statistical Analyses

Correlation analysis was performed to confirm reproducibility of replicate gene expression data sets from GFP+ and GFP- cell populations. Pearson's correlation coefficients (R^2 values) were calculated as a measure of correlation; 1.0 is considered perfect linear correlation. To test significance of differential gene expression levels, the Wald test was used in the DESeq2 tool (Love et al., 2014). Calculated p-values were corrected by the FDR (false discovery rate) method, which is also implemented by DESeq2. The confidence interval was taken as 95% (i.e. p<0.05). Gene ontology analyses and associated disease groups were statistically analyzed for significant enrichment of gene sets in the corresponding category, using Hypergeometric test, which was built into the WebGestalt (Wang et al., 2013) tool. Significance was determined as p<0.05.

Results

Isolation of craniofacial chondrocytes by FACS sorting

To isolate homogenous population of mature chondrocytes, I used 4-dpf (day post fertilization) stage zebrafish larvae. At this stage, most craniofacial chondrocytes passed stacking stage and begun producing collagen-rich ECM (Fig. 2.1A-C). I used a transgenic zebrafish line Tg[Col2a1a:caax-EGFP] (Dale and Topczewski, 2011) expressing EGFP in chondrocytes within the craniofacial skeleton, otic vesicle and pectoral fins in the head. To enrich for craniofacial chondrocytes, I dissected heads in order to eliminate trunk EGFP+ cells in the notochord, hypochord and floor plate. For each experiment we use a pool of over 100 zebrafish larvae.

The main challenge in isolating live chondrocytes from mature cartilage is dissociating cells from the dense ECM while preserving their cell membrane integrity. To achieve this goal, I varied the enzyme concentrations in a protease cocktail containing collagenase, dispase as well as DNase I, and assessed the quality and efficiency of single-cell dissociation by wet-mounted cell suspension and with wide-field epifluorescence microscopy (Fig. 2.1D). Next, I crossed the Tg[Col2a1a:caax-EGFP] line with a nuclear mCherry reporter (H2a-mCherry) line (Luderman et al., 2017), to obtain double labeled chondrocytes. Under epifluorescence, I assessed the integrity of the cell membrane (EGFP) and nucleus (mCherry) using a wide range of protease mixes (Fig. 2.1E). The optimized protease cocktail (see Materials and Methods) yielded the highest number of intact, single cells as detected by microscopy.

Using the optimized protease cocktail on dissected heads, I successfully dissociated live chondrocytes from Tg[Col2a1a:caax-EGFP] intact cartilage tissue (Fig. 2.1F). FACS analysis showed that live EGFP+ chondrocytes constituted 2.6% of the entire population of dissociated cells. I also isolated live EGFP- non-chondrocytic cells for comparative analysis. Stringent gating parameters led us to collect 7.2% of the entire population in the EGFP- cell category mainly containing brain, eye and epithelial cells (Fig. 2.1G).

Total RNA extracted from FACS-isolated EGFP+ cells contain transcripts enriched in chondrocyte markers

To assess the integrity of RNA extracted from FACS isolated GFP+ and GFP- cells, I analyzed an electrophoretic profile that showed strong 18S and 28S rRNA peaks (Fig. 2.2A-B). I measured RNA integrity (RIN) values to assess RNA quality using RNA pico chip (Agilent Technologies). RIN values were between 8.3 and 9.7; 28S/18S rRNA ratios ranged between 1.8 and 2.1 (Fig. 2.2C). I confirmed enrichment of chondrocytes in the GFP+ population by testing transcript expression levels by RT-PCR of chondrocyte-specific markers, such as *col2a1a* and *creb3L2*, which showed higher levels, while *β-actin* levels were comparable between the GFP+ and GFP- populations (Fig. 2.2D). These RNA samples were used to construct two independent replicates of GFP+ and GFP- cDNA libraries for Next Gen Illumina HiSeq 2500 RNA-sequencing as described in the 'Materials & Methods' section.

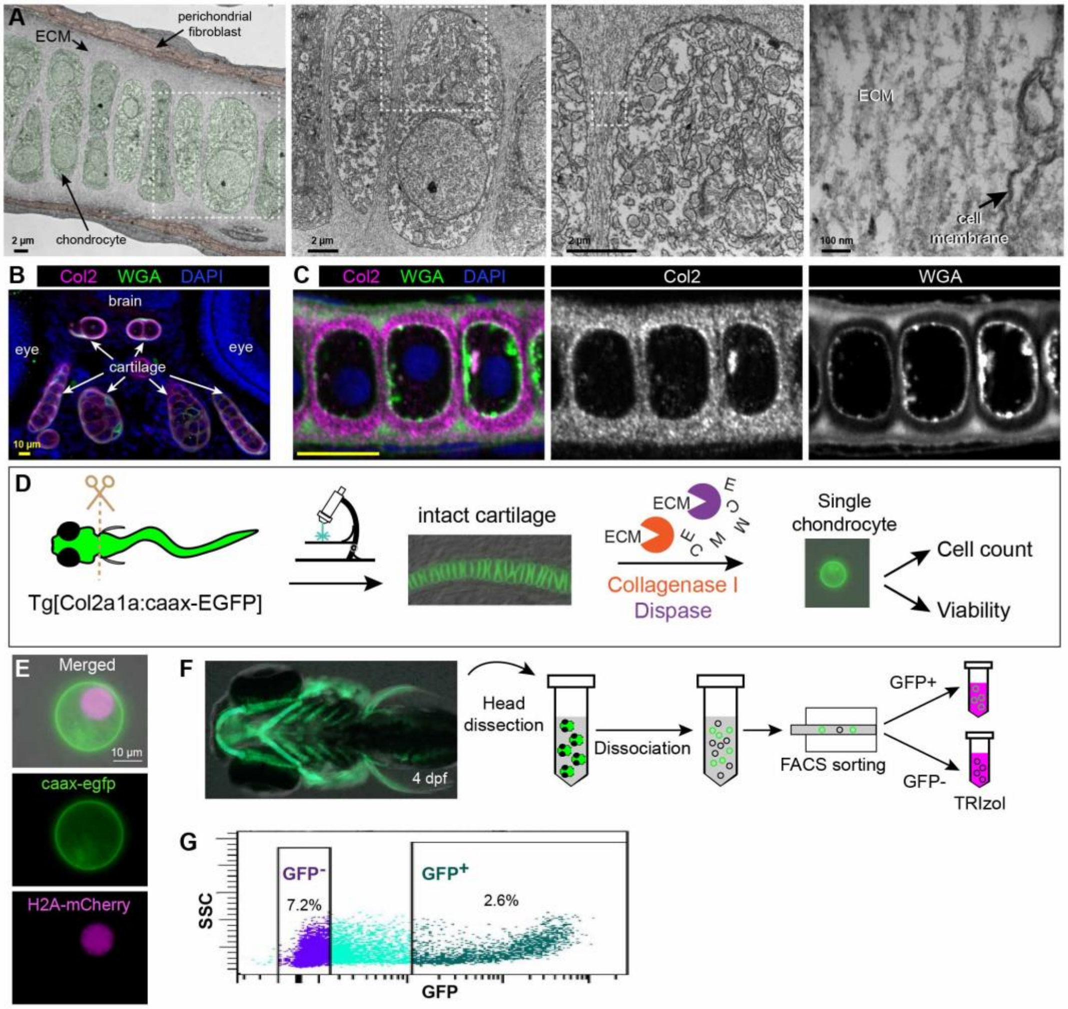


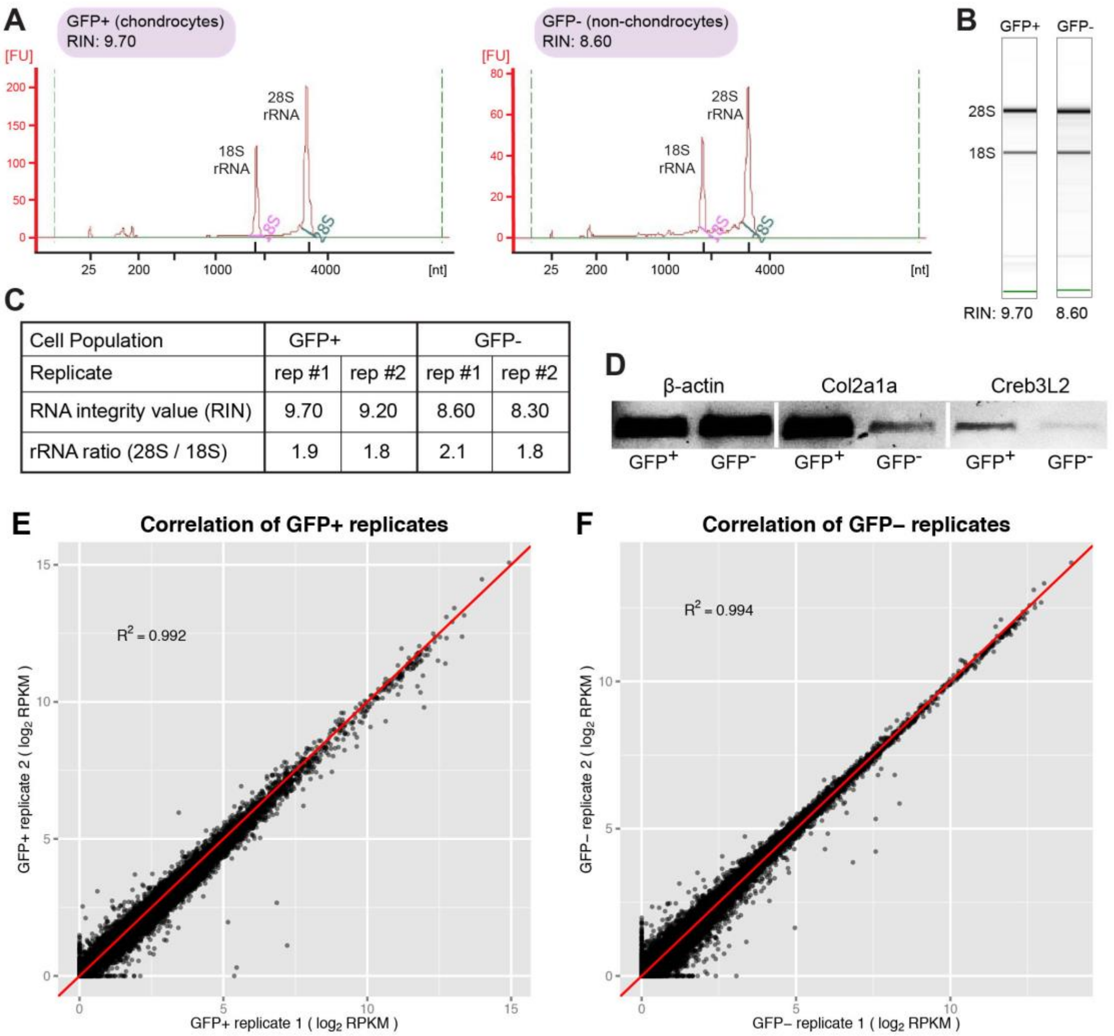
Figure 2.1. Experimental strategy for cartilage dissociation

A. Electron micrographs of 4 dpf zebrafish craniofacial cartilage. Magnified images of boxed regions are displayed on the right-side panels. **B.** Immunohistochemistry image of 4 dpf zebrafish head in coronal section and **C.** its higher objective image. Col2: collagen type-II, WGA: Wheat germ agglutinin. DAPI is used as nuclear stain. **D.** Experimental design to dissociate live chondrocytes from transgenic Tg[col2a1a:caax-EGFP] zebrafish embryos at 4 dpf stage. caax-EGFP transgene marks cell membrane of craniofacial chondrocytes within the head. Dashed line shows where embryo heads were dissected for dissociation. **E.** Representative image of a dissociated single chondrocyte from Tg[col2a1a:caax-EGFP] line (cell membrane in green) with a nuclear mCherry marker (H2a-mCherry, in magenta). **F.** Experimental outline for isolation of GFP+ live zebrafish chondrocytes and GFP- non-chondrocytic cells with FACS. **G.** Gating parameters in FACS to sort GFP+ and GFP- cell populations. Percentage of cells obtained in each gate is represented on the graph. SSC: Side Scatter.

Quality control analysis of RNA-seq reads confirms efficient mapping and high correlation of replicates

We collected more than 30 million reads per sample and included independent replicates for both GFP+ and GFP- populations. Using the TopHat2 algorithm (Kim et al., 2013), we obtained mapping rates varying between 85% to 96% among all samples (Fig. 2.2E). Next, we assessed the quality of sequence run performance with the QPLOT tool (Li et al., 2013) (Fig. 2.2F). We retained up to 20 possible alignments for each read during mapping to the transcriptome and genome using TopHat2. Mapping statistics for each sample with the QPLOT tool showed that more than 30 million reads were mapped to zebrafish transcriptome at high efficiency with TargetMapping values, specifically 89-90%. The percentage of ZeroMapQual and MapQual<10 values were relatively small, which indicated that most of the reads have high mapping quality. All samples passed quality control criteria with 100% Mapping Quality (MQpass) and 0% quality control fail rate (QCFailRate). Base compositions of four nucleotide types (A, C, G, T) showed nearly homogenous distribution among samples varying between 21% and 28%. Both TopHat2 and QPLOT algorithms confirmed high quality sequencing and reads mapping, meeting the criteria required for constructing reliable differential gene expression profiles from RNA-seq data.

Additional criterion for assembling a differential gene expression dataset is correlated replicates (Marioni et al., 2008; McCarthy et al., 2012). I started with independent replicates for both GFP+ (chondrocyte) and GFP- (non-chondrocyte) cell populations.



Cell Population	GFP+		GFP-	
Replicate	rep #1	rep #2	rep #1	rep #2
RNA integrity value (RIN)	9.70	9.20	8.60	8.30
rRNA ratio (28S / 18S)	1.9	1.8	2.1	1.8

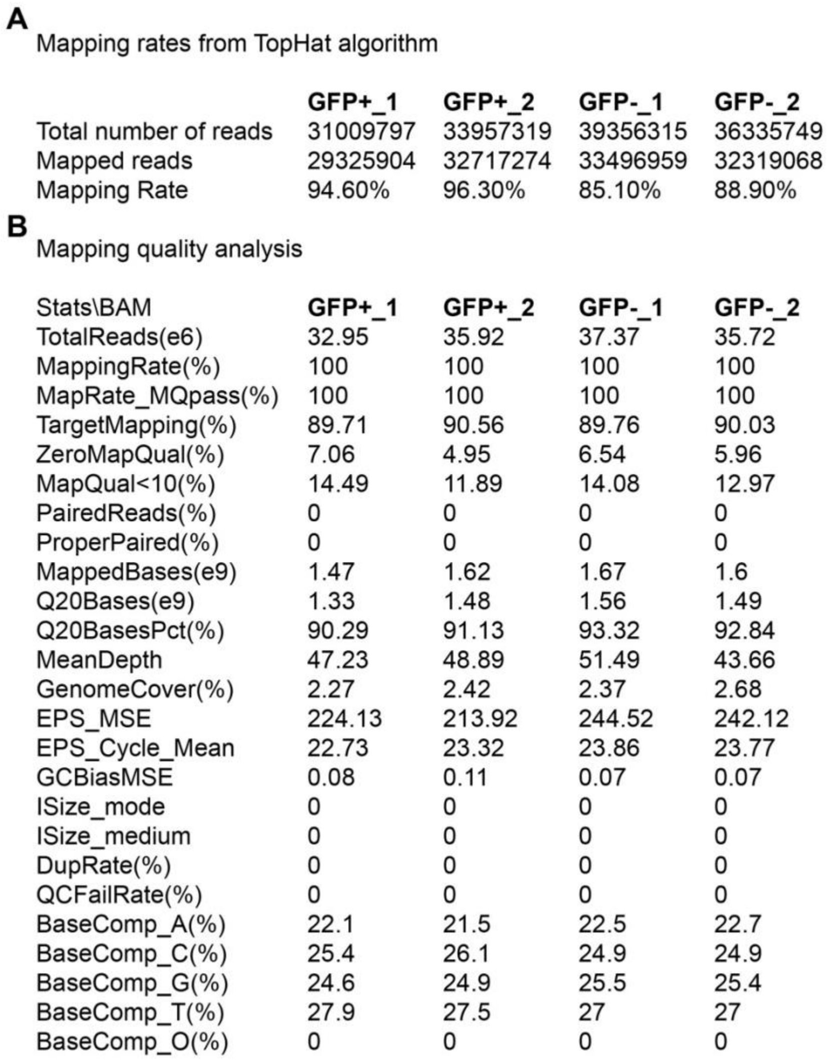
Figure 2.2. Quality control of RNA samples and cDNA libraries

A. Electrophoretic profile overlays that show 18S and 28S rRNA peaks in GFP+ and GFP- cell populations. **B.** Gels displaying abundant 18S and 28S rRNAs. RNA integrity (RIN) values are indicated below. **C.** RIN and rRNA ratio values of each replicate sample used in this study. **D.** RT-PCR showing enriched expression of chondrocyte markers *col2a1a* and *creb3l2* in GFP+ cell population, and the beta-actin control. **E-F.** Scatter plots displaying gene expression levels ($\log_2$RPKM values) from replicate samples of GFP+ population (E) and GFP- population (F). Each gene is represented by an individual dot. Highly expressed genes are closer to upper right and lowly expressed genes are toward lower left corner. Red line marks perfect linear correlation. R^2 values shown on graphs indicate high level of correlation (R^2=1, perfect linear correlation).

To test whether transcriptomes obtained from replicates are correlated among groups, we plotted $\log_2$RPKM values (Reads Per Kilobase of transcript per Million mapped reads) of each gene from each sample on a scatter plot and calculated R^2 values as a measure of correlation. Both GFP+ and GFP- transcriptomes yielded high level of linear correlation with R^2 values of 0.992 and 0.994, respectively (Fig. 2.3A, B). Taken together, our cell isolation protocol, RNA purification, library construction, Next-Gen Sequencing protocol and sequence read analysis allowed us to generate a high-quality dataset for mature chondrocyte transcriptome and non-chondrocyte controls.

RNA-seq analysis revealed known and novel chondrocyte-enriched genes with roles in ECM organization and cartilage development

Bioinformatics analyses identified differentially expressed genes between GFP+ and GFP- cell populations (Fig. 2.4A). Among 9,367 differentially expressed genes, 4,033 were specifically enriched in GFP+ chondrocytes (>2-fold, false discovery rate [FDR]-corrected $p<0.05$) (Fig. 2.4B), and the remaining genes were either enriched in GFP- cells or not significantly enriched in either of the populations. To analyze the nature of differentially expressed genes, we performed a hypergeometric test for associated gene ontology (GO) terms (Rivals et al., 2007), and found over-representation of genes with roles in cartilage and skeletal system development and ECM organization (Fig 2.4C). The ‘cartilage development’ gene category contained 32 previously reported genes such as *sox9a*, *creb3L2*, *fgfr3* and *aggrecan*. The ‘ECM organization’ category included structural ECM genes such as collagens, matrilins and aggrecan.



A

Mapping rates from TopHat algorithm

	GFP+_1	GFP+_2	GFP-_1	GFP-_2
Total number of reads	31009797	33957319	39356315	36335749
Mapped reads	29325904	32717274	33496959	32319068
Mapping Rate	94.60%	96.30%	85.10%	88.90%

B

Mapping quality analysis

Stats\BAM	GFP+_1	GFP+_2	GFP-_1	GFP-_2
TotalReads(e6)	32.95	35.92	37.37	35.72
MappingRate(%)	100	100	100	100
MapRate_MQpass(%)	100	100	100	100
TargetMapping(%)	89.71	90.56	89.76	90.03
ZeroMapQual(%)	7.06	4.95	6.54	5.96
MapQual<10(%)	14.49	11.89	14.08	12.97
PairedReads(%)	0	0	0	0
ProperPaired(%)	0	0	0	0
MappedBases(e9)	1.47	1.62	1.67	1.6
Q20Bases(e9)	1.33	1.48	1.56	1.49
Q20BasesPct(%)	90.29	91.13	93.32	92.84
MeanDepth	47.23	48.89	51.49	43.66
GenomeCover(%)	2.27	2.42	2.37	2.68
EPS_MSE	224.13	213.92	244.52	242.12
EPS_Cycle_Mean	22.73	23.32	23.86	23.77
GCBiasMSE	0.08	0.11	0.07	0.07
ISize_mode	0	0	0	0
ISize_medium	0	0	0	0
DupRate(%)	0	0	0	0
QCFailRate(%)	0	0	0	0
BaseComp_A(%)	22.1	21.5	22.5	22.7
BaseComp_C(%)	25.4	26.1	24.9	24.9
BaseComp_G(%)	24.6	24.9	25.5	25.4
BaseComp_T(%)	27.9	27.5	27	27
BaseComp_O(%)	0	0	0	0

Figure 2.3. Bioinformatics analysis for the quality of RNA-seq reads
A. Table showing mapping rates of RNA-seq reads from GFP+ and GFP- cell populations as calculated by TopHat2 algorithm. **B.** Table containing mapping quality control metrics of GFP+ and GFP- samples as assessed by QPLOT tool.

In addition, proteoglycan processing enzymes such as chondroitin sulfate synthase 1 (*chsy1*), xylosyltransferase I (*xylt1*) and *fam20b* fell into this category (Eames et al., 2011b; Kitagawa et al., 2001; Koike et al., 2009; Schreml et al., 2014). Gene ontology underscored that the EGFP+ cell isolate is highly enriched in chondrocytes. In particular, the significant overrepresentation of cartilage development genes in the GFP+ dataset further validated our isolation protocol. In contrast, non-chondrocyte-enriched GO terms excluded skeletal system processes, rather included more generic categories like 'biological regulation' and 'anatomical structure development' (Fig. 2.4D) since this population is likely comprised of several non-chondrocytic cell types such as neurons, glial cells and myocytes among many others.

ECM components are among the top chondrocyte-enriched genes

Unbiased, whole transcriptome approach is a powerful tool to uncover components of the cartilage matrix. Among the top 50 genes highly enriched in zebrafish chondrocytes, 19 encode structural ECM proteins, most of which are shared with the human cartilage matrix (Table 2.1A). For instance, human homologs of collagen type II alpha 1a (*col2a1a*), matrilin 1 (*matn1*), aggrecan a (*acana*) and aggrecan b (*acanb*) have been extensively used as hallmarks of cartilage matrix (Martin et al., 2001; Ohno et al., 2003; Schibler et al., 2009). In addition, we detected enrichment of other cartilage matrix protein transcripts also abundant in human cartilage such as epiphycan (*epyc*), tenascin X (*tnxb*), collagen types -IX (*col9a1a*, *col9a1b*, *col9a2*, *col9a3*) and -XI (*col11a1a*, *col11a2*) (Chevalier et al., 1994; Knudson and Knudson, 2001; Vornehm et al., 1996).

Besides the conventional ECM components, we detected previously unreported protein transcripts highly enriched in chondrocyte transcriptome, such as *Fibulin-7-like* (CABZ01113374.1). *In silico* domain analyses suggest that this protein contains calcium-binding EGF-like domains and SUSHI repeats (UniProtKB, A0A0G2KY26). It is similar to fibulin-7, an extracellular matrix protein (62% similar, 46% identical). Given its domain structure and homology to fibulin-7, it is likely to be an extracellular matrix protein. The RPKM value for this gene is well below 1 in GFP- cells indicating very low to no expression, while it is expressed with >500 RPKM value (~2000-fold enrichment) in GFP+ chondrocyte population (Table 2.1).

This comprehensive dataset provides expression level information of all chondrocyte-enriched genes and places them in context to age-matched control cells in non-collagen II producing cells of the larval zebrafish head. This dataset lists genes previously unknown to be associated with chondrocyte function and confirms expression of known cartilage genes.

Expression levels of COPII components exhibit paralog-specific enrichment in chondrocytes

ECM macromolecules are synthesized in the endoplasmic reticulum (ER) and exit the site through the vesicular-tubular carriers *en route* to the Golgi complex, where they undergo further modification. ER-to-Golgi trafficking is a critical step in secretion of ECM proteins, and is mediated by coat protein II complex (COPII). Specific roles for COPII components in ECM formation and cartilage development have been previously discussed in the literature (Unlu et al., 2014; Vacaru et al., 2014). Inner COPII coat components,

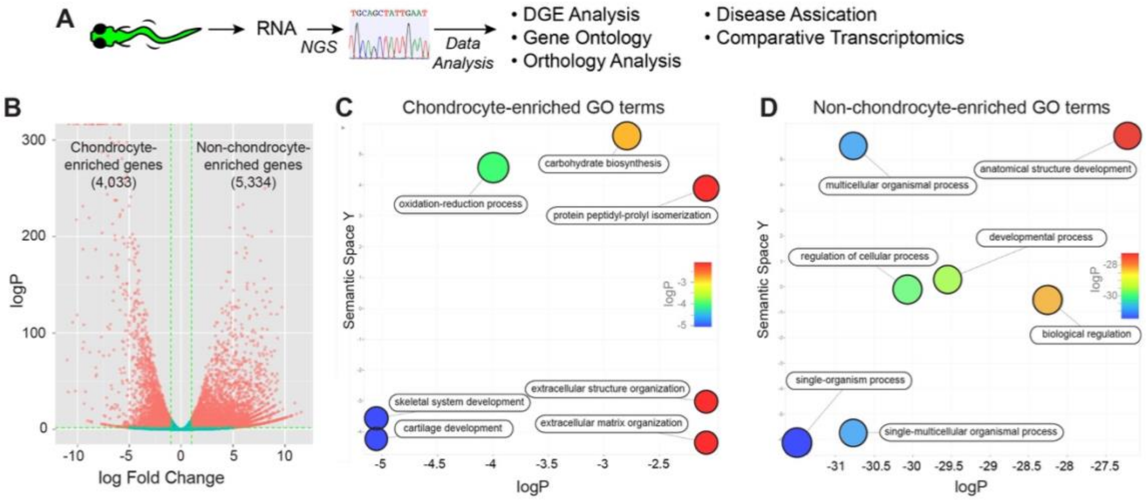


Figure 2.4. Differential Gene Expression Analyses of Chondrocyte and Non-chondrocyte-enriched Gene Sets

A. Experimental outline describing RNA-seq and differential gene expression (DGE) analyses performed. **B.** Volcano plot displaying differentially expressed genes (red dots) between GFP+ and GFP- cells (fold-difference > 2.0 and p<0.05). Genes with non-significant expression patterns are shown in green at the mid-bottom of the graph **C.** GO categories overrepresented in chondrocyte-enriched and **D.** non-chondrocyte enriched gene lists as detected by hypergeometric test of GO terms (p<0.05). Significantly represented GO terms (biological processes) are represented with scatterplot option of REViGO choosing medium resulting (0.7). x-axis indicates log(p-values), i.e. more significantly enriched GO categories are closer to bottom left. Categories are also color coded according to significance levels (see color scale). Similar GO terms were grouped closer in semantic space Y.

Table 2.1. Most highly enriched structural ECM genes in zebrafish chondrocytes

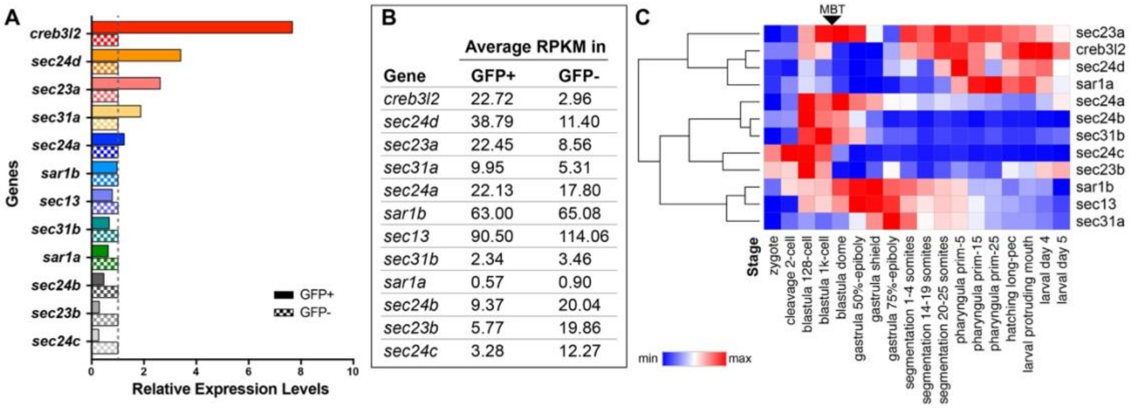
Gene Symbol	Gene Name	Fold Difference	RPKM in GFP+	RPKM in GFP-
CABZ01113374.1	fibulin-7-like	2045.5	587.9	0.4
tnxb	tenascin XBa	1918.9	84.2	0.1
epyc	epiphycan	1841.4	2167.4	2.1
matn1	matrilin 1	1190.2	32910.8	39.5
chad	chondroadherin	1118.0	118.8	0.2
acanb	aggrecan b	728.1	241.3	0.6
acana	aggrecan a	630.8	380.3	1.1
col9a3	collagen, type IX, alpha 3	594.5	5380.3	16.8
col2a1a	collagen, type II, alpha 1a	454.5	19463.4	74.2
CLEC3A	C-type lectin domain family 3, member A	391.9	535.6	0.7
col9a1a	collagen, type IX, alpha 1a	387.4	7820.4	37.4
col9a1b	collagen, type IX, alpha 1b	386.5	66.4	0.3
lect1	Chondromodulin I (leukocyte cell derived chemotaxin 1)	383.4	3341.8	16.1
si:dkey-6n6.1	Similar to ECM protein 2 (Matrix glycoprotein SC1/ECM2)	369.0	607.5	3.1
col9a2	*procollagen, type IX, alpha 2*	353.2	9643.1	50.7
col11a2	collagen, type XI, alpha 2	263.9	2357.1	16.6
ucmab	upper zone of growth plate and cartilage matrix associated b	240.3	9846.1	75.8
col11a1a	collagen, type XI, alpha 1a	226.6	3885.8	31.9
otomp	otolith matrix protein	207.1	19.8	0.2

sec23a and *sec24d*, act in type-II collagen secretion and chondrocyte maturation in zebrafish (Lang et al., 2006; Sarmah et al., 2010) and humans (Boyadjiev et al., 2006; Fromme et al., 2007). (Boyadjiev et al., 2006; Boyadjiev et al., 2011; Garbes et al., 2015; Moosa et al., 2015). Consistent with these findings, RNA-seq analysis revealed approximately 2.6-fold enrichment of *sec23a* in the GFP+ chondrocyte population (Fig. 2.5A, B). On the other hand, expression of its close paralog, *sec23b,* is about 3.4-fold enriched in non-chondrocytic, GFP- cells. Human mutations in *SEC23B* have been reported in congenital dyserythropoietic anemia type II patients (Bianchi et al., 2009; Iolascon et al., 2010). This condition involves gallstone formation, but cartilage deficits were not described.

Sec24 proteins function in cargo selection through direct or adaptor-mediated interaction with COPII cargos such as cartilage matrix components (Wendeler et al., 2007). Vertebrates, including zebrafish, harbor four Sec24 paralogs (Tang et al., 1999). Previous reports have suggested that during evolution Sec24 paralogs diversified and became selective for tissue-specific cargos (e.g. collagen-type II or serotonin transporter [SERT]), although, they can exhibit redundant functions during development (Sarmah et al., 2010). The zebrafish and medaka models, as well as human patient studies, showed that SEC24D functions in ECM cargo traffic (e.g. collagen-II and matrilin) and is essential for normal skeletal development (Garbes et al., 2015; Ohisa et al., 2010; Sarmah et al., 2010). Given its ECM cargo-specific role, *sec24d* was the most enriched amongst *sec24* paralogs, exhibiting 3.5-fold enrichment in GFP+ chondrocytes over GFP- non-chondrocytic population (Fig. 2.5A, B). The expression of *sec24b* and *sec24c*, was enriched in GFP- cell populations mostly of neural origins in the brain and the eye, again consistent with SEC24C's role in serotonin transporter (SERT) traffic (Sucic et al., 2011).

Others and we have previously reported that Creb3L2, an ER-synthesized transcription factor of CREB/ATF family, regulates expression of select COPII components in chondrocytes to activate collagen secretion during chondrocyte maturation in zebrafish and mice(Melville et al., 2011; Saito et al., 2009a). Creb3L2 is highly enriched in chondrocytes (7.7-fold) just like its targets, (Melville et al., 2011; Saito et al., 2009a; Tomoishi et al., 2017) *sec23a* and *sec24d* (Fig. 2.5A, B). These findings offer support for transcriptional regulation of COPII-mediated ER-to-Golgi transport by CREB/ATF family transcription factors as a regulatory mechanism for ECM secretion. Our data indicate that besides ECM components, the machinery required for their synthesis, processing and secretion is also enriched in the EGFP+ cells.

I have shown that COPII-related genes display spatial expression diversity between chondrocytes and non-chondrocyte cell types, which can be attributed to their distinct functional roles as discussed above. To assess whether temporal expression profiles of COPII components implicated in ECM trafficking also correlate with cartilage morphogenesis, we utilized the reference dataset from White et al. study (White et al., 2017). Expectedly, COPII inner coat components *sec23a* and *sec24d*, as well as *creb3L2* transcription factor that controls their expression, are expressed at high levels at post-gastrulation stages, which is when tissues with high ECM demands (e.g. cartilage) start to form (Fig. 2.5C). Hierarchical clustering supported genetic and functional interaction between *sec23a*, *sec24d* and *creb3L2* since they group together as a cluster with regards to their temporal expression patterns (Fig. 2.5C). In contrast, other COPII coat component



	Average RPKM in	
Gene	GFP+	GFP-
creb3l2	22.72	2.96
sec24d	38.79	11.40
sec23a	22.45	8.56
sec31a	9.95	5.31
sec24a	22.13	17.80
sar1b	63.00	65.08
sec13	90.50	114.06
sec31b	2.34	3.46
sar1a	0.57	0.90
sec24b	9.37	20.04
sec23b	5.77	19.86
sec24c	3.28	12.27

Figure 2.5. Expression profiles of COPII-related genes
A. COPII-vesicle component paralogs and creb3L2 transcription factor that selectively regulates their expression levels are displayed. Graph of relative expression levels of selected genes in GFP+ and GFP- cells. Expression level of each gene was normalized against its corresponding expression in GFP- population. **B.** Table displaying actual average RPKM levels of COPII-related genes in GFP+ and GFP- cells**. C.** Hierarchical clustering and heatmap for temporal expression profiles of COPII-related genes across zebrafish developmental stages, from a zygote to free-feeding larvae. MBT (Mid Blastula Transition) marks transition from maternal RNAs to zygotically produced transcripts.

paralogs, namely, *sec24a*, *sec24b*, *sec31b*, *sec24c* and *sec23b* show opposing expression profiles with peak expression points at pre-gastrulation stages, consequently clustering together.

These data strongly reveal that COPII paralogs have diversified to assume distinct expression domains throughout developmental stages, which could be translated into functional diversity between closely-related paralogs, such as *sec24c* and *sec24d*.

Chondrocyte-enriched dataset includes ECM processing enzymes

Heparan sulfate- and chondroitin sulfate proteoglycans (PG) are abundant in the cartilage ECM, and are highly modified with sugar moieties and sulfate side chains. Several PG-processing enzymes act in post-translational modification of cartilage PGs (Luderman et al., 2017). The chondrocyte-enriched transcriptome has revealed highly expressed PG modifying enzymes within top 100 hits, such as chondroitin sulfate N-acetylgalactosaminyl transferase 1a [*csgalnact1a*; enriched 680-fold], uronyl-2-sulfotransferase [*ust*; enriched 663-fold], hyaluronan synthase 3 [*has3*; enriched 296-fold] and carbohydrate (chondroitin 6) sulfotransferase 3a [*chst3a*; enriched 119-fold]. For instance, *CHST3* mutations in humans were reported to result in '*Spondyloepiphyseal dysplasia with congenital joint dislocations'* with diagnostic features including dislocation of hips or knees at birth, joint dysplasia, short stature and progressive kyphosis (Unger et al., 2010). Hence, zebrafish chondrocyte transcriptome can be utilized to study cartilage-specific ECM processing enzymes and pathways.

Comparative transcriptome analysis of zebrafish and human cartilage genes

To gain more insight into how the zebrafish chondrocyte-enriched transcriptome correlates to human cartilage biology and disease states, we performed orthology analysis between zebrafish and human genes. To address this point, we examined genes with more than 5-fold enrichment in the GFP+ population and RPKM>1 ($p<0.05$). These filters yielded 1,449 zebrafish chondrocyte-enriched genes. We then mapped these 1,449 zebrafish genes to their human orthologs. Gene conversion was performed in three categories, *one-to-one*, *one-to-many* and *many-to-many* depending on the number of orthologs found in zebrafish and human genomes for any given gene. The final conversion of the 1,449 zebrafish genes resulted in 1,112 human orthologs that, by gene ontology analysis, fell in to 9 distinct GO biological process categories (Fig. 2.6A). Remarkably, in the independent analysis of the human and zebrafish datasets, ‘extracellular matrix organization’, ‘skeletal system development’ and ‘cartilage development’ GO categories are shared. Molecular function analysis on GO terms emphasized enrichment of ECM structural constituents, proteins with collagen and fibronectin binding capabilities as well as chondroitin sulfotransferases in the human orthologous gene dataset (Fig. 2.6B). Expectedly, most of the human orthologous genes were grouped into the extracellular matrix/ region in cellular component GO analysis (Fig 2.6C). We also identified human orthologs acting in the endoplasmic reticulum, including protein folding and modification enzymes (e.g heat shock proteins, prolyl 4-hydroxylases, EXT family glucosyltransferases etc.) that process secretory ECM proteins.

Of the 1,449 zebrafish genes, 437 genes did not match to any human orthologs. However, this unmapped gene category is mainly composed of un-annotated zebrafish genes, those discovered by *in silico* approaches and pseudogenes. Therefore, it is likely that further pairwise-alignment, sequence-based orthology approaches and functional analyses can identify novel, previously un-annotated cartilage markers currently grouped within this category.

Additionally, comparative analysis of zebrafish and human cartilage gene expression profiles showed that the zebrafish chondrocyte-enriched transcriptome highly resembles that of humans since both share overrepresentation of similar cellular processes. The human orthologous gene dataset displayed enrichment of genes associated with skeletal disorders such as bone diseases, osteochondrodysplasias and joint diseases (Fig. 2.6D), suggesting that novel genes discovered by this study as enriched can help identify unknown players in undiagnosed skeletal disorders.

To test whether the human ortholog gene dataset contained central regulatory pathways of ECM organization and skeletal development genes (the top two GO biological processes), we performed protein association network analysis using the STRING (Szklarczyk et al., 2017) database. Briefly, this analysis connects proteins with known physical or functional interactions/ associations with nodes (lines). Proteins with high regulatory functions are connected to multiple others through nodes, hence forming hubs. Therefore, network analysis helps visualize multiple pathways in one chart, providing a more global look. Network analysis of the ECM organization category placed major cartilage matrix proteins such as collagen types COL2A1, COL9A1, COL27A1 and COL11A1 into major hubs (Fig. 2.7A), confirming that human orthology analysis successfully captured central ECM constituents and their associated partners. Furthermore, network analysis on the skeletal system development category displayed key regulators of

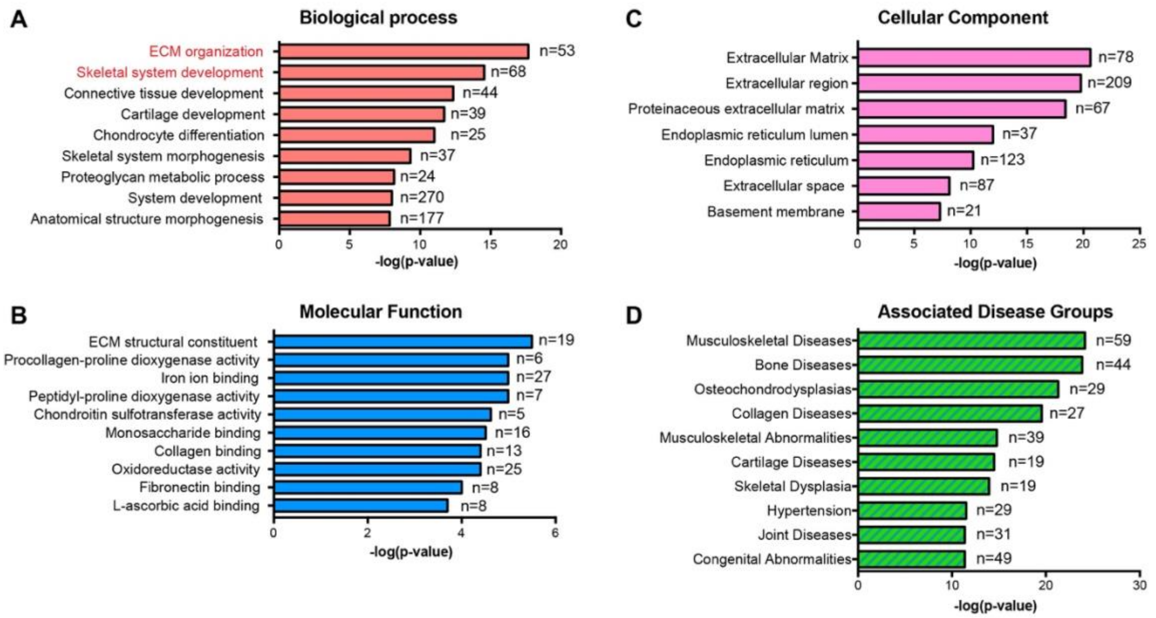


Figure 2.6. Gene set enrichment analysis for human orthologs of highly enriched chondrocyte-selective genes

Significantly represented gene ontology categories (fold change>5, RPKM>1, p<0.05) for the: **A.** biological process, **B.** molecular function, **C.** cellular component and **D.** associated disease groups; n= number of genes contained in each category is shown next to the bar.

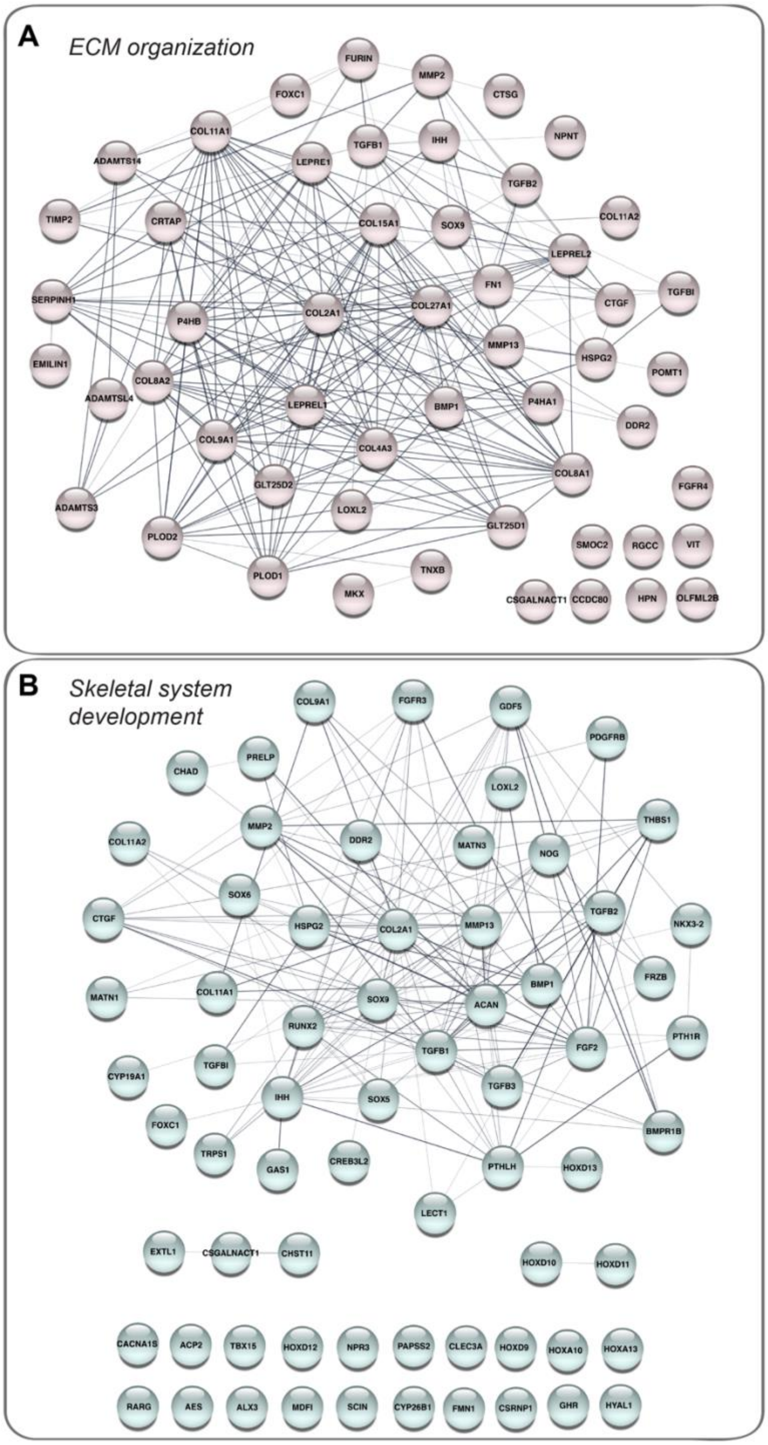


Figure 2.7. Protein association networks of human orthologous genes.
A. ECM organization genes **B.** skeletal system development genes (generated with STRING database v10.5’s protein-protein association tool). Nodes (circles) represent individual proteins; edges (lines) represent currently known associations either physical, genetic or functional. Edge confidence is represented by line thickness. Thicker lines indicate higher confidence of protein-protein biological associations. Biological processes used in association network depiction are shown in Figure 2.5A (red).

skeletal development on major hubs. Particularly, *COL2A1* (major cartilage matrix component), *SOX9* (*SRY-box 9*, transcription factor driving cartilage differentiation and activator of *COL2A1* expression (Lefebvre et al., 1997)), *RUNX2* (*runt-related transcription factor 2*, key transcription factor regulating bone differentiation (Gaur et al., 2005)) and *IHH* (*indian hedgehog*, regulator of chondrocyte proliferation and differentiation (St-Jacques et al., 1999)) formed main hubs (Fig. 2.7B), verifying validity of the orthology analysis and usability of the human orthologous gene dataset for discovering novel skeletal development genes.

Since there is a high genetic and developmental conservation between zebrafish and human cartilage morphogenesis, we reasoned that zebrafish chondrocyte transcriptome could yield potent biomarkers for chondrocytes. To validate and further filter our human orthologue dataset for the most relevant genes, we compared it to human fetal cartilage-selective gene set from Li et al. study (Li et al., 2017). This dataset contains gene expression data from five, independent 14 to 18-week-old human femur growth plate cartilage samples. This dataset reveals 1,205 genes with cartilage-selective expression levels (fold change>5). Comparing 1,112 human orthologs of zebrafish chondrocyte-enriched genes to human fetal cartilage-selective genes, we identified 211 common genes as potent candidates for chondrocyte biomarkers. Common genes are expectedly enriched for GO categories such as skeletal system development, cartilage development and ECM organization (Fig. 2.8A). Human fetal cartilage only genes (994 genes) were enriched for GO categories like signal transduction, programmed cell death, apoptotic process (Fig. 2.8B). High numbers of non-overlapping genes could be attributed to differences in experimental techniques. To generate zebrafish chondrocyte transcriptome, we performed

FACS-based sorting of genetically labeled, live chondrocytes without the need of mechanical dissection of cartilage tissue; while, human fetal cartilage tissues were first mechanically dissected during procedure, then enzymatically digested to collect chondrocytes. Though this collection method generates RNA-seq quality sample, it does not eliminate peripheral, non-chondrocytic cells or dead/ dying chondrocytes from collection. This could explain representation of cell death and apoptotic process-related genes only in human fetal-cartilage dataset. Cell death associated genes in human fetal cartilage-selective geneset included DAP3 (death associated protein 3), ANXA5 (annexin A5), BBC3 (BCL-2 binding component 3), which are established apoptosis markers. Zebrafish chonodrocyte-only geneset showed enrichment of several metabolic processes such as sulfur compound, carbohydrate and organic acid metabolic processes (Fig. 2.8C). These categories house several proteoglycan processing enzymes such as CHSY1 (chondroitin sulfate synthase 1), CHST12 (carbohydrate [chondroitin-4] sulfotransferase 12) and ST3GAL4 (ST3 beta-galactoside alpha-2,3-sialyltransferase 4), suggesting that zebrafish chondrocyte transcriptome is enriched with a greater number of matrix processing enzymes. Differences in representation of proteoglycan processing enzymes could be explained by diversity of matrix proteoglycans in zebrafish craniofacial cartilage and human femur growth plate cartilage.

211 common genes identified independently from zebrafish craniofacial chondrocytes and human fetal growth plate cartilage are potentially involved in conserved pathways defining basic chondrocyte biology and functions. Plotting the common genes on an interaction network based on their genetic and/or direct interaction among each other, it has become clear that majority of these genes form closely connected hubs around

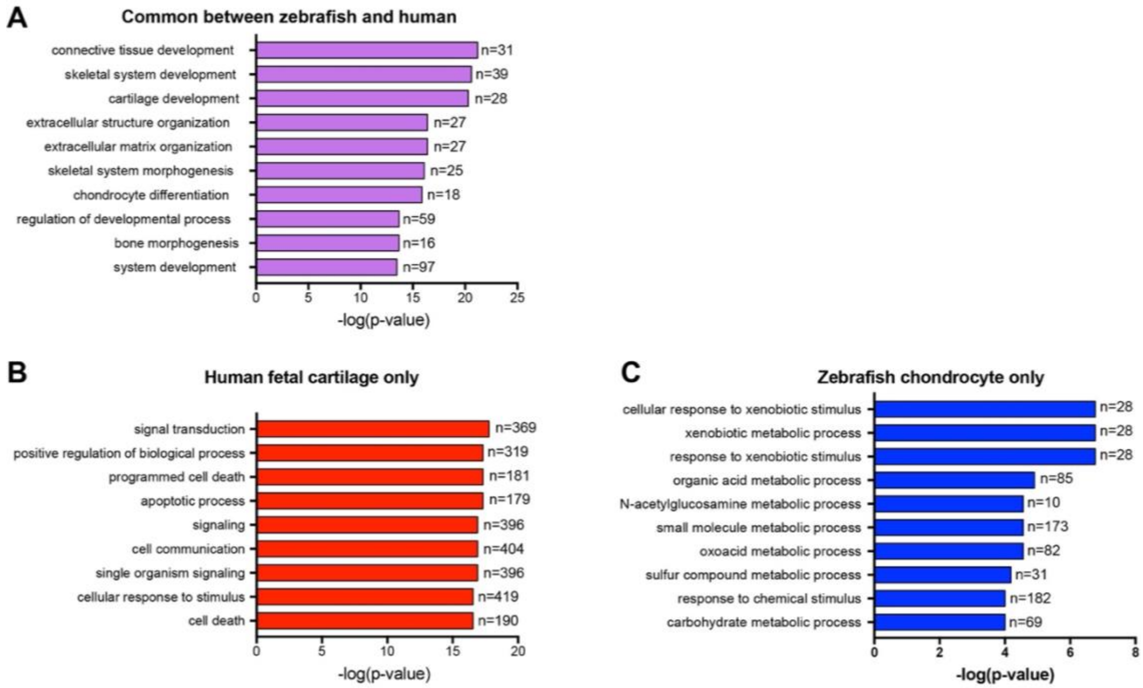


Figure 2.8. Comparative analysis of zebrafish chondrocyte-enriched genes and human fetal cartilage-selective genes.

Gene set enrichment analysis of GO terms for **A.** 211 common genes that are enriched 5-fold in both zebrafish and human datasets; **B.** 1003 genes enriched only in human fetal cartilage; **C.** 910 genes enriched only in zebrafish chondrocytes. Number of genes included in each GO category (n) was indicated next to each corresponding bar.

established cartilage markers such as COL2A1 (collagen type-II α1), COL9A1 (collagen type-IX α1), SOX9 and ACAN (aggrecan) (Fig. 2.9). 79 of the genes do not show any reported interaction, thus remain outside the hubs. Unconnected genes have great potential to uncover yet unidentified pathways regulating cartilage homeostasis.

Discussion

Vertebrate skeletal development relies heavily on chondrogenesis, a process through which mesenchymal stem cells differentiate into mature chondrocytes and secrete ECM, thereby facilitating formation of future endochondral bones (Goldring et al., 2006). The composition of matrix is particularly critical for cartilage homeostasis and endochondral ossification of long bones found in the axial, appendicular and craniofacial skeleton. Thus, identifying the complete spectrum of ECM components expressed by mature chondrocytes is needed to comprehensively study skeletal biology.

Isolation of a pure ECM producing chondrocyte sample is challenging because the dissected tissues such as long bone growth plate contains mixed stages of chondrocytes, from ECM-producing to proliferating and hypertrophic chondrocytes, plus neighboring cell types that are typically included in the sample. To tackle these problems thereby providing a relatively homogeneous source of cells (Kimmel et al., 1998). Second, zebrafish craniofacial chondrocytes can be specifically marked by a transgenic reporter and sorted by e.g. FACS, thus eliminating the need of cartilage dissection.

We have developed a robust protocol for isolation of a high purity and high integrity RNA sample from mature zebrafish craniofacial cartilage at 4 days of age, using FACS-isolated live chondrocytes from the Tg[Col2a1a:caax-EGFP] reporter line. Using Next Generation Sequencing on an Illumina platform, we sequenced the chondrocyte

transcriptome and established baseline expression levels. For comparison, we made an expression library using the remaining head tissue negative for GFP and obtained a complementary dataset.

Comparison of the GFP+ chondrocyte transcriptome with the GFP- non-chondrocytic craniofacial cell population resulted in 4,033 chondrocyte-enriched genes. Among the top 50 highly enriched genes, 19 represented structural cartilage matrix components such as collagen type-II, aggrecan, tenascin X and matrilin-1. And overall, 43

previously reported skeletal system developmental genes, and 32 cartilage developmental genes were detected.

Low cellularity and high proteoglycan / collagen content in cartilage presents a formidable challenge to isolating high quality RNA for RNA-seq experiments. Commercial kits are sufficient for RNA isolation from cells in culture and soft tissues; however, cartilage and bone present a unique challenge for RNA extraction. Harsh homogenization conditions for rigid and dense cartilage ECM are required, often damaging available cellular material for RNA isolation. Thus, RNA-seq datasets from normal chondrocytes are limited, and some of those available might not be entirely representative of physiologically normal cells (Le Bleu et al., 2017). Others' and our own attempts to isolate cartilage samples from human perioperative material generated substandard quality RNA and yielded samples not suitable for RNA-seq experiments. This prompted us to redesign the strategy for identification of the chondrocyte transcriptome and use the

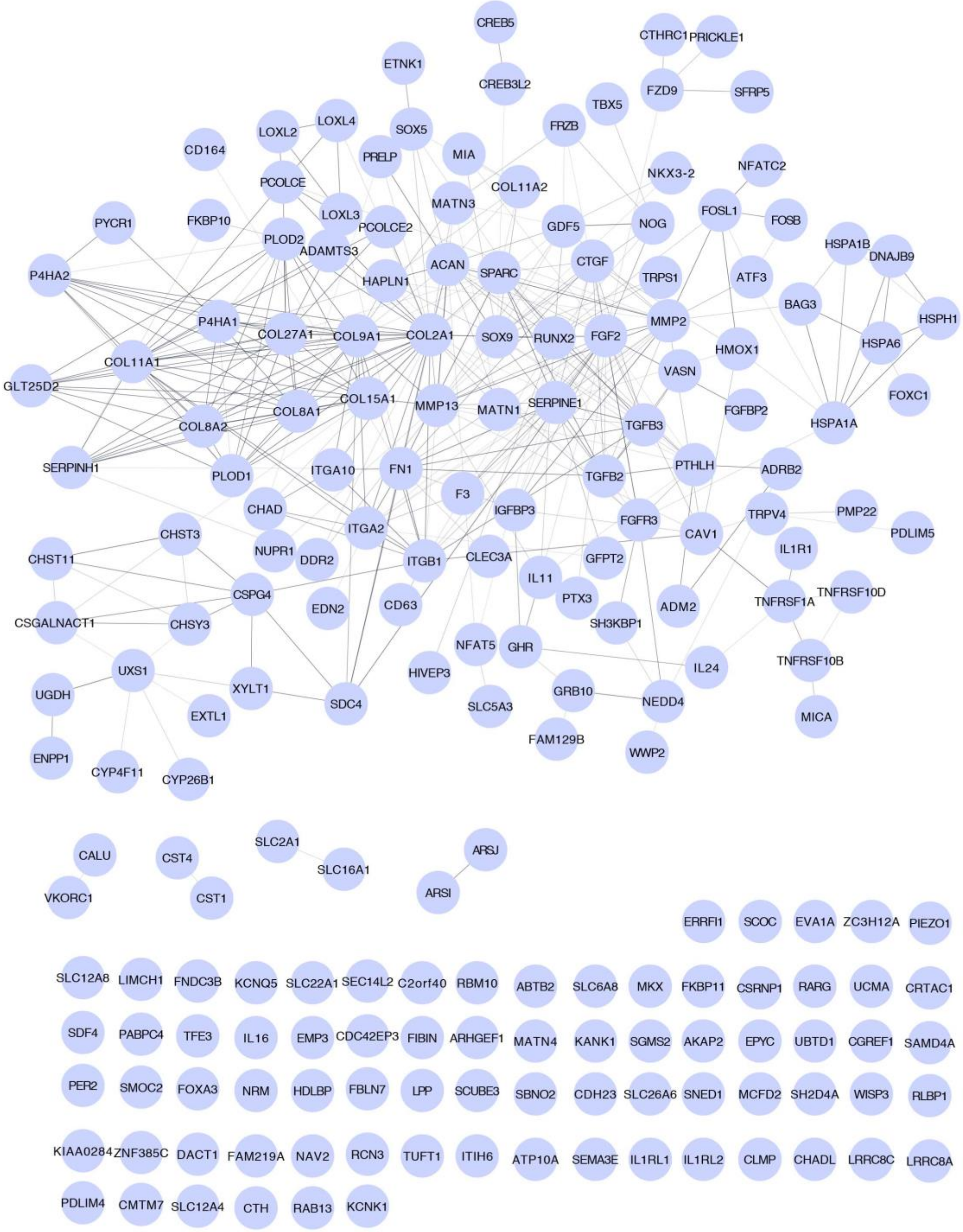


Figure 2.9. Network analysis of common genes enriched in both zebrafish chondrocytes and human fetal cartilage.

Protein association networks for 211 common genes. The 75 single genes are highly expressed in chondrocytes but their biological function is not yet linked to chondrocyte biology.

zebrafish model to develop an isolation protocol for live chondrocytes from mature embryonic cartilage. Zebrafish head skeleton at 4-dpf stage provides a non-calcified tissue housing homogeneous pool of chondrocytes that are already mature, stacked and producing large volumes of the ECM. Other advantages of this model include large number of easily accessible larvae to scale up the RNA preparation and availability of transgenic fish lines with robust fluorescent labeling of *in situ* live chondrocytes.

Forward genetic screens for phenotypes affecting craniofacial development identified numerous zebrafish mutations in genes essential for chondrocyte function (Eames et al., 2011b; Lang et al., 2006; Neuhauss et al., 1996; Piotrowski et al., 1996; Sarmah et al., 2010; Schilling et al., 1996; Yan et al., 2002). These genes include regulators of ECM synthesis and processing, patterning of the head skeleton and the secretory machinery. The chondrocyte transcriptome dataset described here corroborated high expression levels of COPII components essential for secretion of collagen and other ECM proteins (Lang et al., 2006; Melville et al., 2011; Sarmah et al., 2010). Furthermore, Gene Ontology analysis confirmed that the nature of the highly enriched genes in the GFP+ population is associated with skeletal development and function.

Genetic and biochemical studies have suggested that Sec23 and Sec24 paralogs have diverse functions in cargo selection and transport (Boyadjiev et al., 2006; Boyadjiev et al., 2011; Gorur et al., 2017; Lang et al., 2006; Saito and Katada, 2015; Sarmah et al., 2010). However, it has remained elusive whether spatial expression pattern of *sec23a* and *sec24d* paralogs correlates with that of their preferred cargos. Our cell-type-specific transcriptome study has given us a direct way to test this hypothesis. Chondrocyte-specific enrichment of *sec23a* and *sec24d* paralogs corroborated co-expression of these COPII components with their preferred cargo, type II collagen, in chondrocytes. Furthermore, the transcriptome dataset will provide researchers with yet unrecognized COPII cargos to be tested in animal models. Besides ECM components and their cargo adaptors required for ER-to-Golgi transport, this dataset also contains genes essential for synthesis, processing and trafficking in all cellular compartments. Coordinated response to developmental demands or to the repair and healing process is likely orchestrated by transcription factors. Up to date only Creb3L2 was implicated in driving expression of select COPII adaptors. The discovery of the regulatory networks controlling availability of cargo specific trafficking machinery in chondrocytes awaits further work.

Moreover, cell dissociation and sorting protocols developed here can be used to study chromatin dynamics and transcription factor binding during chondrocyte development. Efficient sorting of chondrocytes will be helpful in studies such as ChIP-seq. Alternatively; this method can be adapted to encompass long non-coding RNAs and miRNAs to study broader RNA populations in the context of chondrocyte biology.

Craniofacial and skeletal disorders are prevalent in human population, but not all are associated with a causative gene deficiency. Discovery of causative mutations in human skeletal disorders has relied heavily on genome-wide association studies (GWAS) and whole-exome sequencing (WES) approaches (Albagha et al., 2011; Davis et al., 2013; Gedeon et al., 1999; Hsu and Kiel, 2012). However, both methods leave researchers with several candidate genes to confirm, even after long and tedious genetic analysis steps. We propose that our reference dataset will greatly assist in discovery of novel genes linked to yet undiagnosed skeletal disorders. We found the highly enriched chondrocyte genes to be associated with human syndromes, such as osteo-chondro-dysplasia and joint diseases affecting skeletal system. In particular, comparative transcriptomics approach we took to identify 211 common genes between embryonic zebrafish chondrocytes and human fetal cartilage samples will be of great use to skeletal biology community. Future work using CRISPR/Cas9 genome editing approaches readily available to researchers (Auer and Del Bene, 2014; Jao et al., 2013; Varshney et al., 2015) to study the function of these candidate genes will be instrumental in disease gene modeling and developmental studies.

CHAPTER III

CREB3L2 TARGETS SEC24D AND CELL-AUTONOMOUSLY REGULATES COLLAGEN SECRETION IN CHONDROCYTES

Gokhan Unlu[1,2,3], Joseph H. Breeyear[1,2], Ela W. Knapik[1,2,3]

[1]Division of Genetic Medicine, Department of Medicine; [2]Vanderbilt Genetics Institute, Vanderbilt University Medical Center; [3]Cell & Developmental Biology, Vanderbilt University, Nashville, TN 37232, USA

Abstract

Development and homeostasis of skeletal tissues depend heavily on deposition of collagen-rich extracellular matrix. Due to their unconventionally large and fibrillar structures, collagen oligomers require specialized tubular COPII carriers to be transported from ER-to-Golgi *en route* to secretion to extracellular space. How skeletal cells spatiotemporally regulate the availability of collagen secretion machinery has long been a long-standing question. As an entry point to addressing this question, we and others have recently shown that collagen-specific, COPII inner coat component, *sec23a* is regulated by a multi-domain transcription factor Creb3L2 (cAMP response element binding protein 3-like 2) in mature chondrocytes. However, two critical questions about *in vivo* roles of Creb3L2 remain unanswered: 1) Is N-terminal transcription factor domain of Creb3L2 sufficient for physiological levels of collagen secretion in mature chondrocytes *in vivo*? 2) Does Creb3L2 regulate multiple components of collagen secretion machinery? Generating CRIPSR/Cas9-induced domain-specific zebrafish mutant lines, we found that both N- and C-termini of Creb3L2 need to be intact for *in vivo* levels of collagen secretion in chondrocytes. By chromatin immunoprecipitation and expression analyses in zebrafish embryos, we discovered that Creb3L2 directly activates expression of *sec24d*, a COPII inner coat component and binding partner of Sec23a. These results reveal that Creb3L2 targets multiple components of collagen transport machinery and its full-length form is necessary to provide for full capacity collagen secretion *in vivo*.

Introduction

Chondrocytes are highly specialized cells that secrete large amounts of extracellular matrix (ECM) proteins to build cartilage matrix. One of the hallmarks of chondrocyte differentiation is the onset of collagen II expression, the most abundant component of cartilage ECM. In many systems, it was shown that Sox9 is the transcription factor that activates collagen II expression in differentiated chondrocytes by binding to Col2 promoter (Bell et al., 1997; Lefebvre et al., 1997). Collagen II forms higher order structures in the form of triple helices (Grynpas et al., 1980; Kar et al., 2006) and is post-translationally modified before secretion (Myllyharju and Kivirikko, 2004). Procollagen, like other ECM proteins, is transported from ER to Golgi by coat protein II (COPII) coated vesicles that are composed of Sec23-Sec24-Sar1 inner coat and Sec13-Sec31 outer coat complex (Barlowe et al., 1994; Bonfanti et al., 1998; Lee et al., 2005; Miller et al., 2002; Paccaud et al., 1996; Salama et al., 1997). Traditionally, COPII vesicles have been considered common for all secretory proteins; however, recent evidence shows that bulky cargos like procollagen use a different version of COPII carriers that are as big as 300nm (Bonfanti et al., 1998), while small, globular proteins are transported in 60-90nm vesicles (Mironov et al., 2003). Therefore, regulation of COPII carrier size is the critical step for procollagen transport and secretion.

COPII carrier size can be determined by inner and outer coat components (Fromme et al., 2007; Kim et al., 2012) as well as post-translational modifications of COPII proteins (Jin et al., 2012). Sec23, an inner coat component, has two paralogs in vertebrates, Sec23a and Sec23b. Sec23a has been implicated in procollagen trafficking(Boyadjiev et al., 2011; Fromme et al., 2007). Mutations in *SEC23A* gene are associated with the human disease cranio-lenticulo-sutural-dysplasia (CLSD), an autosomal recessive syndrome with characteristics of facial dysmorphisms and skeletal defects (Boyadjiev et al., 2006). Moreover, a null mutation in zebrafish *sec23a*, known as *crusher*, leads to craniofacial abnormalities and intracellular accumulation of collagen II in chondrocytes (Lang et al., 2006). Further research found that Sec23a is specifically required for formation of mega-size procollagen carrying COPII vesicles (Kim et al., 2012). Since Sec23a is such a critical component for collagen trafficking, its transcriptional regulation in differentiated chondrocytes is equally important. The expression of *sec23a* is directly activated by the transcription factor Creb3L2 (Ishikura-Kinoshita et al., 2012; Saito et al., 2009a). In both mice and zebrafish, Creb3l2 mutations lead to intracellular collagen accumulation in chondrocytes and developmental craniofacial defects (Ishikura-Kinoshita et al., 2012; Melville et al., 2011; Saito et al., 2009a).

Creb3L2 (also known as BBF2H7) is a member of CREB/ATF family of transcription factors (TF) which are ER-resident transmembrane proteins. Members of this family are translocated to Golgi and cleaved there through regulated intramembrane proteolysis (RIP) (Ye et al., 2000). Upon cleavage, N-terminal portion moves to nucleus and acts as a basic leucine zipper (bZIP) transcription factor via binding to cAMP response elements (CRE) within promoter regions of target genes (Fig. 3.1). A recent study showed that ER-luminal C-terminus of Creb3L2 is secreted to ECM and induces chondrocyte proliferation through indian hedgehog (Ihh) signaling in mouse cartilage tissue (Saito et al., 2014). However, following two questions remain in Creb3L2's role in chondrocyte function: 1) whether C-terminal secreted peptide (C-SP) carries a function in mediating

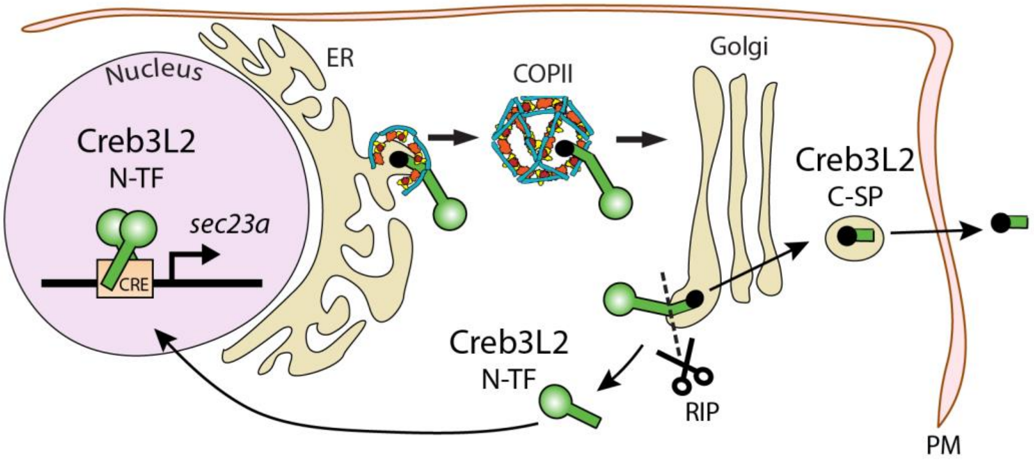


Figure 3.1. Trafficking of and processing of Creb3L2
Creb3L2 is transported from ER to Golgi complex within COPII vesicles. It is cleaved throughout RIP, upon which N-TF translocates to the nucleus, lumenal C-SP is secreted to the extracellular space. Nuclear N-TF dimers bind to CRE sites within DNA and activates expression of target genes (e.g. *sec23a*). ER: endoplasmic reticulum, PM: plasma membrane, RIP: regulated intramembrane proteolysis, N-TF: N-terminal transcription factor, C-SP: C-terminal secreted peptide.

collagen secretion and 2) whether Creb3L2 directly targets transcription of other genes, besides *sec23a*, to promote collagen secretion from cells.

Here, we present evidence using zebrafish genetic tools that C-SP portion of Creb3L2 is required for efficient collagen secretion out of chondrocytes and only full-length Creb3L2 is sufficient to provide physiological levels of collagen secretion. We found that Creb3L2 directly binds to CRE site within *sec24d* promoter in zebrafish chondrocytes, and activates its expression. Our findings uncover comprehensive functions of Creb3L2 in chondrocyte maturation and cartilage matrix production through directly targeting Sec23a-Sec24d dimer complex of inner COPII coat, and consequently promoting collagen trafficking.

Methods

Generationof creb3L2 mutant lines with CRISPR/Cas9 system

CRISPR/Cas9 target sites within zebrafish *creb3L2* gene (GRCz10 assembly) were identified by searching for GGN$_{18}$<u>NGG</u> sites within the exons (either + or – strands). Underlined sequence indicates protospacer adjacent motif (PAM). Following sites were targeted in this study;

creb3L2-g1: GGAGAACCCGGCAGCGACGC<u>TGG</u>

creb3L2-g12: GGTGGTGTAGGAGTCGTGTA<u>TGG</u>

These guide RNA (gRNA) target sites were cloned into pT7cas9gRNA2 vector (Jao et al., 2013). To use as DNA template for *in vitro* transcription, gRNA constructs were linearized with BamHI. Creb3L2-1 and -g12 gRNAs were synthesized with MEGAshortscript T7 transcription kit (ThermoFisher). To generate mutations with CRISPR/Cas9 system, a mixture of 500pg Cas9 mRNA and 100 pg - 500pg gRNA was injected into one-cell stage embryos as described before (Jao et al., 2013). Injected fish were raised to adulthood to generate founders (G_0).

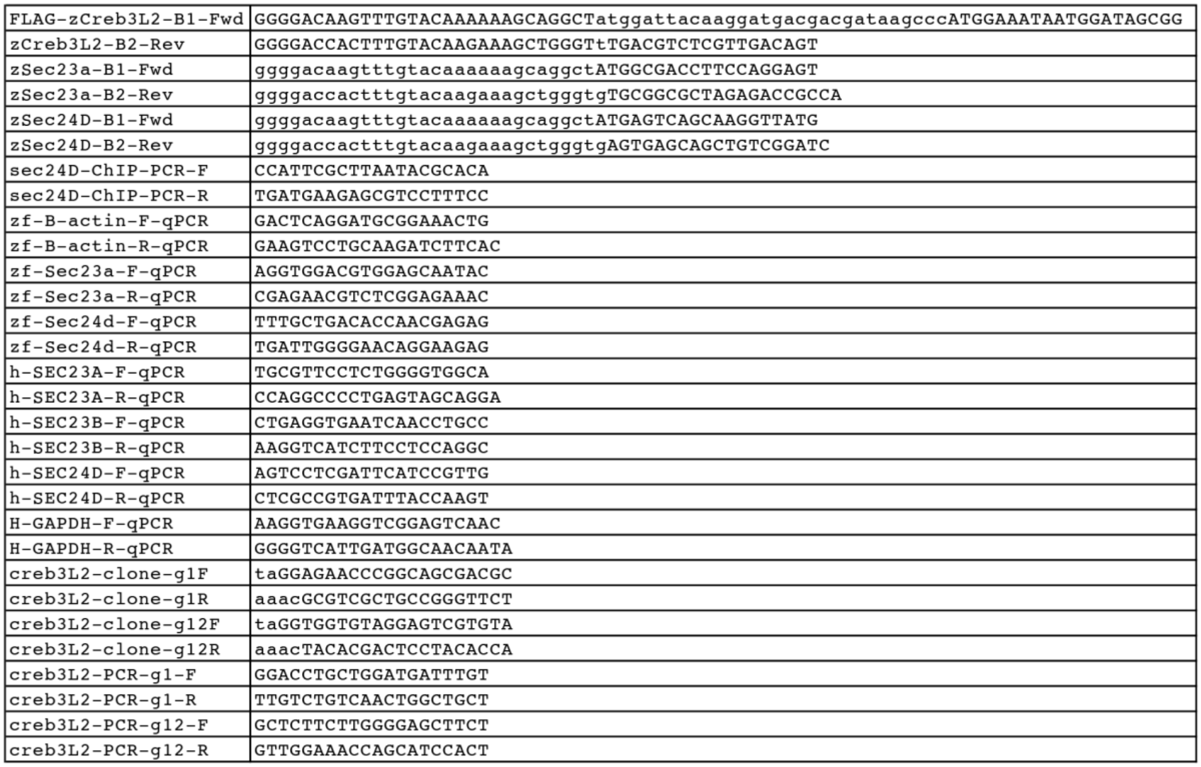
To screen for germline transmission, potential founders were crossed to wild-type (AB) fish. Their embryos were lysed to extract genomic DNA by boiling in lysis buffer (50µM KCl, 10µM Tris pH:8.0, 0.3% Tween 20) at 100ºC for 10 minutes, then digesting with 1µg/µl proteinase K (Fisher Bioreagents) for 30 minutes at 55ºC and boiling at 100ºC for 10 minutes again. CRISPR target sites were amplified from extracted genomic DNA using PCR primer sets listed in Table 3.1. To detect potential heterozygous mutants, heteroduplex formation assay was performed as described previously (Yin et al., 2015). Briefly, PCR products were re-hybridized with the following settings: incubation at 95ºC for 5 minutes, cooling 2ºC per second down to 85ºC, cooling 0.1ºC per second down to 25ºC. Samples were loaded on polyacrylamide gel (10% acrylamide (29:1, 30% stock, Biorad#1610156), 1X TBE, 0.05% ammonium persulfate, 0.05% TEMED) and run at 120V for 2 hours to detect heteroduplexes as slow running, upper bands on the gel. Polyacrylamide gels were soaked in 1 X TBE buffer with ethidium bromide (0.5 µg/ml) for 15 minutes to stain DNA bands. Gel images were acquired by Biorad Gel Doc imaging system.

PCR products of CRISPR targets sites were cloned into pGEM-T easy (Promega) following instructor’s manual. Plasmids isolated from pGEM-T clones were sequenced with SP6 primer via Sanger sequencing method (Genewiz, NJ). Mutant sequences were aligned with WT *creb3L2* sequence with NCBI/NLM's BLAST® tool to detect deletions, insertions and substitutions.

Construction of Gateway Vectors for Tol2kit System

Zebrafish *creb3l2* coding sequence was amplified from pCS2+_Creb3L2 (Melville et al., 2011) using primers containing attB1 (forward) and attB2 (reverse) recombineering sites. FLAG tag sequence was added to forward primer in order to tag N-terminus of Creb3L2. PCR product was purified and cloned into pDONR221 vector by Gateway® BP clonase II enzyme mix. Obtained clone served as middle entry vector for Tol2kit Multistep Gateway Recombineering system (Kwan et al., 2007). FLAG-Creb3L2 middle entry vector was incubated with p5E_1.7kbCol2a1a promoter (Dale and Topczewski, 2011), p3E_v2a-EGFP (shared by Dr. Josh Gamse), pDestTol2pA2 and Gateway® LR clonase II enzyme mix overnight at room temperature. After transformation and colony screening, destination vector that drives expression of FLAG-Creb3L2 coding sequence fused with 'self-cleaving viral 2a peptide' tagged EGFP (v2a-EGFP) under tissue-specific Col2a1a promoter was acquired. Zebrafish *sec24d* was cloned for Tol2kit system using the same protocol, from pCS2+_Sec24d vector (Sarmah et al., 2010). Instead of v2a-EGFP tag, v2a-mCherry was used for sec24d destination construct. Zebrafish Sec23a was cloned likewise, from pCS2+_Sec23a vector (Lang et al., 2006). 5' element p5E-β-actin (Kwan et al., 2007) was utilized in LR cloning step, in order to express Sec23a-v2a-EGFP under β-actin promoter.

Table 3.1. Primer sequences



FLAG-zCreb3L2-B1-Fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTatggattacaaggatgacgacgataagcccATGGAAATAATGGATAGCGG
zCreb3L2-B2-Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTtTGACGTCTCGTTGACAGT
zSec23a-B1-Fwd	ggggacaagtttgtacaaaaaagcaggctATGGCGACCTTCCAGGAGT
zSec23a-B2-Rev	ggggaccactttgtacaagaaagctgggtgTGCGGCGCTAGAGACCGCCA
zSec24D-B1-Fwd	ggggacaagtttgtacaaaaaagcaggctATGAGTCAGCAAGGTTATG
zSec24D-B2-Rev	ggggaccactttgtacaagaaagctgggtgAGTGAGCAGCTGTCGGATC
sec24D-ChIP-PCR-F	CCATTCGCTTAATACGCACA
sec24D-ChIP-PCR-R	TGATGAAGAGCGTCCTTTCC
zf-B-actin-F-qPCR	GACTCAGGATGCGGAAACTG
zf-B-actin-R-qPCR	GAAGTCCTGCAAGATCTTCAC
zf-Sec23a-F-qPCR	AGGTGGACGTGGAGCAATAC
zf-Sec23a-R-qPCR	CGAGAACGTCTCGGAGAAAC
zf-Sec24d-F-qPCR	TTTGCTGACACCAACGAGAG
zf-Sec24d-R-qPCR	TGATTGGGGAACAGGAAGAG
h-SEC23A-F-qPCR	TGCGTTCCTCTGGGGTGGCA
h-SEC23A-R-qPCR	CCAGGCCCCTGAGTAGCAGGA
h-SEC23B-F-qPCR	CTGAGGTGAATCAACCTGCC
h-SEC23B-R-qPCR	AAGGTCATCTTCCTCCAGGC
h-SEC24D-F-qPCR	AGTCCTCGATTCATCCGTTG
h-SEC24D-R-qPCR	CTCGCCGTGATTTACCAAGT
H-GAPDH-F-qPCR	AAGGTGAAGGTCGGAGTCAAC
H-GAPDH-R-qPCR	GGGGTCATTGATGGCAACAATA
creb3L2-clone-g1F	taGGAGAACCCGGCAGCGACGC
creb3L2-clone-g1R	aaacGCGTCGCTGCCGGGTTCT
creb3L2-clone-g12F	taGGTGGTGTAGGAGTCGTGTA
creb3L2-clone-g12R	aaacTACACGACTCCTACACCA
creb3L2-PCR-g1-F	GGACCTGCTGGATGATTTGT
creb3L2-PCR-g1-R	TTGTCTGTCAACTGGCTGCT
creb3L2-PCR-g12-F	GCTCTTCTTGGGGAGCTTCT
creb3L2-PCR-g12-R	GTTGGAAACCAGCATCCACT

Microinjection of Tol2kit constructs

50 pg medaka transposase mRNA and 10pg of destination vector was injected into 1-cell stage zebrafish embryos collected from feelgood^{m662}/AB, creb3L2^{g1}/AB, creb3L2^{g12}, *crusher*^{m299}/AB or bul^{m606}/AB heterozygous crosses.

Cryosectioning and Immunohistochemistry

Injected embryos were allowed to grow at 28.5ºC until the indicated stages and fixed in 4% PFA at 4ºC overnight. Fixative was removed and embryos were moved to 30% sucrose solution (in PBS) for overnight incubation at 4ºC. Embryos were embedded in Cryomatrix™ (ThermoFisher Scientific) and frozen at -80ºC for 15 minutes, then 14 µm-thick sections were taken with cryostat (Leica CM1900) and collected onto Fisherbrand™ Superfrost™ Plus slides. Slides were dried for 30 minutes and rehydrated in PBS before staining. Antigen retrieval was performed by incubating in 20µg/ml proteinase K at room temperature for 5 minutes; permeabilization was performed with 0.5% triton X-100 in PBS for 10 minutes at room temperature. Slides were, in turn, incubated in blocking solution (2% goat serum & 2mg/ml BSA) for 30 minutes at room temperature, followed by overnight primary antibody incubation (1:250 Collagen-II antibody [Rockland 600-401-104] and 1:250 GFP antibody [chicken polyclonal GFP antibody by Vanderbilt Antibody and Protein Resource] or 1:250 mCherry [Clonetech, 632543]) at 4ºC. Slides were then incubated in secondary antibodies (Rabbit-Alexa Fluor®-555 or -488 and Chicken-Alexa Fluor®-488 or Mouse-Alexa Fluor®-555, LifeTechnologies) for one hour, and 1:4000

DAPI for 15 minutes at room temperature. Finally, slides were mounted in ProlongGold® anti-fading agent (ThermoFisher Scientific).

Imaging and Quantification

Slides were imaged with Zeiss AxioImager.Z1. 'Percent collagen area in cytosol' is calculated by the following formula: (Collagen-positive intracellular area / [cytosol area - nucleus area marked by DAPI])*100. ImageJ was used for intracellular area measurements.

HEK293 Cell Culture & Transfection

HEK293 cells were grown in DMEM (GIBCO 11995), supplemented with 10% FBS (GIBCO 26140), 1% GlutaMAX (GIBCO 35050061) and 1% non-essential amino acids (Sigma M7145) at 37ºC with 5% CO_2. HEK293 cells were transfected with either pCMV-SPORT6-humanCREB3L2 (full length) construct (Harvard Plasmid ID) or pCS2+_EGFP as control using Lipofectamine® 2000 (Thermo Fisher Scientific) reagent following manufacturer's instructions.

Quantitative PCR Analysis

Zebrafish: Quantitative real-time PCR (qRT-PCR) was performed as described previously (Sarmah et al., 2010). Total RNA was extracted from 10-15 embryos (per sample) at 5 dpf stage using the TRIzol reagent (ThermoFisher Scientific). 500 ng of total RNA was used as template for reverse transcription to synthesize cDNA using M-MLV reverse transcriptase (Promega) and poly-T primer. Each PCR reaction was performed with 1 µl of cDNA, 1X SYBR Green Real-Time PCR Master Mix and 2 µM of each primer. qRT-PCR reactions were run on CFX96 (Biorad) system. Data were analyzed with –ΔΔCt method via normalizing against β-actin control.

HEK293 cells: RNA was isolated from transfected HEK293 using TRIzol reagent (ThermoFisher Scientific). 500 ng of total RNA was used as template. The same protocol for qRT-PCR setup was used as described above.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) protocol was modified from published protocols (Bogdanovic et al., 2013; Vastenhouw et al., 2010) with following modifications. 175 transgenic (Col2:FLAG-Creb3L2-v2a-egfp) embryos per ChIP sample were collected at 4 dpf stage and transferred to a 15 ml conical tube. Egg water was aspirated and embryos were fixed in 1.86% fresh paraformaldehyde (PFA) at pH 7.4 with gentle shaking at room temperature for 15 minutes. Glycine was added to a final concentration of 0.125 M in order to quench PFA. Embryos were incubated in 0.125 M glycine by rocking end to end for 5 minutes at room temperature. Then, glycine was removed and replaced with ice-cold 1X PBS by rinsing in it three times. After removing PBS, 500 µl of cell lysis buffer (10 mM Tris-HCl pH7.5, 10 mM NaCl, 0.5% IGEPAL) was added. Next, embryos were distributed into 4 microfuge tubes (~40 embryos per tube per 125µl cell lysis buffer). Using microtube homogenizer pestles, embryos were homogenized at every 5 minutes on ice, for total of 15 minutes. Microfuge tubes were centrifuged at 3,500 rpm for 5 minutes at 4ºC. Supernatant was removed and pellet was resuspended in 400µl of nuclei lysis buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% SDS). Pellet was pipetted up and down to lyse the nuclei and tubes were left on ice for 10 minutes. 800 µl (2 volumes) of IP dilution buffer (16.7 mM Tris-HCl pH 7.5, 167 mM NaCl, 1.2 mM EDTA, 0.01% SDS) was added to each tube and sample was split into 3 microfuge tubes (400 µl per tube). Chromatin was sonicated using Bioruptor sonicator bath (Diagenode) at high setting (30 sec ON, 30 sec OFF) for 5 minutes, a total of 4 times. Water bath was replenished with ice at the end of every 5-minute cycle. 32µl of 10% triton X-100 (Sigma #T8787) was added to each tube of 400 µl sonicated chromatin. Tubes were centrifuged at 14,000 rpm for 10 minutes at 4ºC. Supernatant was stored at -80ºC for immunoprecipitation.

For immunoprecipitaion, 25 µl Protein G beads (Dynabeads®, Invitogen #10003D) were washed in 1 ml fresh block solution (0.5% BSA in PBS) per sample and rocked for 5 minutes at 4ºC. Beads were, then, collected by using a magnetic stand. Wash was repeated 2 more times to block the beads. Beads were collected and 100 µl of block solution containing 1 µl antibody was added to each tube. Antibodies used in this study are anti-Flag (Sigma, F3165), anti-IgG (Santa Cruz, SC2027), anti-H3K4me3 (Abcam, ab8580). Antibodies were incubated with magnetic beads overnight at 4ºC to facilitate binding. After overnight incubation, antibodies were removed and magnetic beads were washed with 1 ml block buffer, three times. Beads were collected with magnetic stand and resuspended in 100 µl block buffer & 300 µl of sonicated chromatin. Bead-chromatin mix was incubated overnight at 4ºC on rocker. Next day, magnetic beads were collected with magnetic stand and washed in 1 ml RIPA wash buffer (50 M HEPES pH 7.6, 1 mM EDTA, 0.7 % DOC, 1% IGEPAL, 0.5% LiCl) for 10 minutes on rocker, three times. Beads were collected, resuspended in 1 ml of 1X TBS (50 mM Tris pH 7.5, 150 mM NaCl) and rocked for 3 minutes at 4ºC. Tubes were centrifuged at 3000 rpm for 3 minutes, TBS was removed and

660 µl of elution buffer (50 mM $NaHCO_3$, 1% SDS) was added onto the beads. To elute DNA-protein complex off of the beads, tubes were incubated at 65ºC for 15 minutes while vortexing briefly at every 2 minutes. Then, beads were spun down at 14,000 rpm for 1 minute. 650 µl of supernatant was transferred into a microfuge tube and 46µl of 3M NaCl was added to reverse-crosslink DNA from chromatin, at 65ºC overnight. Next day, RNaseA (Thermo Fisher Scientific, 12091039) was added to a final concentration of 0.33 µg/µl to each tube and incubated for 2 hours at 37ºC. 1 volume of Phenol/ Chloroform/ Isomaylalcohol (25:24:1) was added, mixed and centrifuged at 14,000 rpm for 5 minutes. Upper phase was transferred to a new microfuge tube. To precipitate DNA, 20 µg glycogen and 1/10 volume of 3M sodium acetate were added, and tubes were centrifuged at 14,000 rpm for 5 minutes. Supernatant was removed and DNA pellet was air-dried at room temperature and resuspended in 35 µl of nuclease-free water. 3 µl of obtained ChIP DNA was used to amplify potential Creb3L2 binding sites within *sec24d* promoter (see primer table for sequences) in a 10µl PCR. As input, sonicated chromatin was diluted in 3 volumes of elution buffer and reverse-crosslinked with 0.2 M NaCl likewise, as described above.

Results

In vivo functional analysis of Creb3L2 through domain-specific mutations

To test *in vivo* functions of Creb3L2 functional domains, we created zebrafish genetic mutant lines with CRISPR/ Cas9 system (Jao et al., 2013). Of several guide RNAs (gRNAs) tested, two (i.e. creb3L2-g1 and -g12) showed great efficiency in generating deleterious mutations in transient assays (Fig. 3.2). Exon 2 was targeted with gRNA creb3L2-g1 to generate a full *creb3L2* knockout since it targets upstream of both DNA binding bZIP (basic leucine zipper) domain and C-SP. Exon 9 targeting creb3L2-g12 gRNA was selected to make a specific C-SP deletion as it targets between bZIP and transmembrane domains. Mutations occurring in g12 target region is predicted to leave cytoplasmic Creb3L2 intact, and free of activation requirement by RIP (Fig. 3.2A-C). By microinjecting diluted amounts of creb3L2-g1 and -g12 gRNAs along with Cas9 mRNA into one-cell stage zebrafish embryos, we raised founder fish (G_0) carrying *creb3L2* mutations in their germline. Successive outcrosses with wild-type line (AB) helped clean potential non-specific mutations out of the background and confirmed F2 adult carriers were crossed for functional analyses.

Zebrafish *creb3L2* mutant, *feelgood*^{m662} (*fel*) isolated from a large-scale ENU-screen for craniofacial mutants (Neuhauss et al., 1996) already provides a powerful tool to investigate transcription factor activity of Creb3L2 since it specifically hits DNA binding domain with a missense mutation in a conserved lysine residue (Fig. 3.2A) (Melville et al., 2011). Zebrafish *fel* mutants present with short body stature, malformed jaw and cartilage structures due to defective collagen secretion (Melville and Knapik, 2011; Melville et al., 2011). These results already suggested that transcription factor activity of Creb3L2 was required for collagen trafficking at the cellular level; body length elongation and craniofacial cartilage development at anatomical level. To investigate whether Creb3L2, outside its bZIP transcription factor domain, carries a physiological activity to regulate these processes, we first analyzed the morphology of domain-specific *creb3L2*-mutants generated via CRISPR/Cas9 system. As shown before, we detected significant reduction in body length of *fel*$^{-/-}$ mutants at 5 days post-fertilization (dpf). However, neither of *creb3L2* $^{g1/g1}$ or *creb3L2* $^{g21/g12}$ displayed a statistical difference in body length from wild

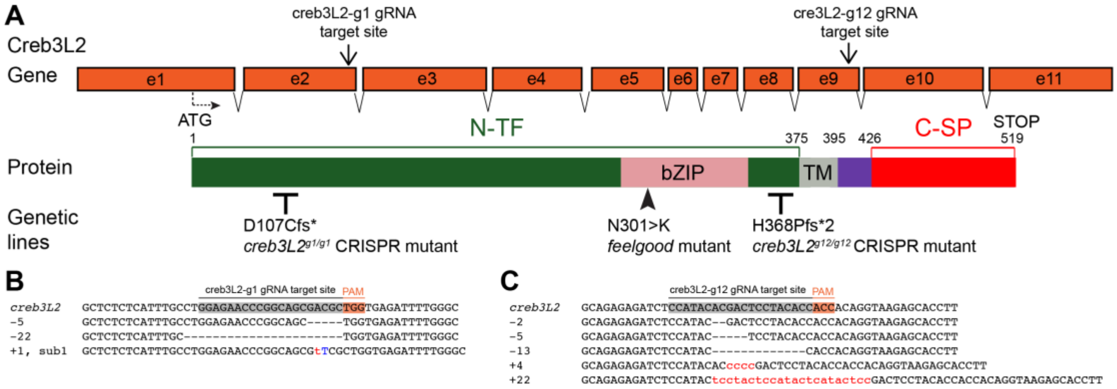

A
Creb3L2
Gene
creb3L2-g1 gRNA target site
cre3L2-g12 gRNA target site
e1
e2
e3
e4
e5
e6
e7
e8
e9
e10
e11
ATG
N-TF
C-SP
STOP
1
375
395
426
519
Protein
bZIP
TM
Genetic lines
D107Cfs*
creb3L2^g1/g1 CRISPR mutant
N301>K
feelgood mutant
H368Pfs*2
creb3L2^g12/g12 CRISPR mutant
B
creb3L2-g1 gRNA target site
PAM
creb3L2 GCTCTCTCATTTGCCTGGAGAACCCGGCAGCGACGCTGGTGAGATTTTGGGC
-5 GCTCTCTCATTTGCCTGGAGAACCCGGCAGC-----TGGTGAGATTTTGGGC
-22 GCTCTCTCATTTGC----------------------TGGTGAGATTTTGGGC
+1, sub1 GCTCTCTCATTTGCCTGGAGAACCCGGCAGCGtTCGCTGGTGAGATTTTGGGC
C
creb3L2-g12 gRNA target site
PAM
creb3L2 GCAGAGAGATCTCCATACACGACTCCTACACCACCACAGGTAAGAGCACCTT
-2 GCAGAGAGATCTCCATAC--GACTCCTACACCACCACAGGTAAGAGCACCTT
-5 GCAGAGAGATCTCCATAC-----TCCTACACCACCACAGGTAAGAGCACCTT
-13 GCAGAGAGATCTCCATAC-------------CACCACAGGTAAGAGCACCTT
+4 GCAGAGAGATCTCCATACACccccGACTCCTACACCACCACAGGTAAGAGCACCTT
+22 GCAGAGAGATCTCCATACtcctactccatactcatactccGACTCCTACACCACCACAGGTAAGAGCACCTT

Figure 3.2. CRISPR/Cas9 target sites on genomic structure of creb3L2
A. Schematics of gene and protein structures for Creb3L2. Guide RNAs (gRNAs) targeting exon 2 (e2) and e9 are marked on the gene structure. Genetic mutant lines and predicted changes in the primary sequence of Creb3L2 protein are indicated below the protein structure. **B-C**. Sequenced mutations generated with CRISPR/Cas9 system using **B.** *creb3L1-g1* and **C.** *creb3L2-g12* gRNAs. Dashes (-) indicate deletions. Red sequence mark insertions; blue mark substitutions (sub).

type embryos (Fig 3.3A-B). Transcription-factor domain-mutant *fel*$^{-/-}$ embryos have significantly reduced head sizes even though their trunk lengths do not show statistically significant deviation from WT (Fig. 3.3B). This observation indicated that craniofacial development is more severely affected by the loss of Creb3L2 TF activity than body length. Since both *creb3L2* $^{g1/g1}$ and *creb3L2* $^{g21/g12}$ failed to recapitulate reduction in head size (Fig. 3.3B), we sought out to investigate craniofacial cartilage structures more closely to be able to detect potentially subtler morphological changes. Alcian blue staining at 5 dpf for cartilage matrix proteoglycans did indeed display morphological alterations in both *creb3L2*$^{g1/g1}$ and *creb3L2*$^{g21/g12}$ mutants, though not as severe as *fel*$^{-/-}$ (Fig. 3.4A). Specifically, ceratohyal cartilage elements of all *creb3L2* mutants were significantly shorter in size than WT (Fig. 3.4B). These data reveal that both N-terminal TF and C-SP portions of Creb3L2 are essential for proper cartilage growth.

Full-length Creb3L2 is required for cell-autonomous collagen secretion in chondrocytes

Cartilage growth at 5 dpf stage zebrafish relies heavily on deposition of collagen-II-rich ECM. Transcription-factor domain-mutant *fel*$^{-/-}$ chondrocytes were shown to have collagen transport defects, and consequently compromised cartilage matrix. To assess whether cartilage growth defects in *creb3L2*$^{g1/g1}$ and *creb3L2*$^{g12/g12}$ mutants result also from collagen transport defects, we performed immunofluorescence for collagen-II. Initial analyses corroborated collagen transport defects in all *creb3L2* mutants just like in *fel*$^{-/-}$; thus, we pursued to investigate whether Creb3L2 acts cell autonomously to regulate collagen secretion in chondrocytes. To be able to run mosaic rescue analysis *in vivo*, we

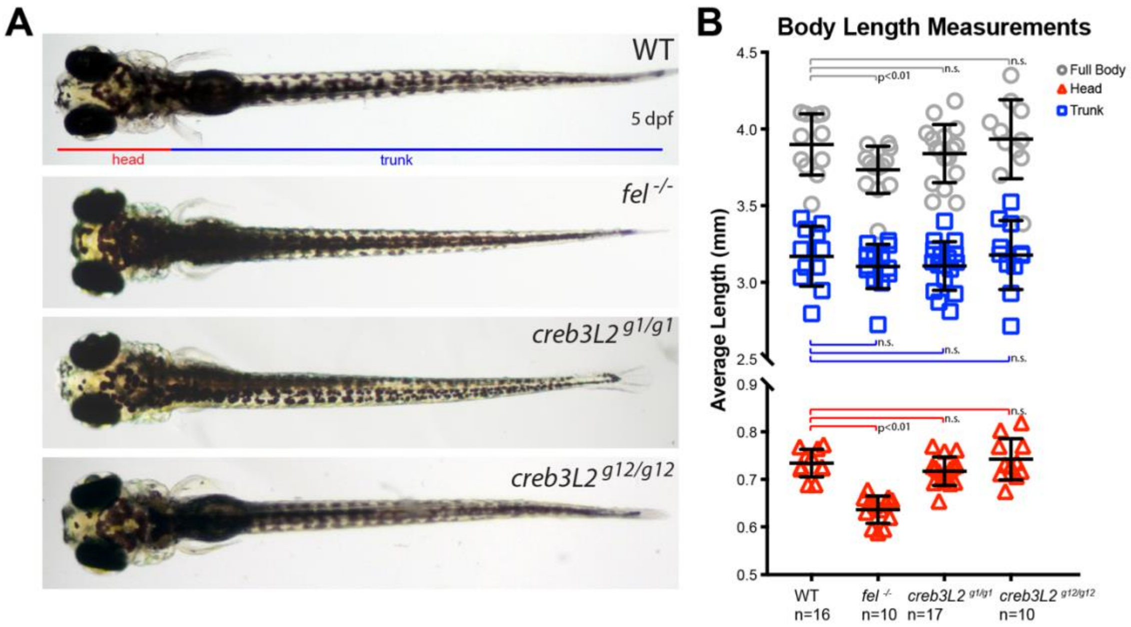


Figure 3.3. Body length elongation phenotype in domain-specific *creb3L2* mutants
A. Representative images of dorsal views of 5 dpf stage zebrafish larvae with indicated domain-specific *creb3L2* mutations. **B.** Full body, head and trunk length measurements of multiple larvae with indicated *creb3L2* genotypes.

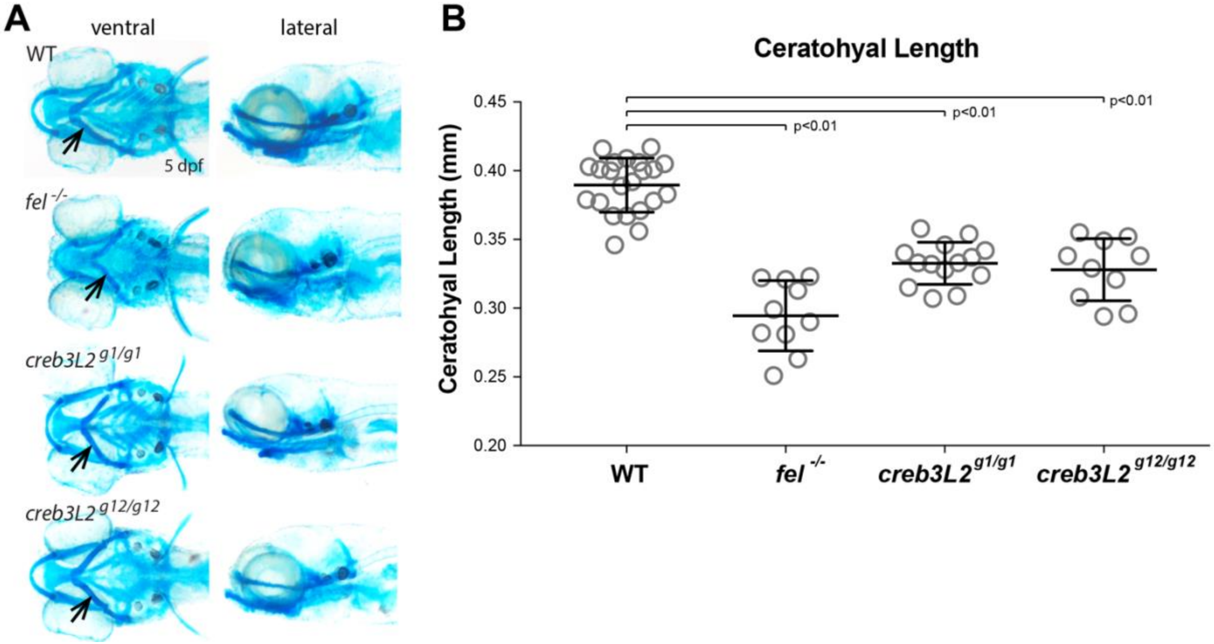


Figure 3.4. Craniofacial cartilage growth is reduced in *creb3L2* mutants

A. Alcian blue staining of 5 dpf zebrafish larvae displaying craniofacial cartilage elements. Arrow indicates ceratohyal cartilage. **B.** Ceratohyal length measurements of *creb3L2* mutants. All mutant groups show significant reduction in ceratohyal size compared to WT.

devised a mosaic overexpression approach using self-hydrolyzable V2a-EGFP reporter (Fig. 3.5A). Utilizing transposon-based Tol2kit system (Kwan et al., 2007), we mosaically overexpressed N-terminally FLAG-tagged full-length Creb3L2 in Col2-positive craniofacial chondrocytes. While *creb3L2*-deficient chondrocytes (GFP^{-}) contain intracellular collagen deposits, neighboring *creb3L2*-mutant cells expressing FLAG-Creb3L2 (marked with V2a-EGFP [GFP+]) have secreted collagen to ECM, thus are devoid of intracellular collagen staining, just like WT chondrocytes (Fig. 3.5B). Quantification of percentage of collagen occupying area within individual chondrocytes show significant levels of rescue with full-length Creb3L2 construct at the cellular level (Fig. 3.5C). These data suggested that Creb3L2 is required in a cell-autonomous manner for efficient collagen transport.

C-SP-only knockout (*creb3L2*$^{12/g12}$) chondrocytes contained similar levels of intracellular collagen accumulation to the levels detected in full knockout (*creb3L2*$^{g1/g1}$) and transcription-factor domain-mutant, *fel*$^{-/-}$ chondrocytes. This finding suggested that C-terminal portion might very well play a role in collagen secretion activity of Creb3L2. Rescue experiments revealed that C-terminus' effect on collagen secretion is cell-autonomous since collagen accumulation in C-SP-lacking *creb3L2*$^{g12/g12}$ mutant chondrocytes could be rescued cell-autonomously with overexpression of full-length Creb3L2 (GFP+ cells), whereas their immediate neighbors do have noticeably large collagen backlogs (Fig. 3.5B, bottom panel).

Overall, these data suggest that only full-length Creb3L2 is sufficient to modulate physiological levels of collagen secretion in mature chondrocytes. Neither N-terminal TF nor C-terminal C-SP portions of Creb3L2 alone can compensate for this function *in vivo*.

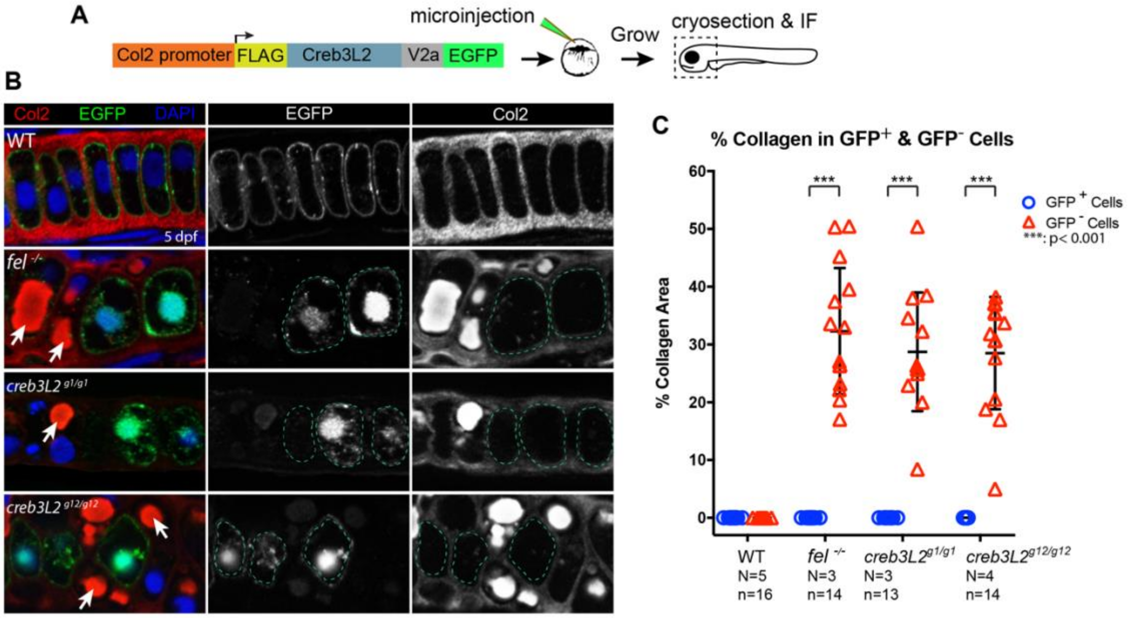


Figure 3.5. Creb3L2 acts cell-autonomously to promote collagen secretion
A. Experimental design to mosaically overexpress FLAG-tagged full-length Creb3L2 with V2a-EGFP tag in zebrafish chondrocytes and analyze by IF on cryosections. **B.** IF for Col2 and EGFP on cryosectioned chondrocytes mosaically overexpressing the FLAG-Creb3L2-V2a-EGFP construct. EGFP marks transgenic cells with construct expression in all *creb3L2* mutants. WT cells were marked with stably expressed caax-EGFP to label the cell membrane and secreted Col2 signal beyond the membrane. Arrows point to intracellular collagen accumulations. **C.** Quantification of percentage of intracellular collagen in cells expressing FLAG-Creb3L2 construct (GFP$^+$) and neighboring non-expressing cells (GFP$^-$). Cells expressing FLAG-Creb3L2 construct close to 0% intracellular collagen across all groups. Non-transgenic chondrocytes in all three groups of *creb3L2* mutants exhibit, on average, about 30% intracellular collagen.

Creb3L2 targets multiple genes to regulate collagen secretion

It was shown in both zebrafish and mouse studies that Creb3L2 depletion resulted in downregulation of inner COPII coat component, *sec23a*, thereby failing procollagen-II to exit ER (Melville et al., 2011; Saito et al., 2009a). Furthermore, Creb3L2 has been reported to directly activate *sec23a* expression in cultured mouse chondrocytes. Yet, it has not been investigated *in vivo* whether Sec23a is the only and sufficient target of Creb3L2 to activate collagen secretion. Mosaic overexpression approach we developed in this study, using V2a-EGFP reporter, has become useful in addressing this question in zebrafish chondrocytes *in vivo* (Fig 3.6A). To focus solely on transcriptional targets of Creb3L2, we chose to use transcription-factor domain-mutant *fel*$^{-/-}$ embryos (named *creb3L2*$^{-/-}$ hereafter) in mosaic rescue experiments. Overexpression of FLAG-Creb3L2 cell-autonomously rescued collagen transport defects in *creb3L2*$^{-/-}$ chondrocytes in this independent experimental setup (Fig. 3.6B) as shown before in Fig 3.5. In contrast, *sec23a* overexpression failed to restore collagen secretion in *creb3L2*$^{-/-}$ cells (Fig. 3.6B, bottom panel). These findings suggest that Creb3L2 targets multiple factors other than *sec23a* to regulate collagen secretion.

Creb3L2 directly targets sec24d expression

Sec23 and Sec24 act synergistically and form heterodimers to build inner coat of COPII carriers (Fath et al., 2007; Matsuoka et al., 1998). Our lab has identified two COPII inner coat component paralogs, *sec23a* and *sec24d*, to be essential for efficient collagen secretion in zebrafish chondrocytes *in vivo* (Lang et al., 2006; Sarmah et al., 2010).

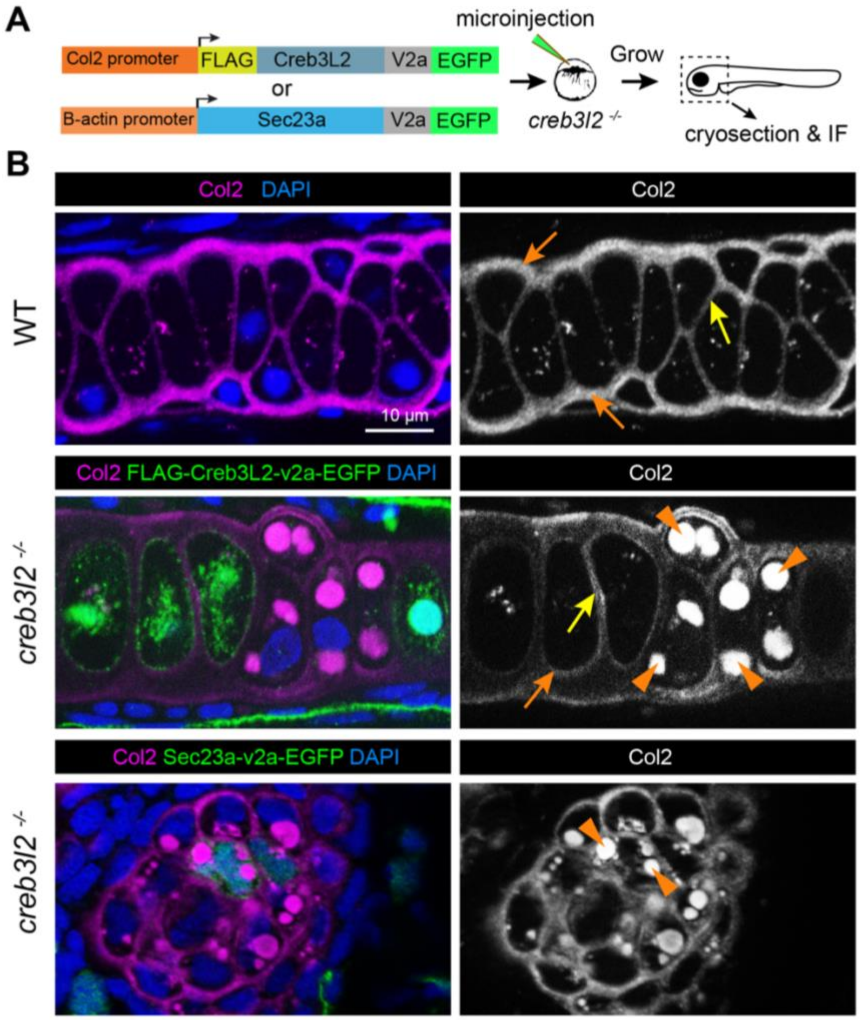


Figure 3.6. Mosaic overexpression of sec23a is not sufficient to restore collagen secretion in *creb3L2*$^{-/-}$ mutants.

A. Experimental design to mosaically overexpress FLAG-Creb3L2 or Sec23a with V2a-EGFP tag and analyze collagen secretion by chondrocytes via IF on cryosections. **B.** IF for Col2 and EGFP on cryosectioned chondrocytes. WT panel shows plain WT cartilage with collagen secretion from chondrocyte to both outer edges of cartilage (orange arrows) and intercellular regions (yellow arrow). Middle panel shows that FLAG-Creb3L2 expressing *creb3L2*$^{-/-}$ (*fel*$^{-/-}$) chondrocytes are devoid of intracellular collagen accumulations and have secreted collagen to ECM, just like WT counterparts (orange and yellow arrows). Neighboring non-transgenic mutant cells still contain collagen accumulations (arrowheads). Overexpression of *sec23a* failed to restore collagen secretion in *creb3L2*$^{-/-}$ chondrocytes (arrowheads, bottom panel).

Functional cooperation between Sec24d and Sec23a led us to propose that expression of these genes could be under similar transcriptional programs in chondrocytes. In particular, we hypothesize that Creb3L2 might activate *sec24d* expression, like its binding partner *sec23a*, in chondrocytes to promote physiological levels of collagen secretion. To test this hypothesis, we first assayed whether *sec24d* expression pattern followed a parallel trend to that of *creb3l2*. Taking advantage of transcription-factor domain-mutant *creb3L2*$^{-/-}$ and *creb3L2*-specific morpholino (MO-2, (Melville et al., 2011)), we found out that expression levels of *sec24d* reduced to its ~50% levels under *creb3L2*-depleted conditions (Fig. 3.7A). In an opposite experimental setup, we overexpressed human *CREB3L2* in cultured HEK293 cells and detected significant upregulation of *SEC24D* expression (Fig. 3.7B). *SEC23A* expression also showed a trend toward upregulation even though it did not reach significance in this setup. In contrast, *SEC23B* expression remained unchanged, suggesting that CREB3L2's effect was specific to *SEC24D* and *SEC23A*. These two experiments demonstrate that *sec24d* and *creb3L2* show parallel patterns, thus strengthening the possibility that Creb3L2 transcriptionally targets *sec24d* and activates its expression.

To directly test binding of Creb3L2 to *sec24d* promoter, we decided to conduct chromatin immunoprecipitation (ChIP) assay and generated a stable zebrafish line expressing FLAG-tagged full-length Creb3L2 (Fig. 3.8A) using the Tol2kit construct which we showed to be functional in rescue experiments (Fig. 3.5). Then, we performed an *in silico* search for CRE sites within 2.5 kb promoter region of zebrafish *sec24d* gene to identify potential Creb3L2 binding sites. This search yielded 4 potential CRE sites throughout the promoter (Fig. 3.8B) and they were ranked on the basis of the frequency score of empirically confirmed sites (Fig. 3.8.C). All CRE sites had relative scores higher

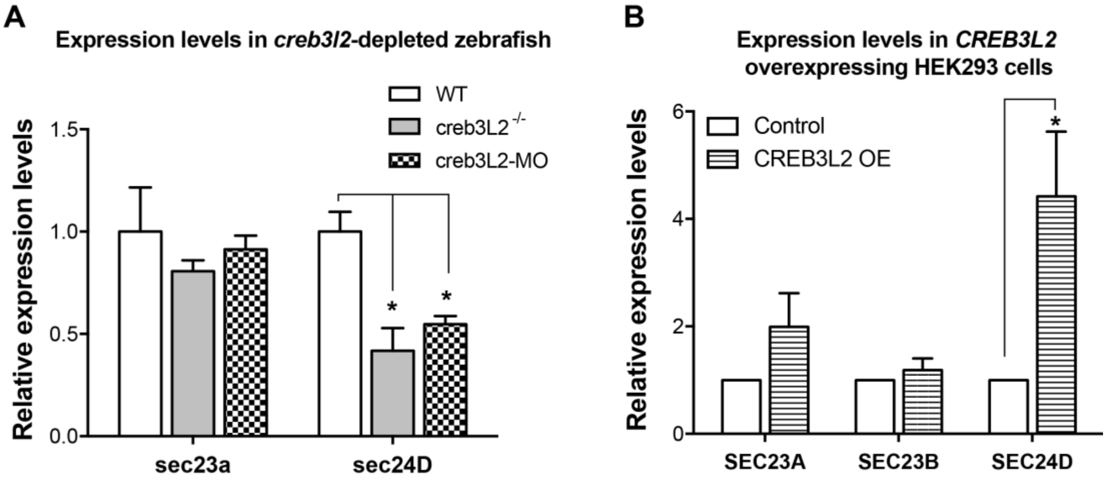
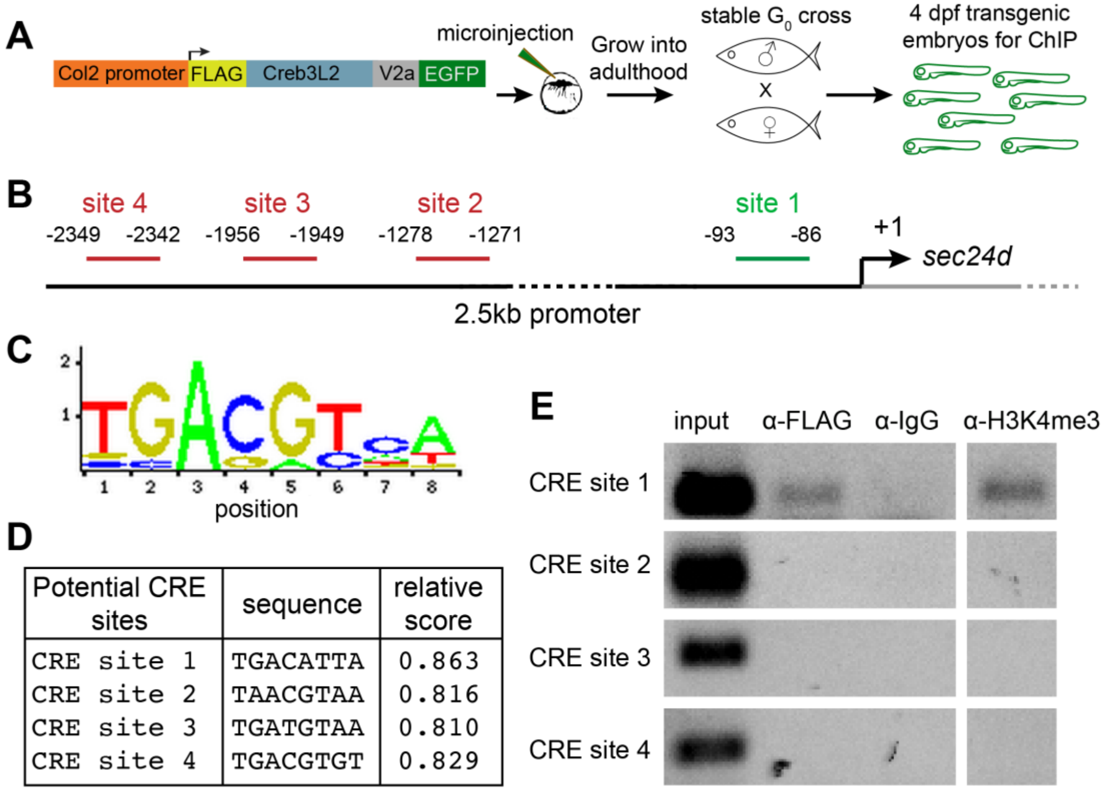


Figure 3.7. ***CREB3L2*** **and** ***SEC24D*** **show parallel expression patterns**

A. Relative expression levels of sec23a and sec24d in 5 dpf zebrafish larvae by qPCR. WT, *creb3L2*$^{-/-}$ (*fel*$^{-/-}$) mutants and creb3L2-morphants are presented. **B.** Relative expression levels of *SEC23A*, *SEC23B* and *SEC24D* in cultured HEK293 cells overexpressing (OE) full-length human *CREB3L2*. Data have been analyzed from 3 independent experiments. *: $p<0.05$.



Potential CRE sites	sequence	relative score
CRE site 1	TGACATTA	0.863
CRE site 2	TAACGTAA	0.816
CRE site 3	TGATGTAA	0.810
CRE site 4	TGACGTGT	0.829

Figure 3.8. Creb3L2 directly binds to sec24d promoter

A. Experimental design for generating stable transgenic zebrafish line overexpressing FLAG-Creb3L2-V2a-EGFP to use in ChIP assay. **B.** CRE sites (tested in this study) within 2.5 kb promoter region of zebrafish *sec24d* gene. **C.** Sequence logo of CREB binding site. Sizes of nucleotides are drawn to scale regarding the base preference at the indicated position, based on empirical data (modified from JASPAR CORE database). **D.** Table displaying potential CRE sites and their predicted relative scores. E. Agarose gel showing ChIP-PCR amplifications for tested CRE sites (1-4) using DNA templates immunoprecipitated with α-FLAG, α-IgG and α-H3K4me3 antibodies. Only CRE site-1 shows positive FLAG-Creb3L2 binding among those tested.

than 0.80 (1.0 is the perfect score), site-1 having the highest (Fig. 3.8D). ChIP assay revealed that site-1 within *sec24d* promoter can indeed be bound by FLAG-Creb3L2 (Fig. 3.8E). The region harboring site-1 was also positive for open chromatin marker, H3K4me3 (Azuara et al., 2006), confirming active transcription through this region. We did not detect FLAG-Creb3L2 binding to either of the other candidate sites, i.e. site-2, -3 and -4. Likewise, these sites were also negative for H3K4me3, suggesting that these sites may be closed for transcriptional activity. Taken together, expression analyses with qPCR and direct binding assay with ChIP suggest that Creb3L2 can directly bind to *sec24d* promoter and activate its expression.

Sec23a and Sec24d act cell-autonomously to facilitate collagen transport

Essential roles of inner COPII coat components, Sec23a and Sec24d, in ER-to-Golgi transport of collagen have been reported investigating human patients (Boyadjiev et al., 2006; Boyadjiev et al., 2011; Fromme et al., 2007) and animal models (Lang et al., 2006; Melville and Knapik, 2011; Ohisa et al., 2010; Sarmah et al., 2010) with skeletal defects. Specifically, zebrafish studies revealed requirement of *sec23a* and *sec24d*-dependent collagen transport during chondrocyte maturation. However, it remains elusive whether their function is cell-autonomously required for physiological levels of collagen secretion.

We have shown that Creb3L2 acts cell-autonomously to activate collagen secretion in chondrocytes. Thus, we expected its targets, Sec23a and Sec24d, to also have cell-autonomous functions in collagen transport machinery. To test this, we designed a mosaic overexpression experiment in zebrafish chondrocytes, taking advantage of self-hydrolyzable V2a-EGFP tag (Kim et al., 2011a) (Fig. 3.9A, C). By microinjecting one-cell stage *sec23a*$^{-/-}$ mutant embryos with transposon based transgenesis construct, we obtained mosaic clones of chondrocytes within embryos at 3.5 dpf stages. Transgenic cells expressing wild-type *sec23a* gene are marked with v2a-EGFP signal. Immunofluorescence analysis revealed that wild-type chondrocytes secrete collagen-II to extracellular space, while *sec23a*$^{-/-}$ mutant chondrocytes retain them in intracellular compartments. Clone of EGFP-positive cells that overexpress wild-type copy in *sec23a*$^{-/-}$ mutant background are able to secrete collagen-II to extracellular space, in a cell-autonomous manner (Fig. 3.9B). This result suggests that Sec23a is cell-autonomously required for collagen transport during chondrocyte maturation. We took a similar approach for Sec24d by overexpressing it in a fused form to V2a-mCherry (Fig. 3.9C). *sec24d*$^{-/-}$ mutant chondrocytes overexpressing 'Sec24d-V2a-mCherry' construct displayed secreted collagen-II with halo-pattern around the cell periphery similar to wild-type counterparts. Nevertheless, neighboring, non-transgenic *sec24d*$^{-/-}$ cells contained intracellular collagen-II deposits (Fig. 3.9D). These results indicate that both inner coat components, Sec23a and Sec24d, are required cell-autonomously for efficient collagen secretion during chondrocyte maturation.

Discussion

Here we present a comprehensive *in vivo* study investigating the functions of Creb3L2 in collagen secretion during chondrocyte maturation. By genetic and cell biological approaches, we showed that both N-terminal TF domain and C-terminal ER-luminal

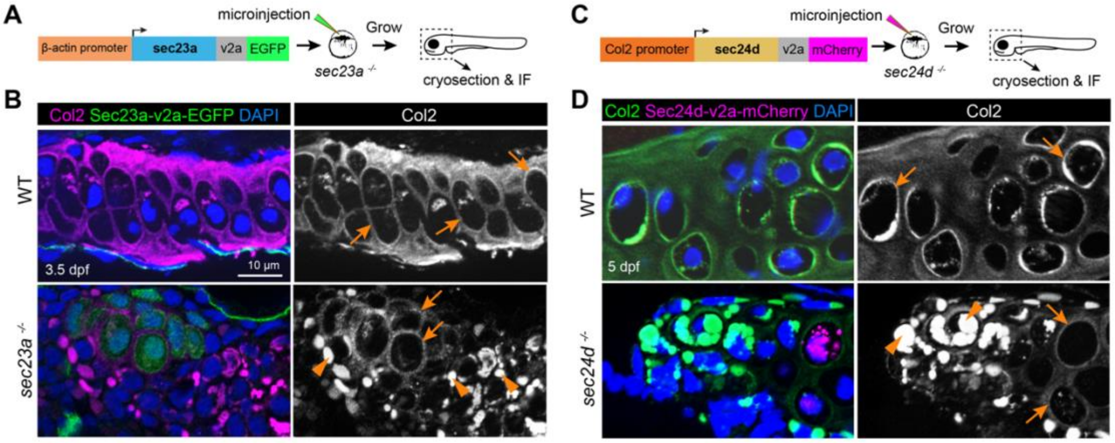


Figure 3.9. Sec23a and Sec24d are cell-autonomously required for collagen secretion in chondrocytes

A. Experimental design to mosaically overexpress Sec23a with V2a-EGFP tag under ubiquitously active β-actin promoter and analyze collagen secretion by chondrocytes via IF on cryosections. **B.** IF for Col2 and EGFP on cryosectioned chondrocytes in *sec23a*$^{-/-}$ WT siblings and homozygous mutants. WT chondrocytes (3.5 dpf) show extracellular collagen around them (arrows), an indication of proper secretion. *sec23a*$^{-/-}$ chondrocytes overexpressing sec23a (marked by EGFP) restore collagen secretion cell-autonomously, as indicated by halo-shaped extracellular collagen pattern around cells (arrows) and lack of intracellular accumulations, which are present in neighboring non-transgenic cells (arrowheads). **C.** Experimental design to mosaically overexpress Sec24d with V2a-mCherry tag under Col2 promoter and analyze collagen secretion by chondrocytes via IF on cryosections. **D.** IF for Col2 and mCherry on cryosectioned chondrocytes in *sec24d*$^{-/-}$ WT siblings and homozygous mutants. 5 dpf chondrocytes show extracellular collagen labeling, with strong pericellular signal around chondrocytes (arrows). *sec24d*$^{-/-}$ mutants contain intracellular collagen accumulations (arrowheads). Sec24d overexpression cell-autonomously restored collagen secretion, indicated by pericellular Col2 staining (arrows) and lack of intracellular accumulations.

domain of Creb3L2 are required to provide for physiological levels of collagen secretion in zebrafish chondrocytes *in vivo*. ChIP and qPCR assays led to the identification of Creb3L2 as a direct transcriptional activator of COPII inner coat component, Sec24d.

It was suggested previously that N-terminal TF activity was sufficient to rescue collagen secretion in cultured mouse *Creb3L2*$^{-/-}$ chondrocytes (Saito et al., 2009a). However, it was not possible to test this in an *in vivo* organismal setting before CRISPR/Cas9 system was adapted to vertebrate species. Here, for the first time, we directly tested whether N-TF activity of Creb3L2 suffices to cope with physiological collagen secretion demand of chondrocytes. Through CRISPR/Cas9 mediated mutations, we generated a zebrafish line (*creb3L2*$^{g12/g12}$) introducing an early termination codon just upstream of transmembrane and C-terminal ER-luminal domains of Creb3L2, but leaving N-TF domain unedited. Contradictory to what was expected, *creb3L2*$^{g12/g12}$ mutant chondrocytes displayed collagen secretion defects, which suggested that N-terminal TF domain itself was not sufficient for collagen secretion from chondrocytes *in vivo*. Stop codon was incorporated upstream of transmembrane domain, so Creb3L2 is predicted to be not incorporated into ER membrane, thus free to translocate to nucleus from cytoplasm, potentially a constitutively active TF. On the contrary, *creb3L2*$^{g12/g12}$ chondrocytes exhibited intracellular collagen accumulations, indicating loss of Creb3L2 activity. A possible explanation could be that truncated Creb3L2 may not fold as efficiently and quickly as WT, which would reduce relative abundance of Creb3L2 at any given time. Another explanation for compromised collagen secretion in *creb3L2*$^{g12/g12}$ mutant chondrocytes could be that trafficking through ER and Golgi stacks has a positive effect on TF activity of Creb3L2. One could speculate that a post-translational modification gained in Golgi or binding of a partner within secretory pathway could enhance TF activity of Creb3L2. Further biochemical and structural studies could address these possibilities directly.

With this study, we established Creb3L2 as a cell-autonomous factor for collagen secretion in chondrocytes, through activation of the multiple target genes, i.e. *sec23a* and *sec24d*. Here we present evidence, for the first time, that Sec24d is a direct target of Creb3L2. Like Creb3L2, its transcriptional activator, Sec24d also acts cell-autonomously to modulate collagen transport. These data lend support to the ‘secretory code model’ (Unlu et al., 2014) which has proposed that stoichiometric availability of specific inner COPII coat component paralogs is ensured through regulatory factors active in tissues where their select cargo is expressed. Creb3L2-Sec23a-Sec24d trio provides the prime example of this model. Creb3L2 (regulatory factor) is active in chondrocytes expressing high amounts of collagen (select cargo) and activates expression of Sec23a and Sec24d (specific inner COPII coat component paralogs) to ensure efficient transport of collagen cargo. Such examples of cargo-specific COPII component paralogs have been reported in literature. For instance, specific requirements of SEC24C for trafficking of serotonin transporter (SERT) (Sucic et al., 2011) and Sec24B for Vangl2 (Merte et al., 2010; Wansleeben et al., 2010; Yang et al., 2013) have been well documented. It is yet to be discovered whether these SEC24 paralogs partner with certain SEC23 paralog to form a cargo-specific inner COPII coat, and whether there are specific regulatory mechanisms controlling availability of these paralogs in target tissues.

Recently, mutations in *SEC24D* was linked to a syndromic form of osteogenesis imperfecta., presenting with multiple fractures with prenatal onset, facial dysmorphism and ossification defects in the skull (Moosa et al., 2015). Craniofacial phenotypes of patients with *SEC24D* mutations are largely shared by CLSD, a disorder caused by *SEC23A* mutations. These clinical findings corroborate animal model studies and collectively suggest that SEC23A-SEC24D inner coat has critical functions during human craniofacial development. Hence, it is plausible to postulate that their transcriptional activator *CREB3L2* is involved in regulation of skeletal and craniofacial development. Future clinical genetics studies will reveal if *CREB3L2* mutations could underlie some of the yet undiagnosed skeletal syndromes.

CHAPTER IV

RIC1 MUTATIONS AND INTEGRATED USE OF PHEWAS IN A BIOBANK REVEAL NOVEL COLLAGEN TRAFFICKING MECHANISM IN COMMON AND RARE DISEASES

Gokhan Unlu[1,2,3], Kinsey Qi[1,2], Amy R. Rushing[1,2], Eric Gamazon[1,2,4,5], David B. Melville[1,6], Nisha Patel[7], Mais Hashem[7], Abdullah AlFaifi[8], Fowzan S. Alkuraya[7], Nancy J. Cox[1,2,4], Ela W. Knapik[1,2,3]

[1]Department of Medicine, Division of Genetic Medicine, [2]Vanderbilt Genetic Institute, Vanderbilt University Medical Center, Nashville, TN 37232, USA; [3]Department of Cell and Developmental Biology, [4]Data Science Institute, Vanderbilt University, Nashville, TN 37232, USA; [5]Clare Hall, University of Cambridge, Cambridge, CB3 9AL, UK; [6]Department of Molecular and Cellular Biology, Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA, 94720, USA, [7]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia, [8]Department of Pediatrics, Security Forces Hospital, Riyadh, Saudi Arabia

Running Title: Genetic models in zebrafish and EHR identify a novel RIC1-linked craniofacial syndrome

Key words: Collagen, RIC1, Rab6, CATIFA, zebrafish

Abstract

Discovery of genotype-phenotype relationships remain a major challenge in clinical medicine and basic biology. Here, we examined three sources of phenotypic data to uncover a novel pathomechanism for rare and common diseases resulting from procollagen secretion deficits. Using a zebrafish phenotype-driven forward genetic screen and positional cloning of the *round* mutations, we identified *ric1* gene to be essential for normal skeletal biology. We show that activation of Rab6a GTPase by the Ric1-Rgp1 GEF complex is required for procollagen secretion through the Trans-Golgi-Network (TGN) en route to extracellular matrix. Using a transcriptome-wide association study (TWAS) in the EHR-linked BioVU biobank, we show genetically predicted reduced expression of *RIC1* is associated with common disease phenome that includes skeletal and dental conditions. Furthermore, we identified families with a rare variant in *RIC1* gene and describe a novel syndrome we named CATIFA (Cataract, cleft lip, Tooth abnormality, Intellectual disability, Facial dysmorphism, ADHD). Collectively, the rare and common diseases corroborate the animal model phenotypic findings. Similarly to zebrafish, CATIFA patient skin fibroblasts accumulate intracellular procollagen in TGN, and the human *RIC1* R1265P variant fails to rescue collagen secretion in zebrafish *round* mutants, unlike wild type *RIC1*, establishing pathogenicity of the variant and evolutionary conservation of this procollagen secretory pathway. The iterative strategy allows to determine contribution of a specific genotype/genome to a spectrum of rare and common disease phenome, and it can be broadly deployed to discover other basic biological functions and disease pathomechanisms.

Introduction

Biomedical research seeks to understand the mechanisms underlying normal biology and what goes awry in disease, with an aim to predict and prevent disease occurrence and to deliver personalized, precision health care (Klein and Gahl, 2018). Despite decades of studies in animal models of individual gene knockouts, Mendelian disease positional cloning, and Genome-Wide Association Studies (GWAS) to unravel complex diseases, progress towards this goal has been slow. The question is whether these three, typically independent, approaches can be integrated to refine discovery process for common and rare disease pathomechanisms and unknown biology.

Collagens are essential components of extracellular matrices and their disrupted production or secretion has been linked in patients and animal models to general skeletal and craniofacial dysmorphology (Arnold and Fertala, 2013; Luderman et al., 2017; Maddirevula et al., 2018; Unlu et al., 2014). Large extracellular matrix (ECM) cargos, such as procollagen, are synthesized in the endoplasmic reticulum (ER) and post-translationally modified and trafficked out to the ECM (Canty and Kadler, 2005).

Zebrafish have been successfully used to understand the consequences of mutations in genes acting at each of these steps (Driever et al., 1996; Knapik, 2000; Latimer and Jessen, 2010). Variants in proteins at ER-to-Golgi secretory pathway, including the coat proteins II (COPII) SEC23A and SEC24D (Lang et al., 2006; Levic et al., 2015; Melville et al., 2011; Sarmah et al., 2010) and procollagen modifying enzymes (Eames et al., 2010; Eames et al., 2011a) cause skeletal and craniofacial defects in zebrafish. Mutations in *SEC23A* lead to cranio-lenticulo-sutural-dysplasia (CLSD), a disease characterized by craniofacial and skeletal defects (Boyadjiev et al., 2006), and mutations in *SEC24D* lead to a syndromic form of osteogenesis imperfecta called Cole-Carpenter syndrome (Garbes et al., 2015). These studies have established zebrafish as a powerful tool to study the genetics and biology of procollagen transport and to model skeletal conditions. Even though ER-to-Golgi transport of procollagen is relatively well-studied in genetic and biochemical models, there is still a longstanding knowledge gap on how procollagen is transported from the Golgi complex to plasma membrane and to the extracellular space (Malhotra et al., 2015; Wakana et al., 2012). Overall, little is known about the factors regulating intra- and post-Golgi transport of procollagen and the medical phenome (Roden, 2017) resulting from their dysfunction.

We used a three-pronged approach to identify novel components of the procollagen secretory pathway in the Golgi complex and their associated medical phenome. We found that Ric1 and its binding partner, Rgp1, are required to activate Rab6a for normal procollagen transport through the TGN. We describe the novel CATIFA developmental syndrome from patients biallelic for the *RIC1* R1265P variant, presenting with multi-organ involvement including craniofacial dysmorphology and abnormal tooth development. We show, using Transcriptome-Wide Association Study (TWAS) that genetically predicted reduced expression of *RIC1* gene is associated with craniodental anomalies in largely common disease BioVU biobank. Our study identifies novel axis for procollagen secretion, a novel genetic syndrome, and an associated common disease phenome with their corresponding disease pathomechanism. Our study shows how integrating animal model such as zebrafish, EHR-based TWAS, and Mendelian rare disease may accelerate discovery of disease pathomechanism, thereby informing diagnosis and ultimately clinical care.

Results

RIC1 is required for normal skeletogenesis

In search for novel components of the procollagen secretory pathway with essential roles in skeletal biology, we have characterized the zebrafish craniofacial *round* (*rnd* $^{m641, m713, m715}$) mutations, which were identified in a large-scale chemical mutagenesis screen(Neuhauss et al., 1996). We genetically mapped and positionally cloned the *round* locus to identify mutations in the *ric1* gene (KIAA1432, ENSDARG00000063362 in Zv9), (Fig. 4.1, Fig. 4.2). By direct sequencing of complementary DNA from the three independent *rnd* alleles, we identified a missense mutation in a conserved residue (R882C) in *rnd*^{m641}; a nonsense mutation (Q64X) in *rnd*^{m713}; and loss of splice donor site between exons 2 and 3, resulting in a frameshift mutation (A84fs*17) in *rnd*^{m715} (Fig. 4.1a). Ric1 is a highly evolutionary conserved protein sharing 71% identity from zebrafish to human. The yeast and human homologs of Ric1 protein and its binding partner Rgp1 have been shown to act as a guanine nucleotide exchange factor (GEF) for Rab6 GTPase (Pusapati et al., 2012; Siniossoglou et al., 2000). However, the role of Ric1-Rgp1-Rab6a in the context of vertebrate development and physiology has not been established.

The zebrafish *ric1*$^{-/-}$ mutations (*rnd* $^{m641, m713, m715}$) present with flattened head, lack of jaw protrusion past the eyes, kinked pectoral fins and short head/body length index at 5 days post-fertilization (dpf) (Fig. 4.1b). The microcephaly-like phenotype was corroborated by smaller head size indices compared to age-matched wild-type (WT) siblings (Fig. 4.1c). Glycosaminoglycan (GAG) staining of sulfated and carboxylated glycoproteins by Alcian blue and Alizarin red revealed that craniofacial cartilage elements and newly formed ossification centers are present in *ric1* mutant embryos, but they are smaller and malformed as compared to WT controls (Fig. 4.1d).

We asked whether microcephaly-like phenotype is a consequence of cell death, but ruled it out by performing a TUNEL assay (data not shown) and a lack of widespread cell death. To characterize the architecture of craniofacial cartilage, we examined fluorescently marked transgenic animals by live, confocal microscopy (Fig. 4.1e**)**. We performed organ shape analysis by examining a specific jaw-supporting element called the hyosymplectic cartilage (HC), due to its easily identifiable key-like shape and good accessibility for *in vivo* imaging. We collected ~80-µm deep Z-stacks encompassing the entire HC from 3 dpf embryos using laser scanning confocal microscopy (Fig. 4.1e**)**, and used Imaris 8 software (Bitplane) to create 3D reconstructions (Fig. 4.1f**)**. The 3D reconstructed images showed irregularities including thicker and misshapen arm and body of the HC, as well as a collapsed foramen, an anatomical passage point (Fig. 4.1g, h). High magnification imaging of the single-cell layered HC arm revealed smaller, constricted mutant chondrocytes, especially at the cell midline, and loss of the characteristic stack-of-coins shape (Fig. 4.1e, f). We measured volumes of five HC chondrocytes at the corresponding positions of age-matched WT and mutants, and found that the average WT volume is ~350 µm^3, while in mutants the volume is reduced to ~185 µm^3 (Fig. 4.1i). We conclude that Ric1 function is required for normal cell shape that further conveys organ shape and, ultimately, the overall craniofacial morphology.

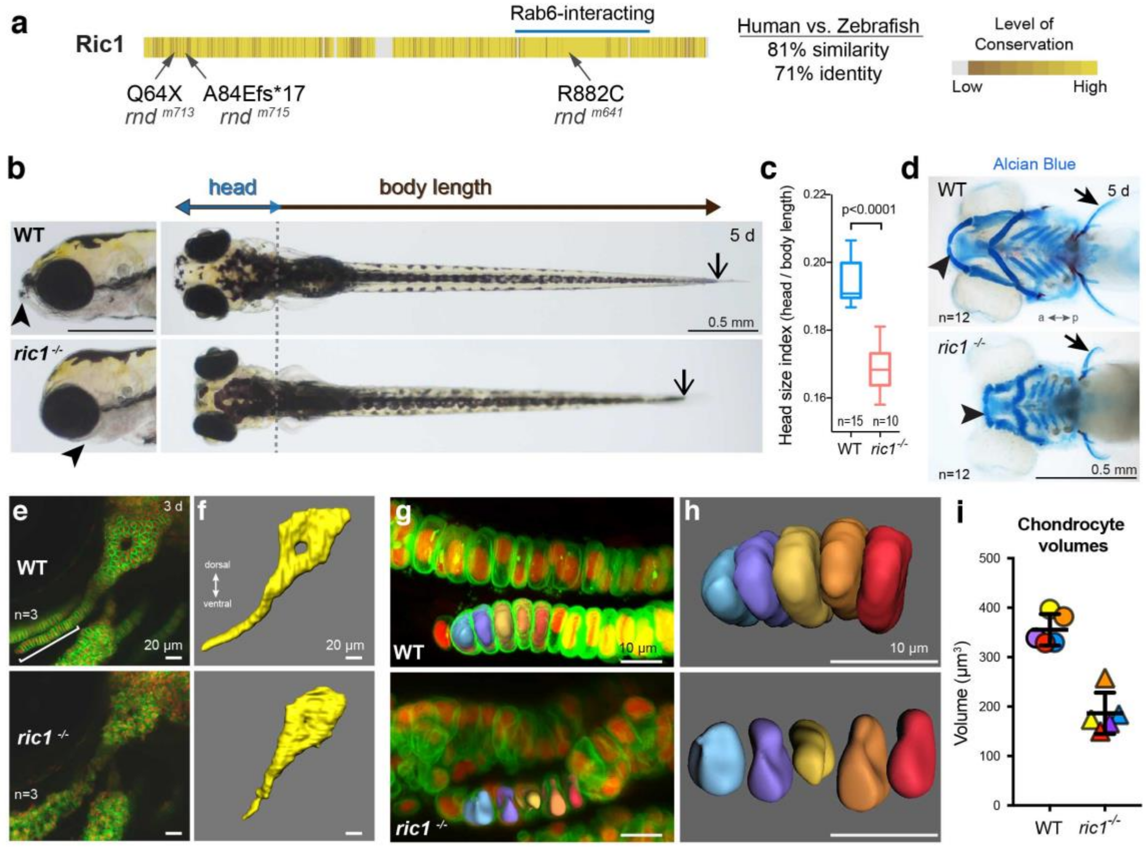


Figure 4.1. Ric1 is required for craniofacial skeleton development and shape. **(a)** Zebrafish Ric1 protein is highly conserved with 81% similarity to human RIC1 (using Clustal Omega, EMBL-EBI). Position of mutations in *rnd* alleles (m641, m713, m715) is indicated with arrows **(b)** Live images of 5 dpf larvae show failure of jaw protrusion in *ric1*$^{-/-}$ mutant embryo (arrowheads on lateral views, left panel) and shortened body length (dorsal views, arrows). **(c)** Box-whisker plot for quantification of head size index (head to total body length ratio) as a measure of microcephaly. Mann-Whitney U test (two-tailed) was used for statistical comparison; 95% confidence interval (CI). **(d)** Alcian blue and Alizarin red staining of the *ric1*$^{-/-}$ mutant shows shorter, malformed craniofacial cartilage elements (arrowheads) and shorter, kinked pectoral fins (arrows). All three alleles exhibit indistinguishable phenotypes and fail to complement one another in genetic complementation assays (data not shown). **(e)** Maximum intensity projections of live z-stack images by confocal microscopy of hyosymplectic cartilage (HC) in WT and *ric1*$^{-/-}$ embryos. Double transgenic Tg(Col2a1a:caax-EGFP; Col2a1a:H2A-mCherry) zebrafish mark plasma membrane (green) and nucleus (red), white bar marks the symplectic arm. **(f)** 3D rendered structure of HC shapes (constructed using Imaris 8) highlights the malformed and short cartilage shape in *ric1*$^{-/-}$. **(g)** Magnified views of symplectic arm chondrocytes. 3D rendered volumes are overlaid onto maximum intensity projection views. **(h)** 3D chondrocyte shapes (Imaris 8) underscore the dysmorphic shape and reduced volume **(i)** of *ric1*$^{-/-}$ chondrocytes. Volume measurements by Imaris 8, color-coded to match the cells in (h).

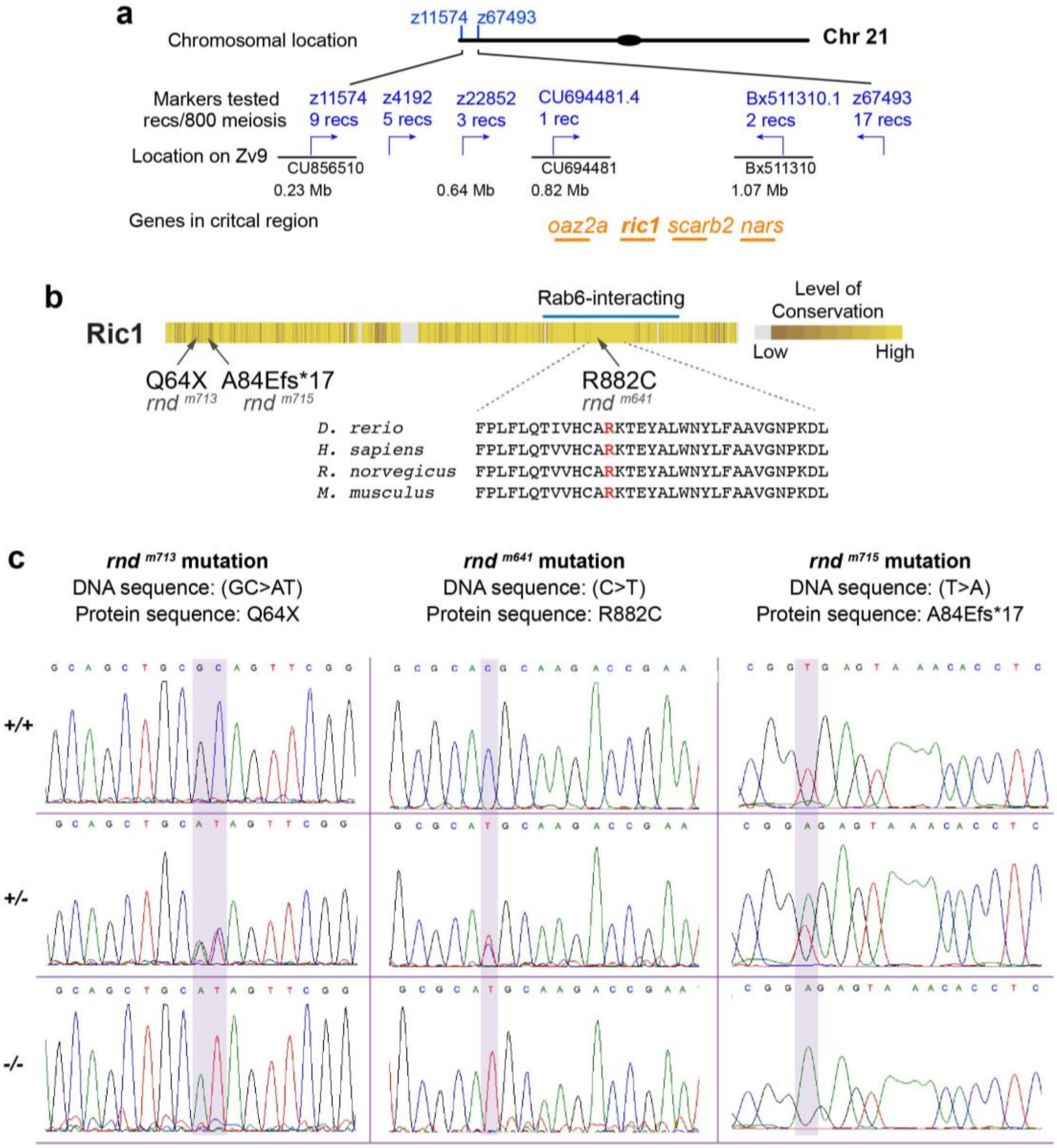


Figure 4.2. Zebrafish *round* mutations map to *ric1* locus.
(a) Genetic linkage analysis and positional cloning map of *rnd* mutation to zebrafish chromosome 21. Microsatellite markers and number of recombination events (recs, blue) in critical region that contains 4 protein-coding genes (orange). **(b)** Sequence conservation analysis (using Clustal Omega, EMBL-EBI) of human and zebrafish Ric1 proteins exhibit 81% similarity, and 71% identity. Positions of mutations detected with direct sequencing of *rnd* alleles (m641, m713, m715) are shown with arrows (electropherograms in Suppl. Fig. 1c). Multiple sequence alignment shows highly conserved R882 residue (red) of the Rab6-interacting domain across vertebrate species analyzed, i.e. zebrafish, rat, mouse and human. **(c)** Electropherograms of direct sequencing results from genomic DNA of homozygous WT (+/+), heterozygous (+/-) and mutant (-/-) embryos for all three alleles, shading highlights mutation site.

PheWAS in BioVU Biobank revealed craniodental and skeletal phenotypes associated with predicted reduced RIC1 expression

To compare the zebrafish *ric1*$^{-/-}$ phenotype to the corresponding human common disease phenome, we examined genetically predicted reduced expression of *RIC1* in the BioVU Electronic Health Records (EHR) and DNA biobank(Denny et al., 2010; Ritchie et al., 2010; Roden et al., 2008) using a TWAS (transcriptome-wide association study) tool, *PrediXcan* (Gamazon et al., 2015). Analysis revealed traits (Phecodes) significantly correlated with genetically determined reduced expression of *RIC1* across multiple organ systems, including skeletal, sensory organs (eye, e.g. corneal opacity), nervous (ADHD), respiratory (asthma), digestive (diverticulosis) and cardiovascular (stricture of artery), (Fig. 4.3a, Table 4.1). The most notable skeletal phecodes included 'fracture of unspecified bones', 'acquired deformities of limbs', and 'disorders of tooth development', specifically, 'tooth eruption and development' (Fig. 4.3b).

The initial phenotypic assessment of the zebrafish *ric1*$^{-/-}$ mutants at earlier developmental stages (5 dpf) revealed craniofacial dysmorphology and deformities of the pectoral fins (homologous to upper extremities). Notably, the TWAS findings drew our attention to zebrafish dentation. To test whether tooth development is compromised in *ric1*$^{-/-}$ mutants, we examined the pharyngeal teeth in older larvae at 7 dpf when they can be easily visualized (Fig. 4.3c, d). Alizarin red staining (calcified tissue) showed the presence of pharyngeal teeth and dermal bones such as cleithrum in wild types, whereas *ric1*$^{-/-}$ larvae lacked pharyngeal teeth, and had smaller calcification domains that failed to elongate to fully formed bone elements (Fig. 4.3c, d). These findings confirmed a highly evolutionarily

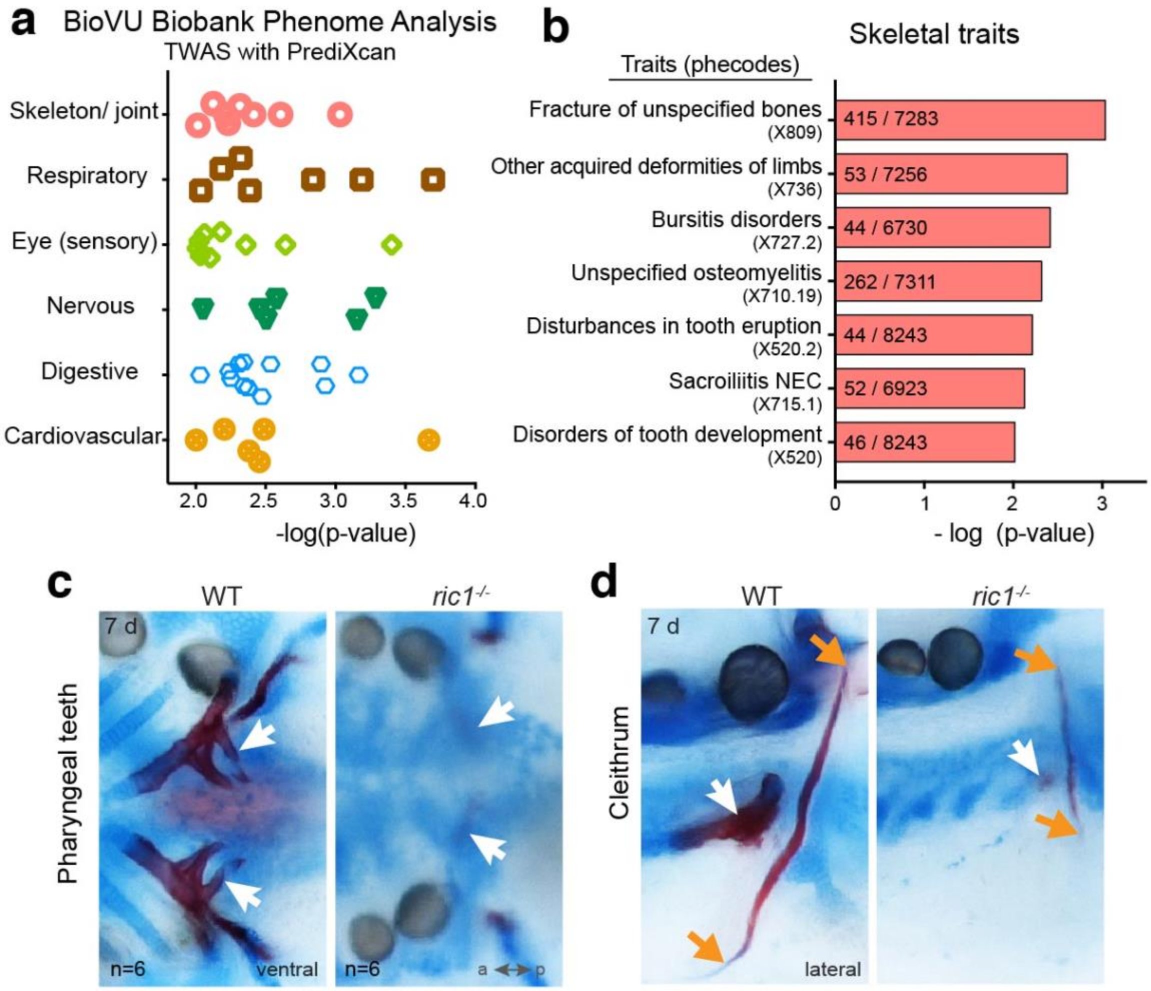


Figure 4.3. Human common disease phenome is significantly associated with genetically predicted reduced *RIC1* expression in BioVU Biobank.
(a) Summary graph of traits (Phecodes) significantly associated with predicted reduced expression of *RIC1* in patients from BioVU biobank, as analyzed by PrediXcan algorithm. Complete dataset is in Table 4.1. Traits are categorized into systems (y-axis), and significance is displayed on x-axis. **(b)** Individual skeletal phecodes from panel (a) are listed on y-axis. 'Number of cases / total cases analyzed' is indicated as insets within bars. x-axis shows significance levels. **(c)** Ventral view of the teeth (arrows) stained by Alcian blue (cartilage) and Alizarin red (ossification), **(d)** lateral view of the pharyngeal teeth (white arrows) on 7th pharyngeal arch. Orange arrows point to lack of cleithrum bone elongation in *ric1*$^{-/-}$ mutant embryo. 7 d: 7 days post-fertilization; a: anterior, p: posterior.

Table 4.1. PrediXcan analysis on BioVU for common disease traits (phecodes) associated with predicted genetically determined reduced expression of *RIC1*

	phecode	beta	p-value	trait	# of cases	# of controls
Skeleton/joint	X809	-0.475223	0.00092842	Fracture of unspecified bones	415	7283
	X736	-34.7626	0.00247298	Other acquired deformities of limbs	53	7256
	X727.2	-4.49326	0.00385264	Bursitis disorders	44	6730
	X710.19	-0.893042	0.00482576	Unspecified osteomyelitis	262	7311
	X858	-0.679979	0.00584572	Complication of internal orthopedic device	251	6685
	X520.2	-1.18055	0.0061075	Disturbances in tooth eruption	44	8243
	X715.1	-2.78064	0.00748377	Sacroiliitis NEC	52	6923
	X520	-1.09126	0.00962552	Disorders of tooth development	46	8243

	phecode	beta	p-value	trait	# of cases	# of controls
Respiratory	X479	-0.758489	0.0002008	Other upper respiratory disease	373	6193
	X465.2	-0.869407	0.00065235	Acute pharyngitis	239	6770
	X519.8	-2.60976	0.00144397	Other diseases of respiratory system, NEC	39	8048
	X495	-1.45524	0.00409617	Asthma	688	6820
	X509.1	-0.489867	0.00476317	Respiratory failure	952	5061
	X513.3	-1.72531	0.00654239	Hypoventilation	101	7861
	X480.3	-1.82948	0.00917754	Pneumonia due to fungus (mycoses)	49	6217

	phecode	beta	p-value	trait	# of cases	# of controls
Eye (sensory)	X364.1	-3.07509	0.00039723	Corneal opacity	32	7724
	X368	-0.780057	0.00228059	Visual disturbances	353	8248
	X374.1	-2.48029	0.00436748	Ectropion or entropion	34	7798
	X362.4	-1.25305	0.00658015	Retinal vascular changes and abnomalities	78	7724
	X367.2	-1.6987	0.00790205	Astigmatism	58	8335
	X362.23	-1.82774	0.00866093	Cystoid macular degeneration of retina	61	7724
	X370	-1.04187	0.00925949	Keratitis	159	7798
	X371.21	-22.982	0.00946209	Allergic conjunctivitis	108	7798
	X362.21	-1.56525	0.00972394	Macular degeneration, dry	71	7724

	phecode	beta	p-value	trait	# of cases	# of controls
Ear (sensory)	X386.9	-0.76795	0.00333873	Dizziness and giddiness (Light-headedness and vertigo)	740	7565
	X381	-0.68503	0.00933308	Otitis media and Eustachian tube disorders	377	8088

	phecode	beta	p-value	trait	# of cases	# of controls
Nervous	X345.11	-2.3214	0.00051637	Generalized convulsive epilepsy	87	6544
	X313.1	-39.9298	0.00070352	Attention deficit hyperactivity disorder	49	8646
	X907	-3.6265	0.00264146	Injuries to the nervous system	79	8942
	X855	-2.53898	0.00312341	Complication of nervous system device, implant, and graft	81	6685
	X313	-29.6269	0.00349694	Pervasive developmental disorders	75	8646
	X317	-1.99595	0.00886056	Alcohol-related disorders	262	7008

	phecode	beta	p-value	trait	# of cases	# of controls
Skin	X691	-3.58012	0.00093336	Congenital anomalies of skin	41	7590

	phecode	beta	p-value	trait	# of cases	# of controls
Digestive	X574.12	-2.53668	0.00068347	Cholelithiasis with other cholecystitis	51	8056
	X562	-0.61594	0.00118303	Diverticulosis and diverticulitis	656	5956
	X532	-1.06687	0.0012721	Dysphagia	896	5130
	X441	-3.80605	0.00291628	Vascular insufficiency of intestine	75	6676
	X260.1	-4.58406	0.00337277	Cachexia	45	6138
	X281.12	-0.895602	0.00422658	Other vitamin B12 deficiency anemia	68	4684
	X559	-1.09202	0.00453522	Ileostomy status	146	5956
	X557	-3.61421	0.00454055	Intestinal malabsorption (non-celiac)	72	5956
	X558	-0.912489	0.00492777	Noninfectious gastroenteritis	328	5956
	X530.12	-2.18981	0.00560592	Ulcer of esophagus	38	5130
	X441.1	-4.94963	0.00576294	Acute vascular insufficiency of intestine	34	6676
	X562.1	-0.525919	0.00928902	Diverticulosis	569	5956

	phecode	beta	p-value	trait	# of cases	# of controls
Cardiovascular & blood	X429.9	-3.70572	0.00021493	Cardiac complications, not elsewhere classified	49	6286
	X818.2	-3.32811	0.00323705	Subarachnoid hemorrhage (injury)	37	8639
	X433.11	-1.36599	0.00350315	Occlusion of cerebral arteries, with cerebral infarction	79	6934
	X285.22	-0.442774	0.00416995	Anemia in neoplastic disease	421	4684
	X447	-0.832855	0.00623695	Other disorders of arteries and arterioles	102	6676
	X447.1	-1.30061	0.00991462	Stricture of artery	38	6676

	phecode	beta	p-value	trait	# of cases	# of controls
Kidney & urogenital tract	X604	-2.50783	0.00042731	Disorders of penis	73	8178
	X624.9	-0.696299	0.00085374	Stress incontinence, female	172	8364
	X189.11	-1.20153	0.00568228	Malignant neoplasm of kidney, except pelvis	233	8503
	X614.5	-0.883838	0.00714291	Inflammatory disease of cervix, vagina, and vulva	142	8581
	X580.2	-1.96667	0.00774004	Nephrotic syndrome without mention of glomerulonephritis	72	5644
	X603.1	-2.90234	0.00808993	Hydrocele	43	8178
	X401.22	-0.702133	0.008471	Hypertensive chronic kidney disease	1016	3058

	phecode	beta	p-value	trait	# of cases	# of controls
Muscular	X292.1	-2.38112	0.00327582	Aphasia/speech disturbance	219	6654
	X770	-1.7722	0.00598723	Myalgia and myositis unspecified	488	8127

	phecode	beta	p-value	trait	# of cases	# of controls
Endocrine	X240	-0.845132	0.00640546	Simple and unspecified goiter	68	7001
	X255	-1.47383	0.00828002	Disorders of adrenal glands	243	7558

	phecode	beta	p-value	trait	# of cases	# of controls
Neoplasm & infection	X204.22	-5.25424	0.00127348	Myeloid leukemia, chronic	48	8016
	X199	-1.22075	0.00161967	Neoplasm of uncertain behavior	570	6274
	X568.1	-0.767425	0.0045824	Peritoneal adhesions (postoperative) (postinfection)	96	7242
	X980	-1.85357	0.00480169	Encounter for long-term (current) use of antibiotics	443	8129
	X110.12	-1.12748	0.00897679	Althete's foot	50	7391

conserved function of *RIC1* in skeletal biology and tooth development. They also revealed a phenotypic spectrum in human patients affected by *RIC1* dysfunction.

CATIFA syndrome is characterized by cataract, cleft lip, tooth abnormality, intellectual disability facial dysmorphism and attention deficit hyperactivity disorder (ADHD)

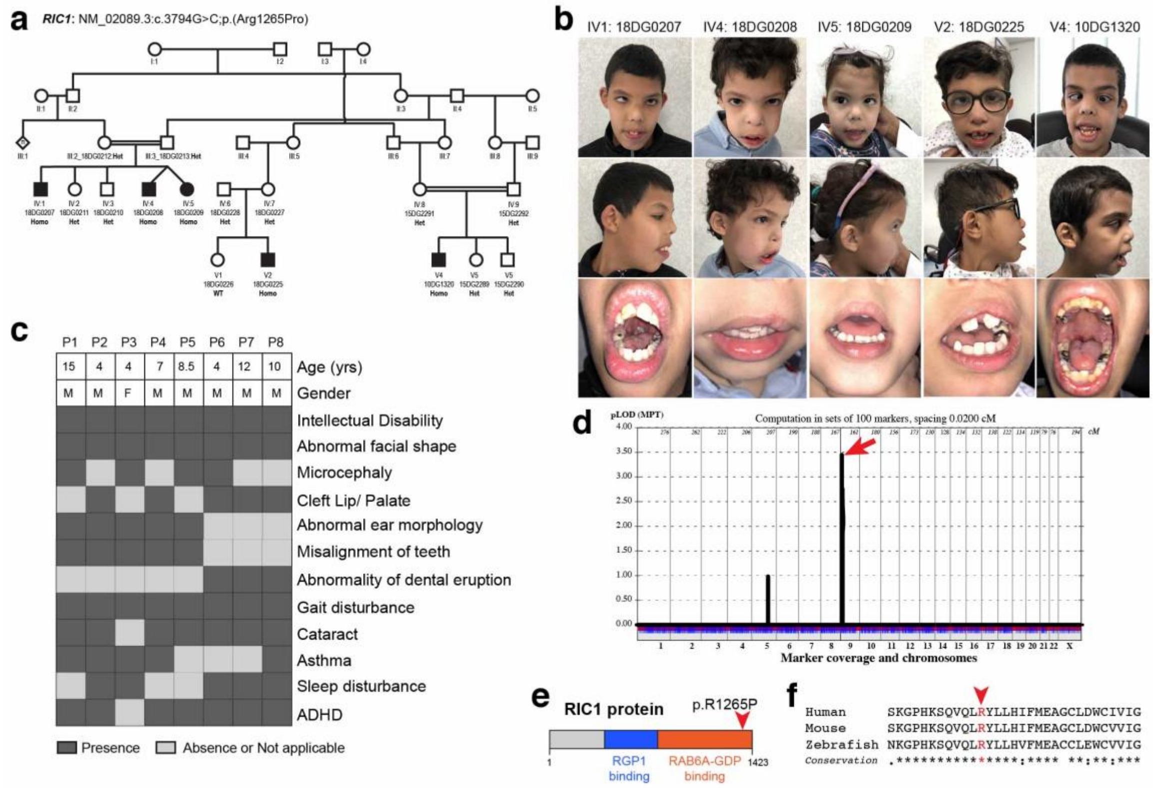
Our prior whole exome sequencing (WES) of the pediatric cataract cohort identified a missense mutation in the RIC1 gene(Patel et al., 2017). With limited evaluation of these patients for other organ system involvement and mounting evidence from the TWAS and the animal model data supporting a complex phenotype of RIC1 deficiency, we pursued reevaluation of the patients.

We identified 8 persons in two extended, consanguineous families (Fig. 4.4a, Fig. 4.5) with a characteristic set of symptoms. All pediatric patients share distinct craniofacial dysmorphology including elongated face, short broad upturned nose with anteverted nares, and long philtrum. Five of the eight children had cleft lip/palate, and all had tooth eruption and/or alignment deficits with extensive caries (Fig. 4.4b, c). Neurological evaluation revealed global developmental delay and intellectual disability, manifesting with motor, speech and cognitive deficits ranging from mild to severe. All males show behavioral abnormalities with mild to severe ADHD. Sensory organ examination revealed visual impairment due to cataract, strabismus, and poor visual tracking (Fig. 4.4b, c). Other symptoms included sleep disturbance (3/8), bronchial asthma (5/8), small ears (5/8), and clumsy, hypotonic gait (8/8). We collectively refer to these features as CATIFA syndrome (Cataract, cleft lip, Tooth abnormality, Intellectual disability, Facial dysmorphism, ADHD). The clinical data of CATIFA patients closely matched the human phenome predicted by TWAS/PrediXcan in BioVU, corroborating the range of medical signs and symptoms associated with *RIC1*.

Using genetic linkage mapping, we linked the CATIFA phenome to a mutation on Chromosome 9 with a supporting LOD score of 3.44 (Fig. 4.4d). Genotyping and Sanger sequencing confirmed a single nucleotide substitution (c.3794C>G) predicted to cause a missense mutation R1265P in the RIC protein (Fig. 4.4e, f) that fully segregated with the phenotype in each of the two families in a strictly autosomal recessive fashion.

Ric1 is an essential component of the procollagen secretory pathway

We sought to identify the basic biological function of Ric1 in vertebrates and the pathomechanisms underlying CATIFA syndrome. Since Ric1 was implicated in cargo traffic through the Golgi complex(Pusapati et al., 2012), and mutations described here affect development of highly secretory chondrocytes, we began by examining cellular localization of procollagen, the predominant secreted cargo in skeletal tissues. By antibody labeling we found that *ric1*-deficient zebrafish chondrocytes retain procollagen II (Col2) intracellularly in large inclusions, while WT cells secrete it across the plasma membrane to the ECM (Fig. 4.6a, b). Quantification showed that normally intracellular collagen occupies <5% of the cytosolic area, whereas in *ric1*$^{-/-}$ chondrocytes it takes up approximately 10 times more surface area (Fig. 4.6c). We found similar results in fibroblast-like, notochord sheath cells (part of the axial hydrostatic skeleton of a vertebrate embryo), where intracellular accumulation of procollagen deposits was accompanied by malformation of notochord basement membrane (Fig. 4.6d, Fig. 4.7). Retention of



P1	P2	P3	P4	P5	P6	P7	P8	
15	4	4	7	8.5	4	12	10	Age (yrs)
M	M	F	M	M	M	M	M	Gender
■	■	■	■	■	■	■	■	Intellectual Disability
■	■	■	■	■	■	■	■	Abnormal facial shape
■	□	■	□	■	■	□	□	Microcephaly
□	■	□	■	□	■	■	■	Cleft Lip/ Palate
■	■	■	■	■	□	□	□	Abnormal ear morphology
■	■	■	■	■	□	□	□	Misalignment of teeth
□	□	□	□	□	■	■	■	Abnormality of dental eruption
■	■	■	■	■	■	■	■	Gait disturbance
■	■	□	■	■	■	■	■	Cataract
■	■	■	■	□	□	□	■	Asthma
□	■	■	□	□	■	■	■	Sleep disturbance
■	■	□	■	■	■	■	■	ADHD

■ Presence □ Absence or Not applicable

Figure 4.4. The R1265P *RIC1* variant segregates in families with a distinct recessive syndrome characterized by Cataract, cleft lip, Tooth abnormality, Intellectual disability Facial dysmorphism and Attention deficit hyperactivity disorder, which we named CATIFA.

(a) Pedigree of a large multiplex, consanguineous family. A double bar represents parental consanguinity. Males are represented by squares, females by circles and affected individuals by shading. Genotypes are listed for tested individuals. Het: Heterozygous, Homo: Homozygous. **(b)** Photographs of patients with CATIFA syndrome. Written consent for the use of photographs was obtained from the parents of affected individuals. **(c)** Human Phenotype Ontology heat map of patients' common clinical features. Blank boxes represent either absent or unreported symptoms. **(d)** Linkage analysis and LOD score value of CATIFA syndrome-linked RIC1 c3794G>C variant on chromosome 9 (arrow). **(e)** Schematic of RIC1 protein marking the R1265 variant site within the Rab6 binding domain. **(f)** RIC1 protein sequence is highly conserved among vertebrates, human, mouse and zebrafish, including the highly conserved arginine (R1265) residue.

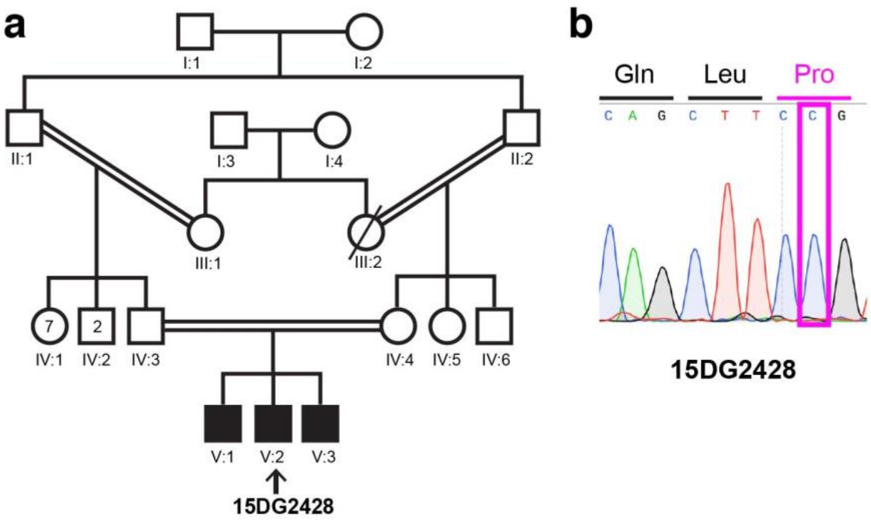


Figure 4.5. Genetic analysis of additional CATIFA patients.
(a) Pedigree of the additional family contributing to this study (modified from Patel et al., 2016, Human Genetics). A double bar represents parental consanguinity. Males are represented by squares, females by circles and affected individuals by shading Arrow points to patient 15DG2428 whose skin biopsy fibroblasts were sequenced **(b),** and analyzed by cellular and molecular methods. Note the mutation site (boxed) resulting in Arg to Pro substitution in the protein sequence.

procollagen in fibroblast-like sheath cells and craniofacial chondrocytes suggests that the secretion defect is not cell-type specific, but may be cargo-selective. To test this hypothesis, we analyzed other cargos such as matrilin, fibronectin, and β-catenin, and found that they all are normally trafficked (Fig. 4.8). We conclude that Ric1 is required for trafficking of procollagen.

To test whether the effect on collagen secretion is cell-autonomous, we used CRISPR/Cas9 genome editing to generate embryos with mosaic loss of *ric1* function. We injected a low concentration of guide RNA (gRNA) to generate *ric1* mutant clones of cells (*ric1*^{gRNA}) in WT background (Fig. 4.9). Ric1-deficient clones of cells recapitulated the *round* phenotype and accumulated intracellular collagen, whereas the neighboring WT cells secreted collagen normally. These data suggest that Ric1 is cell-autonomously required for collagen transport.

To examine the subcellular organization of the collagen inclusions in chondrocytes, we conducted transmission electron microscopy (TEM) analysis and found *ric1*$^{-/-}$ cells contained vesicles of various sizes (Fig. 4.6e; Fig. 4.10; Fig.4.11), while WT cells were devoid of any larger vesicular structures. High magnification images revealed striated material in the core of the large vesicles, reminiscent of collagen fibrils that are normally formed in ECM as procollagen is processed by enzymes that cut N- and C-termini (Fig. 4.6e-e”).

Consistent with collagen secretion defects, cartilage matrix analysis by TEM showed reduced ECM cross-linking in *ric1*$^{-/-}$ as compared to WT at 3 dpf; and by 4 dpf, *ric1*$^{-/-}$ matrix seems devoid of fibrillar cross-links (Fig. 4.6f, Fig. 4.10c, d). Taken together,

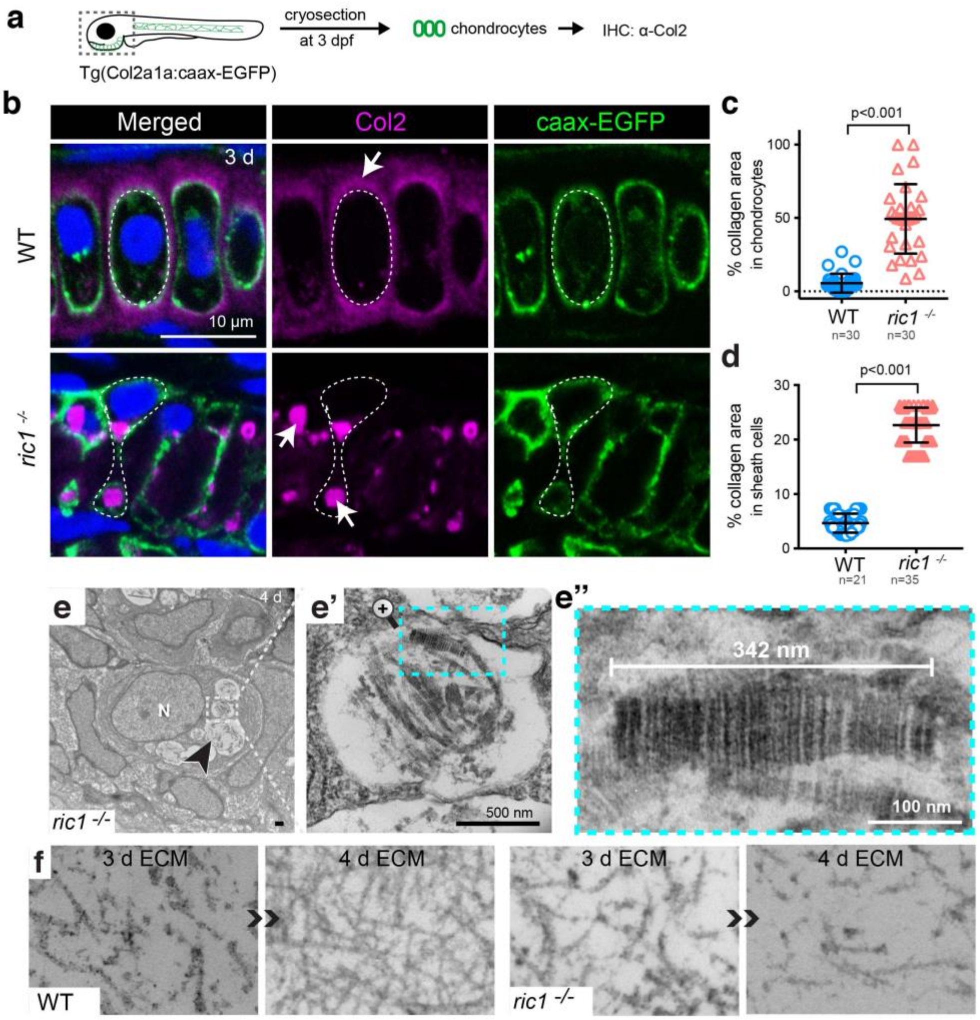


Figure 4.6. Ric1 modulates procollagen transport.
(a) Experimental design for immunohistochemistry (IHC) analysis with Col2 antibody on cryosections (14 µm-thick) in Tg(Col2a1a:caax-EGFP) transgenic zebrafish marking plasma membrane in green. Chondrocytes in the head and notochord sheath cells in the trunk region were analyzed. **(b)** Representative images of WT and *ric1*$^{-/-}$ cartilage stained for collagen-II (Col2, magenta) and EGFP (green). Col2 signal is outside caax-EGFP boundaries (dashed line) in extracellular space, whereas *ric1*$^{-/-}$ chondrocytes show intracellular (inside caax-EGFP boundaries) accumulation (arrows). **(c)** Distribution of collagen accumulation in chondrocytes. Percentage of cytoplasmic area occupied by Col2 signal was measured in each cell with ImageJ. N=3 embryos per group, n=30 cells per group. **(d)** Quantification in notochord sheath cells: N=4 embryos per group, n=21 cells for WT, n=35 cells for *ric1*$^{-/-}$. Data in **c, d,** were analyzed with Mann-Whitney U test, CI= 95%. Mean and SD values are indicated with bars. **(e)** TEM images of 4 dpf *ric1*$^{-/-}$ craniofacial chondrocytes show vacuolar accumulations (arrow), **(e')** higher magnification of boxed area, and **(e")** further magnification showing striated ultrastructure of intracellular collagen fibrils. Control images in **Supplementary Fig. 6**. **(f)** Progression of matrix crosslinking in WT cartilage ECM from 3 dpf to 4 dpf. Note the ECM paucity and deficit in crosslinking at 4 dpf in *ric1*$^{-/-}$ ECM. N: nucleus, ECM: extracellular matrix.

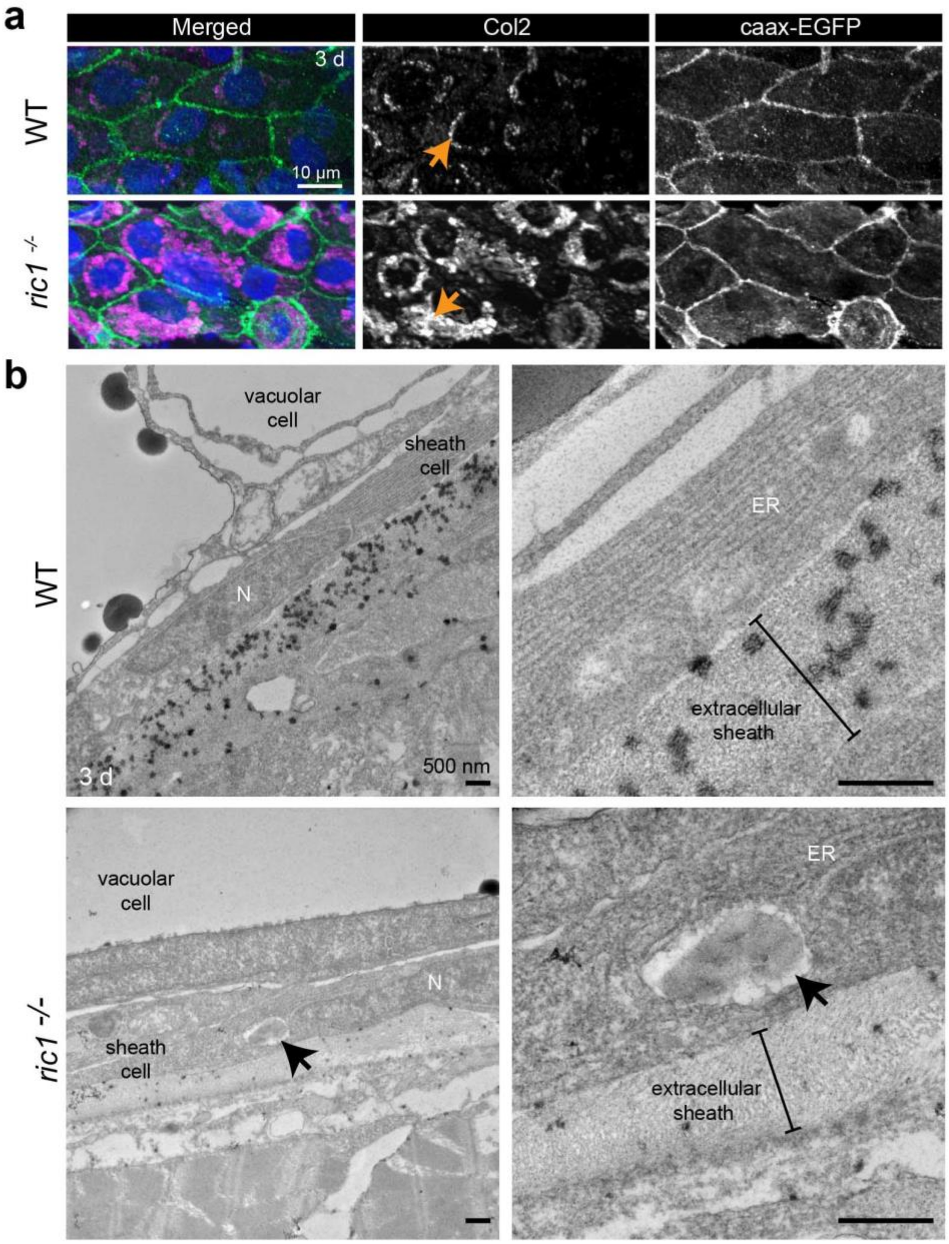


Figure 4.7. Notochord basement membrane (BM) and vacuolar inclusions accumulation in *ric1*$^{-/-}$ notochord sheath cell.

(a) Immunohistochemistry analysis with Col2 antibody on cryosections (14 µm-thick) in Tg(Col2a1a:caax-EGFP) line marking plasma membrane (green) in notochord sheath cells at the trunk level. Representative images of WT and *ric1*$^{-/-}$ notochords stained for Col2 display accumulation of intracellular Col2 in mutant sheath cells (arrows). **(b)** TEM images of notochord tissue at 3 dpf. Magnified views of sheath cells and extracellular sheath / BM are displayed on the right panels. Arrows point to intracellular inclusions, ER: endoplasmic reticulum, N: nucleus. Size bars = 500 nm.

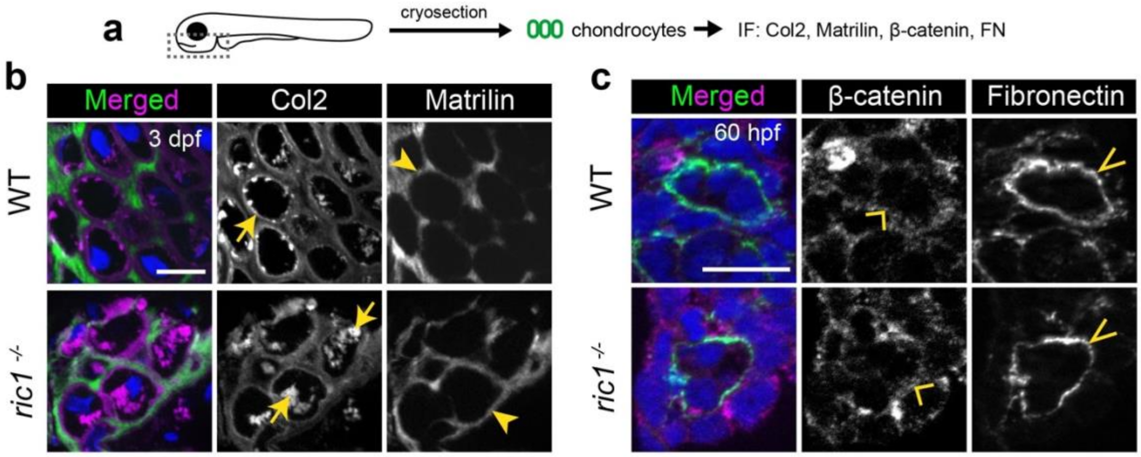
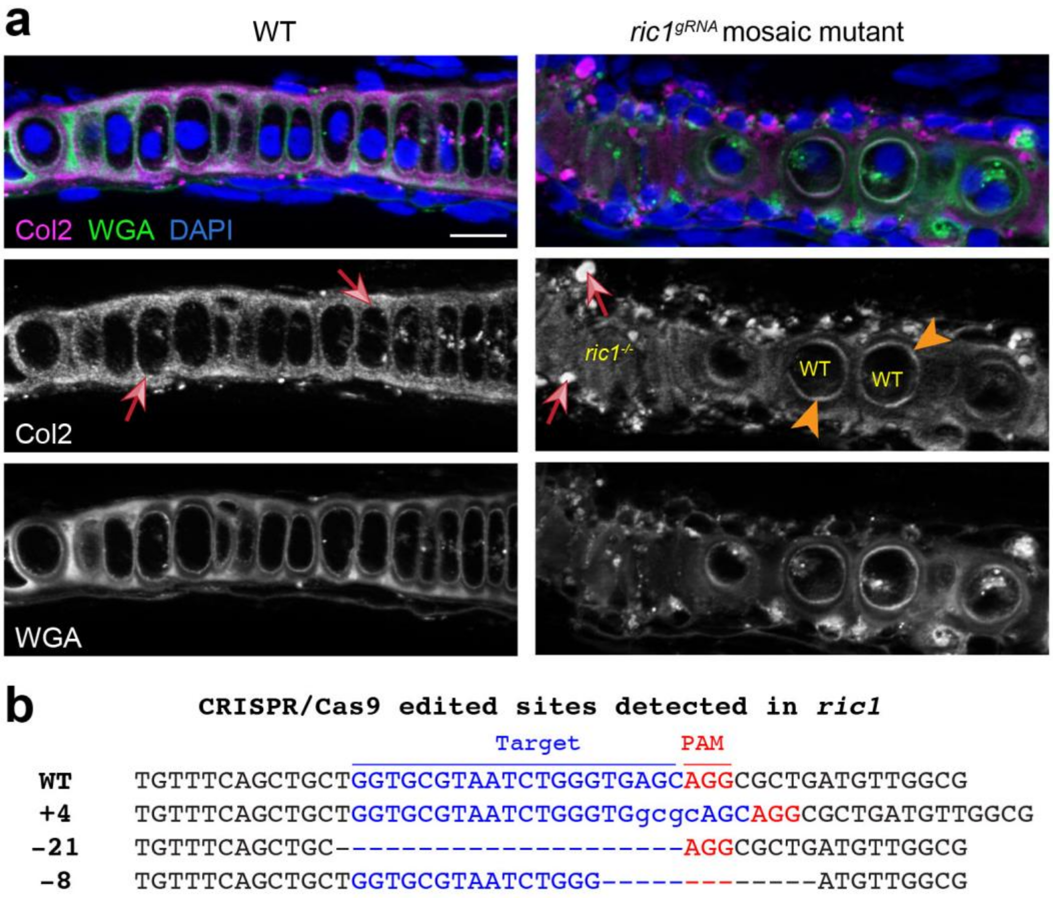


Figure 4.8. Collagen II fails to be transported out of Ric1-deficient chondrocytes, while other ECM cargos are secreted normally.

(a) Experimental design for immunohistochemistry (IHC) analysis with antibodies against Col2, matrilin, β-catenin and fibronectin epitopes on cryosections **(b)** Col2 (arrows) and Matrilin (arrowheads) co-immunostaining in 3 dpf chondrocytes. **(c)** β-catenin (wide arrowhead) and Fibronectin (narrow arrowhead) co-immunostaining at 60 hours post fertilization (hpf) stage. DAPI (blue) marks nuclei.



CRISPR/Cas9 edited sites detected in *ric1*

```
                            Target        PAM
WT    TGTTTCAGCTGCTGGTGCGTAATCTGGGTGAGCAGGCGCTGATGTTGGCG
+4    TGTTTCAGCTGCTGGTGCGTAATCTGGGTGgcgcAGCAGGCGCTGATGTTGGCG
-21   TGTTTCAGCTGC---------------------AGGCGCTGATGTTGGCG
-8    TGTTTCAGCTGCTGGTGCGTAATCTGGG-------------ATGTTGGCG
```

Figure 4.9. CRISPR/Cas9 genome editing-mediated depletion of *ric1* recapitulates collagen secretion defects in a cell-autonomous manner

(a) Co-immunostaining for collagen type-II (Col2) and WGA-labeled glycosylated matrix proteins at 3 dpf zebrafish cartilage. Arrows in *ric1^{gRNA}* mosaic mutant (*ric1$^{-/-}$* chondrocyte clones in the WT embryo injected with gRNA targeting *ric1*) point to collagen accumulations. Arrowheads indicate secreted, extracellular Col2. DAPI (blue) marks nuclei. **(b)** Representative mutations detected in *ric1^{gRNA}* mosaic mutants by direct sequencing. PAM: Protospacer adjacent motif.

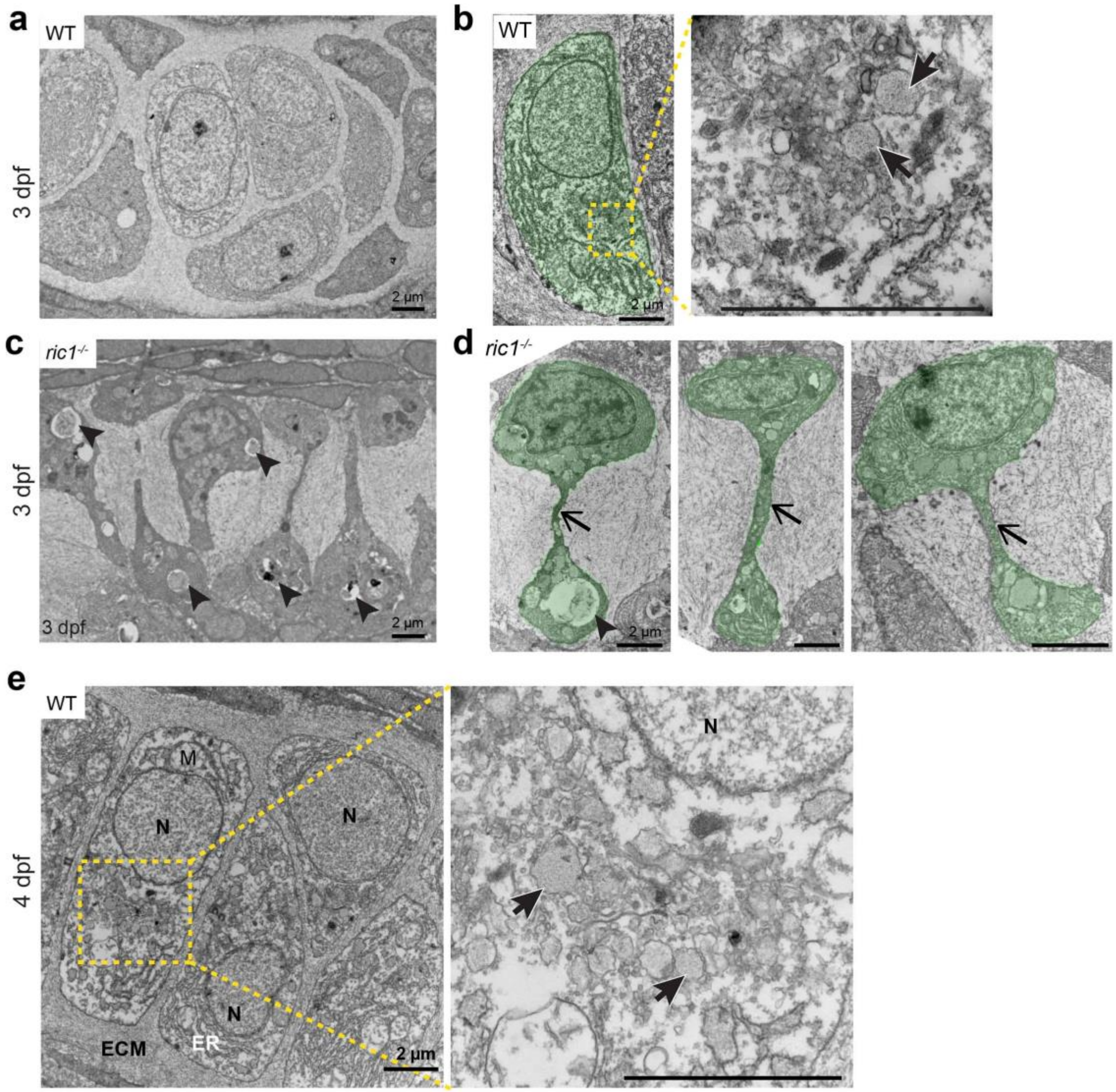
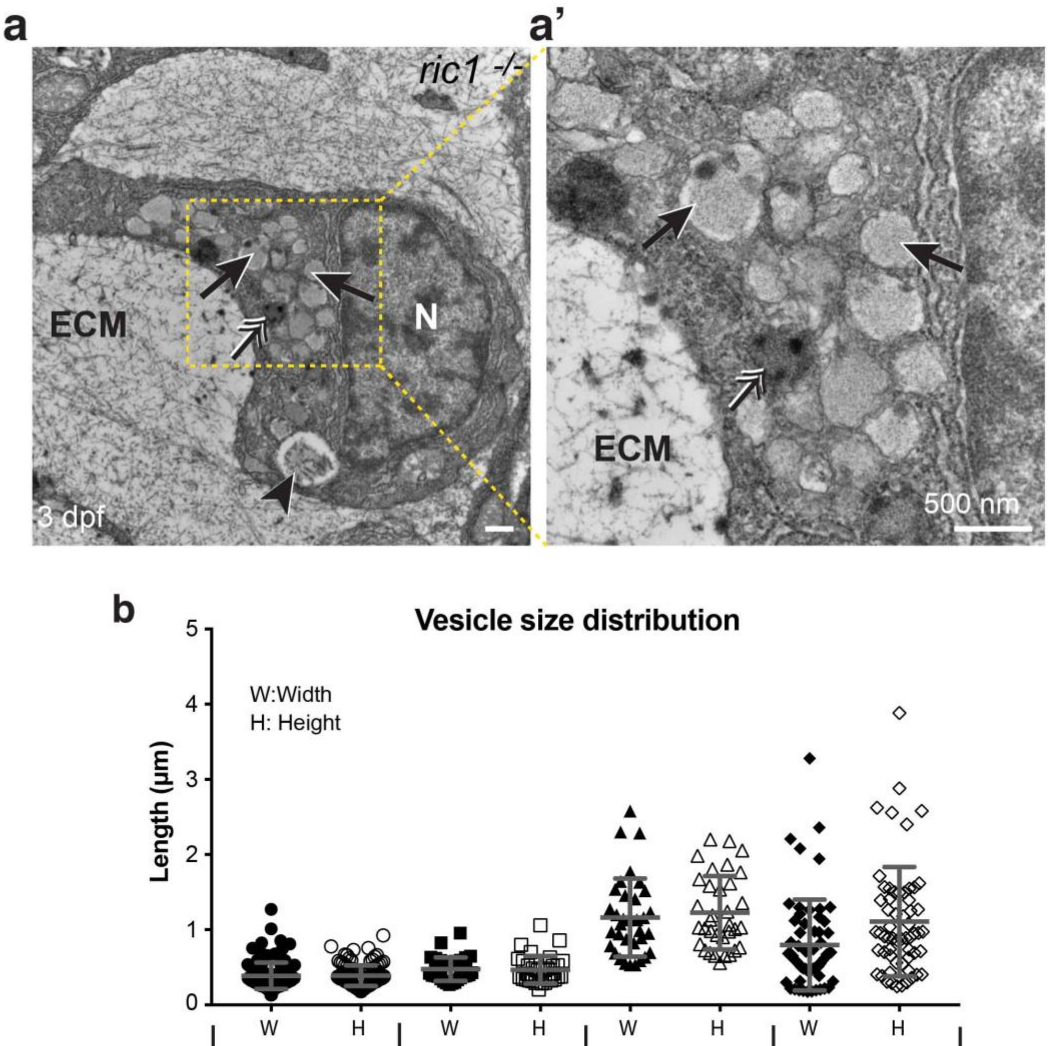


Figure 4.10. Constricted cell shapes of *ric1$^{-/-}$* chondrocytes at 3 dpf and normal ultrastructure of WT cells by TEM at 3 and 4 dpf.

(a) TEM image of WT craniofacial cartilage showing normal chondrocyte shapes. **(b)** Higher magnification reveals vesicular structures associated with normal secretion (arrows). **(c)** *ric1$^{-/-}$* craniofacial cartilage chondrocytes contain large inclusions (arrowheads). **(d)** Representative examples of constricted at the midline chondrocytes (arrows), green overlays on **b** & **d** are drawn after image acquisition to help with cell shape demarcation. **(e)** A representative TEM image of WT chondrocytes at 4 dpf. Magnified view of boxed area is shown on the right, arrows point to normal vesicular compartments. N: nucleus, M: mitochondrion, ER: endoplasmic reticulum, ECM: extracellular matrix.



Type of vesicle	Low density	Medium density	Vacuolar/ High density	Vacuolar / High density
Avg.# of vesicles / cell	5.2	1.4	1.4	5.9
Stage	3 dpf	3 dpf	3 dpf	4 dpf

Figure 4.11. Intracellular vesicles accumulate in *ric1*$^{-/-}$ chondrocytes.
(a) TEM micrograph of representative 3 dpf craniofacial chondrocyte displaying 3 types of vesicles, classified based on their electron densities: low density (arrows), medium density (double headed arrow) and vacuolar/high density (arrowheads). **(a')** Magnified view of the boxed region from a. **(b)** Size distribution and average number of vesicles in 3 and 4 dpf *ric1*$^{-/-}$ chondrocytes. (n≥10 cells/ stage). Vesicle diameters were measured both in parallel (W: width) and perpendicular (H: height) dimension to image plane; and plotted as length (µm) values. N: nucleus, ECM: extracellular matrix. Mean and SD values are indicated with bars.

TEM data corroborate antibody-labeling results of intracellular retention of collagen in *ric1*$^{-/-}$ cells and compromised ECM ultrastructure.

Ric1-Rgp1 GEF complex regulates procollagen transport by activating Rab6a

Prior work in vitro has shown that Ric1-Rgp1 proteins interact and activate Rab6a by facilitating GTP loading (Fig. 4.12a). This model predicts first, that depletion of Rgp1 would result in a deficit similar to the *ric1* loss-of-function phenotype, and, second, that overexpression of constitutively active Rab6a would bypass the Ric1-Rgp1 requirement and rescue procollagen traffic in *ric1*$^{-/-}$ mutant embryos.

To test the first prediction, we used CRISPR/Cas9 genome editing to deplete *rgp1* in zebrafish embryos (Fig. 4.13). By injecting a low concentration of *rgp1*-targeting gRNA, we obtained mosaic mutant embryos that contained clones of *rgp1*-mutant cells in WT background (referred to as *rgp1*^{gRNA} hereafter). Mosaic *rgp1*^{gRNA} embryos resembled *ric1*$^{-/-}$ mutants in their skeletal phenotypes, such as flattened face, lack of protruding jaw, malformed craniofacial cartilage elements and shortened body length (Fig. 4.13). These structural changes were corroborated by Col2 antibody and wheat germ agglutinin (WGA, a lectin that binds to glucosaminoglycans enriched in cartilage ECM) staining, revealing that *rgp1*$^{-/-}$ clones contained intracellular collagen accumulation similar to that in *ric1*$^{-/-}$ chondrocytes (Fig. 4.12b). These results support the hypothesis that Rgp1 and Ric1 act together to regulate procollagen secretion and skeletal morphogenesis.

To test the second prediction that Ric1-Rgp1 GEF activation of Rab6a regulates procollagen secretion, we overexpressed constitutively active, GTP-bound, Rab6a (Q72L) (Martinez et al., 1994) or WT Rab6a in notochord sheath cells in both WT and *ric1*$^{-/-}$

embryos (Fig. 4.12c). Specifically, we tagged the N-terminus of zebrafish Rab6a with EGFP (White et al., 1999) and mosaically overexpressed it under *col2a1a* promoter using Tol2kit (Kwan et al., 2007) transposon mediated transgenesis (Fig. 4.12d). Immunofluorescence analysis of mosaic, transgenic embryos revealed that *ric1*$^{-/-}$ cells overexpressing constitutively active Rab6a (Q72L) cleared collagen accumulation to WT levels in a cell-autonomous fashion (Fig. 4.12e). Collagen in rescued cells appeared as small punctae throughout the cytoplasm, in a similar pattern as seen in WT cells. However, neighboring, non-transgenic, EGFP-negative *ric1*$^{-/-}$ cells contained large deposits of intracellular collagen. Overexpression of WT EGFP-Rab6a did not rescue collagen secretion, suggesting that GTP-bound Rab6a can bypass the requirement for the GEF, but mere overexpression (OE) of Rab6a is not sufficient to rescue collagen transport.

Pathogenic RIC1 variant leads to collagen accumulation in patient fibroblasts

To confirm that the *RIC1* R1265P variant is pathogenic and interferes with collagen transport, we analyzed fibroblasts isolated from skin biopsy of an affected patient. By TEM, we found distended vesicular compartments, similar to those observed in *ric1*$^{-/-}$ zebrafish mutants that were notably absent in control skin fibroblasts (Fig. 4.14a). To identify the cargo accumulating in patient fibroblasts, we labeled cells with collagen-I antibody (Col1, predominant fibrillar collagen in dermal fibroblasts). We found significantly expanded Col1 staining throughout the patient fibroblasts as compared to control cells (Fig. 4.14a, Fig. 4.15). Rab6a is known to act at various steps of secretory and exocytic pathways (Del Nery et al., 2006; Miserey-Lenkei et al., 2010; Valente et al., 2010). To identify the intracellular compartment in which RAB6a GEF (RIC1-RGP1)

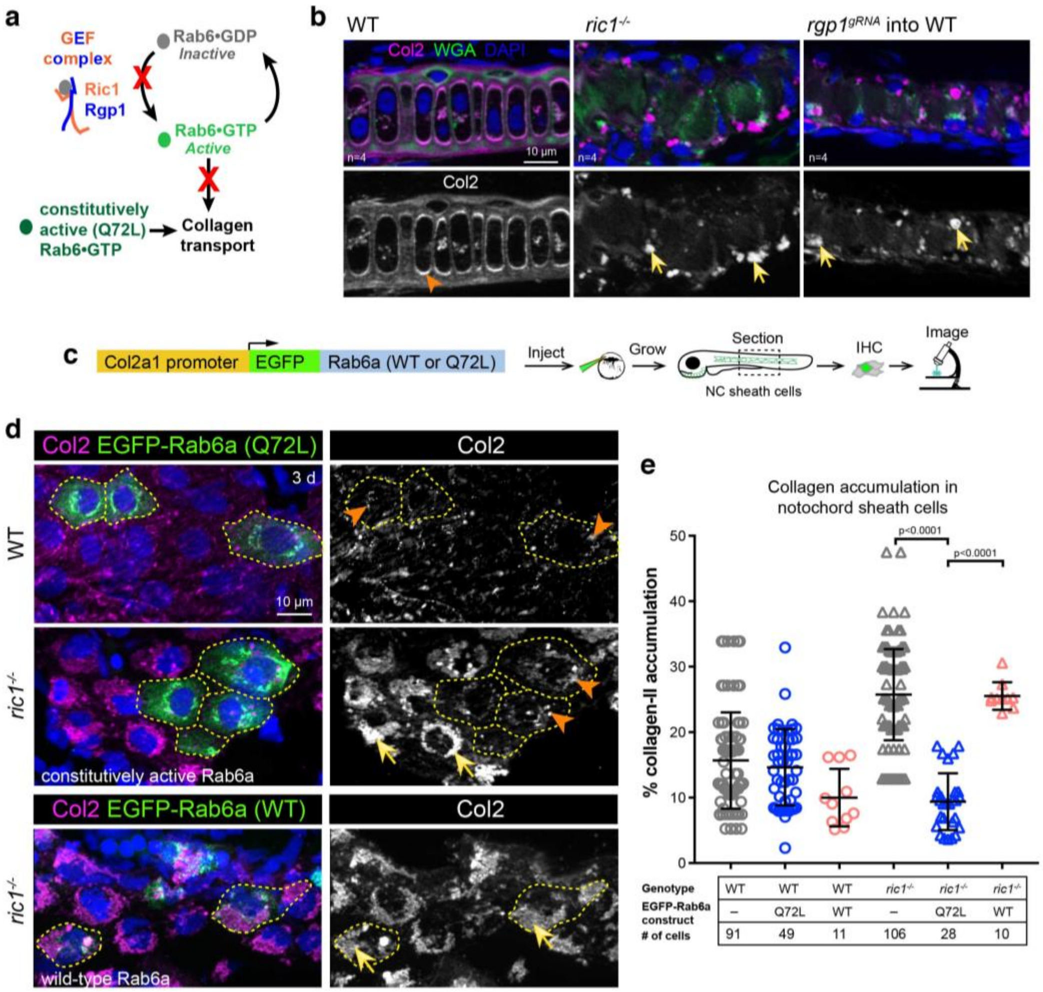


Figure 4.12. Ric1-Rgp1 GEF complex regulates procollagen transport via Rab6a activation.

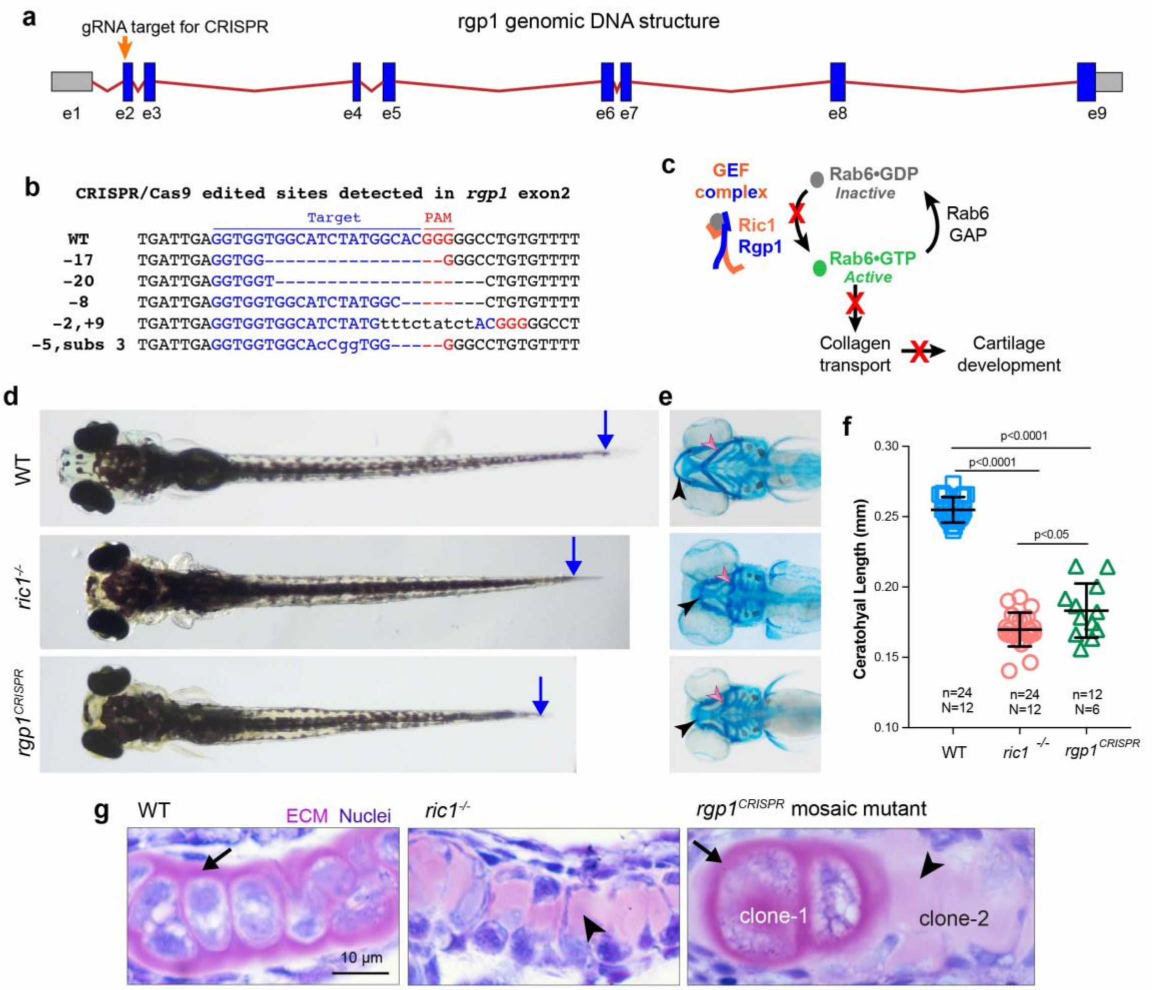
(a) Model of Rab6a activation by the Ric1-Rgp1 GEF complex to regulate collagen transport, and constitutively active Rab6a bypassing GEF requirement for collagen transport. **(b)** IF labeling with Col2 antibody and wheat germ agglutinin (WGA) in 3 dpf zebrafish cartilage shows disrupted cell shape and tissue organization in *ric1*-mutants and *rgp1*$^{\mathrm{gRNA}}$ (*rgp1*-guide RNA for CRISPR-Cas9 genome editing) injected WT embryos. Yellow arrows point to intracellular collagen accumulation, arrowhead points to extracellular Col2. **(c)** Experimental design for mosaic overexpression of EGFP-Rab6a (WT form or Q72L constitutively active mutant) fusion protein in zebrafish notochord sheath cells. **(d)** Col2 and EGFP co-immunostaining of sheath cells over-expressing constitutively active Rab6a (Q72L) mutant and wild type forms (bottom panel). Arrows point to intracellular accumulations; arrowheads mark intracellular collagen in WT and rescued cells. **(e)** Quantification of intracellular collagen accumulation within the cytoplasm, represented as percentage of Col2-stained area over cell surface by measuring areas from maximum intensity projections using ImageJ ‘Measure’ tool. Overexpression of constitutively active Rab6a (Q72L) rescues intracellular collagen accumulation to WT levels (CI=95%, Mann-Whitney U test); whereas WT Rab6a does not. Mean and SD values are indicated with bars.

is required in procollagen transport, we tested co-localization of col1 with compartment-specific markers. The Trans Golgi Network (TGN) markers p230 and Golgin-97 partially co-localized with Col1 staining (Fig. 4.14b-d, Fig. 4.15).

These data indicate that procollagen transport is blocked at the TGN, placing the RIC1-RGP1 GEF activity in the anterograde secretory pathway for procollagen secretion.

Highly evolutionarily conserved zebrafish and human *RIC1* genes are 71% identical and 81% similar at the sequence level. To test whether human RIC1 function is also evolutionarily conserved in regulating collagen transport, we designed a genetic replacement experiment in zebrafish, where we overexpressed human RIC1 (hRIC1) fused to EGFP by a self-hydrolyzable viral v2A tag(Kim et al., 2011a). The Tol2kit-based mosaic expression system allowed us to express hRIC1 in isolated clones of cells in either WT or *ric1*$^{-/-}$ background (Fig. 4.14e). Cleavage of the v2A connector detaches hRIC1 from EGFP, thereby labeling transgenic cells with diffused intracellular EGFP. We found that OE of the wild type, human RIC1 in zebrafish embryos rescued intracellular collagen accumulation in *ric1*$^{-/-}$ cells in a cell-autonomous manner, while it had no effect on collagen expression in WT fish (Fig. 4.14f, Fig. 4.16).

To determine pathogenicity of the RIC1 R1265P variant, we tested whether this variant can also rescue collagen traffic in zebrafish *ric1*$^{-/-}$ mutants. We introduced the R1265P variant into hRIC1-v2A-EGFP construct by site-directed mutagenesis and assayed for collagen deposits. Mosaic OE of the R1265P variant failed to fully restore collagen secretion in Ric1-deficient zebrafish notochord sheath cells, unlike the WT hRIC1



```
CRISPR/Cas9 edited sites detected in rgp1 exon2
                           Target          PAM
WT          TGATTGAGGTGGTGGCATCTATGGCACGGGGGCCTGTGTTTT
-17         TGATTGAGGTGG-----------------GGGCCTGTGTTTT
-20         TGATTGAGGTGGT-------------------CTGTGTTTT
-8          TGATTGAGGTGGTGGCATCTATGGC--------CTGTGTTTT
-2,+9       TGATTGAGGTGGTGGCATCTATGtttctatctACGGGGGCCT
-5,subs 3   TGATTGAGGTGGTGGCAcCggTGG-----GGGCCTGTGTTTT
```

Figure 4.13. Rgp1-depletion recapitulates *ric1$^{-/-}$* mutant phenotype.

(a) Schematic for genomic structure of *rgp1* gene. Exons are shown as boxes, protein coding in blue and untranslated in gray, introns as red lines. Orange arrow marks guide RNA target site within exon 2. **(b)** Deleterious mutations detected in mosaic *rgp1^{CRISPR}* embryos by direct sequencing. **(c)** Proposed model for Ric1-Rgp1 dependent activation of Rab6a on consequent collagen secretion, which modulates cartilage development. **(d)** Live images of 5 dpf WT, *ric1$^{-/-}$* (*rnd^{m713}*) and *rgp1^{CRISPR}* (mosaic mutant embryo generated with CRISPR/Cas9 genome editing). Blue arrows on dorsal views point to body length differences. **(e)**. Alcian blue staining of craniofacial head skeletons at 5 dpf, black arrowheads point to Meckel's cartilage protrusion and pink arrowheads to ceratohyal cartilage element. **(f)** Graph for ceratohyal length measurements shows a range of lengths, likely due to mosaic nature of *rgp1^{CRISPR}* mutants. Both left and right elements are plotted. N: number of embryos, n: number of cartilage elements analyzed. (Student's t-test, CI= 95%) **(g)** Histological analysis of zebrafish cartilage at 3 dpf with Toluidine blue staining. Arrows point to the secreted ECM surrounding WT chondrocytes, arrowheads to diffused matrix staining around mutant cells. Clone-1 in *rgp1^{CRISPR}* mosaic mutant is inferred to be composed of un-edited WT cells, while clone-2 is inferred to be of *rgp1*-mutant chondrocytes due to phenotypic resemblance to *ric1$^{-/-}$* cells in the middle panel.

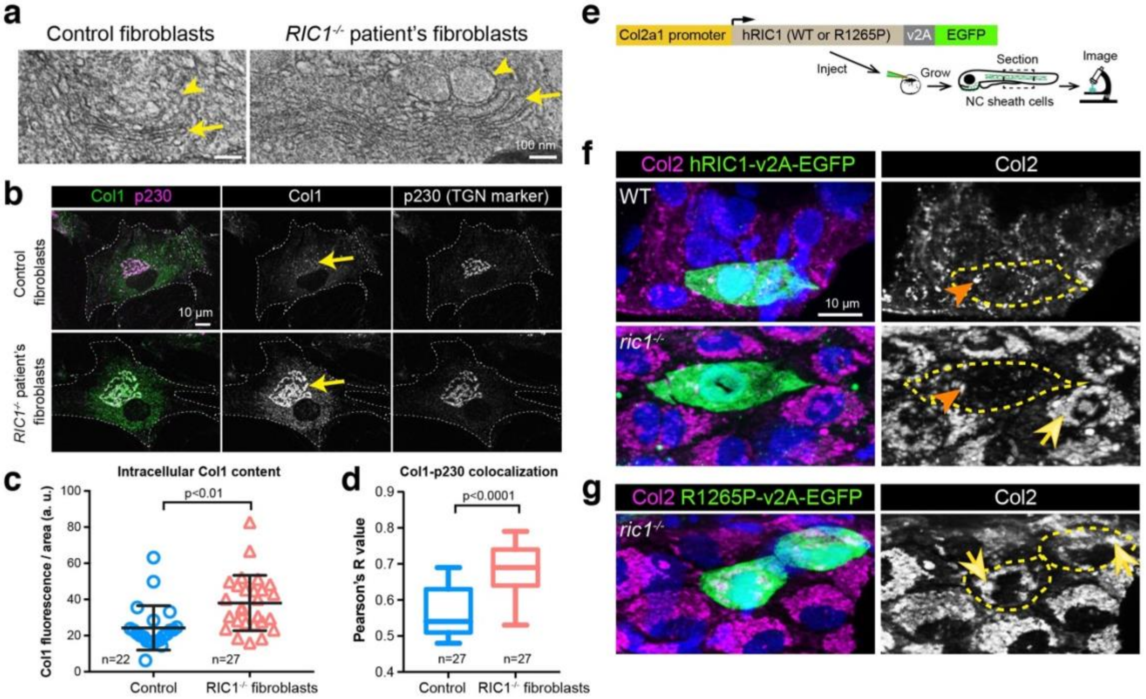


Figure 4.14. Pathogenic *RIC1* variant leads to collagen accumulation in CATIFA fibroblasts.

(a) TEM images of control BJ fibroblasts and *RIC1*$^{-/-}$ patient's dermal fibroblasts show Golgi complex stacks (arrows). Arrowhead points to enlarged Golgi lumen and post-Golgi structures in patient fibroblasts. **(b)** Collagen 1 (Col1) immunofluorescence shows that intracellular collagen in *RIC1*$^{-/-}$ cells co-localizes with the TGN marker p230 (arrow). **(c,d)** Quantification of the intracellular Col1 content, and Col1 co-localization with p230 based on images in (b). (Student's t-test, CI= 95%). **(e)** Experimental design to mosaically overexpress human *RIC1* (hRIC1) gene in zebrafish. EGFP is linked to hRIC1 via self-dissociating viral 2A (v2A) peptide marking *hRIC1* expressing cells. **(f)** Antibody labeling for Col2 and EGFP in notochord sheath cells shows collagen deposits (arrows). hRIC1 overexpressing cell (demarcated by dashed line) has fewer deposits (arrowhead). **(g)** Mosaic overexpression of CATIFA-linked R1265P variant of hRIC1 only partially clears collagen deposits (arrows) while neighboring cells continue to retain Col2 and serve as a *ric1*$^{-/-}$ controls.

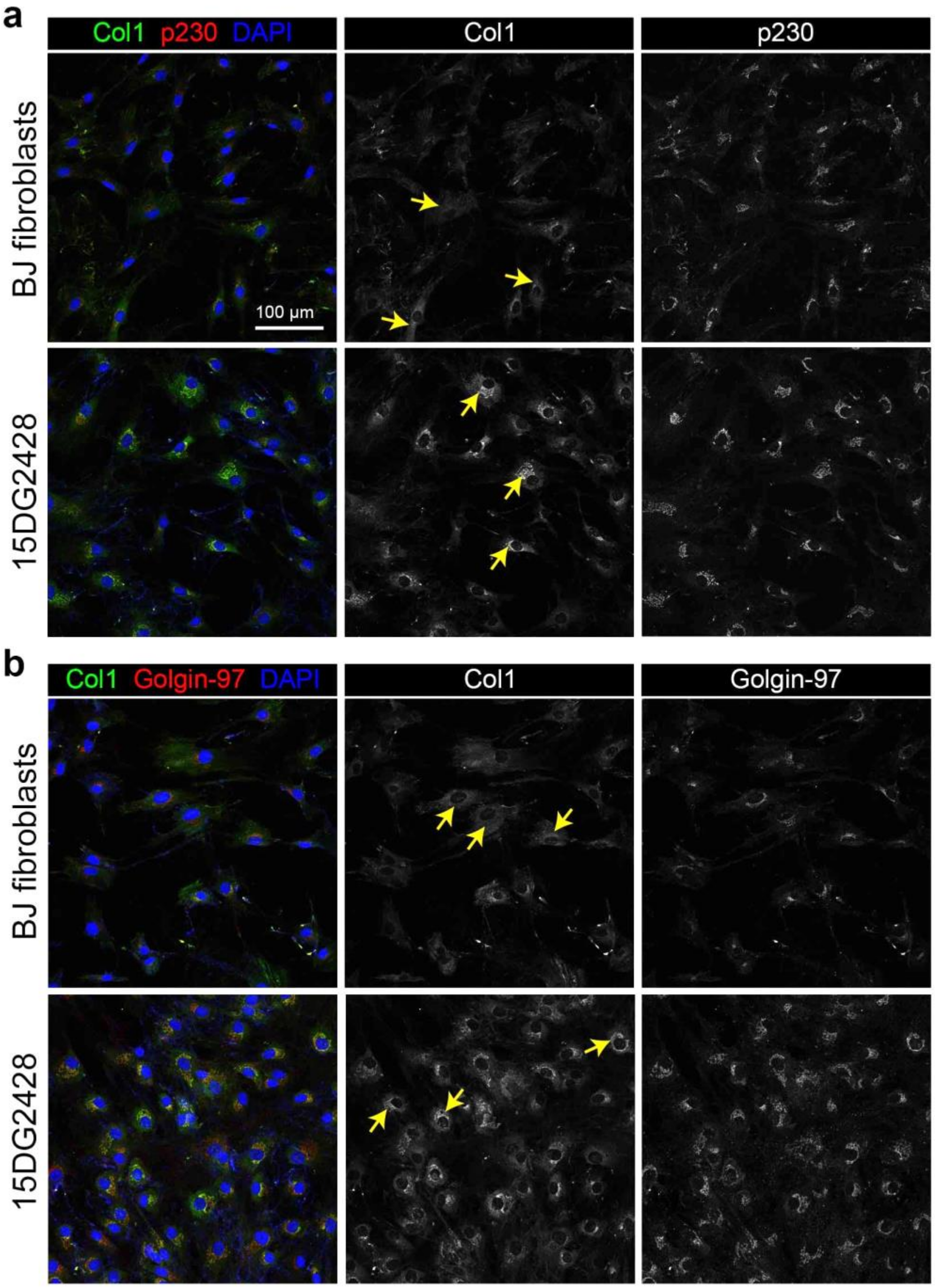


Figure 4.15. Collagen-I accumulates in TGN-associated compartments in *RIC1*$^{-/-}$ patient's fibroblasts.

(a) Representative images of co-IF of BJ fibroblasts (control) and 15DG2428 (*RIC1*$^{-/-}$ patient's dermal fibroblasts) using antibodies against collagen-I and p230 (TGN marker). Confocal images at low magnification (20x/0.80 Plan-Apochromat, WD=0.55mm) are presented. **(b)** Representative low-magnification images of co-IF with Col1 and alternative TGN marker Golgin-97 taken under the same conditions. Arrows point to TGN-associated Col1 signal.

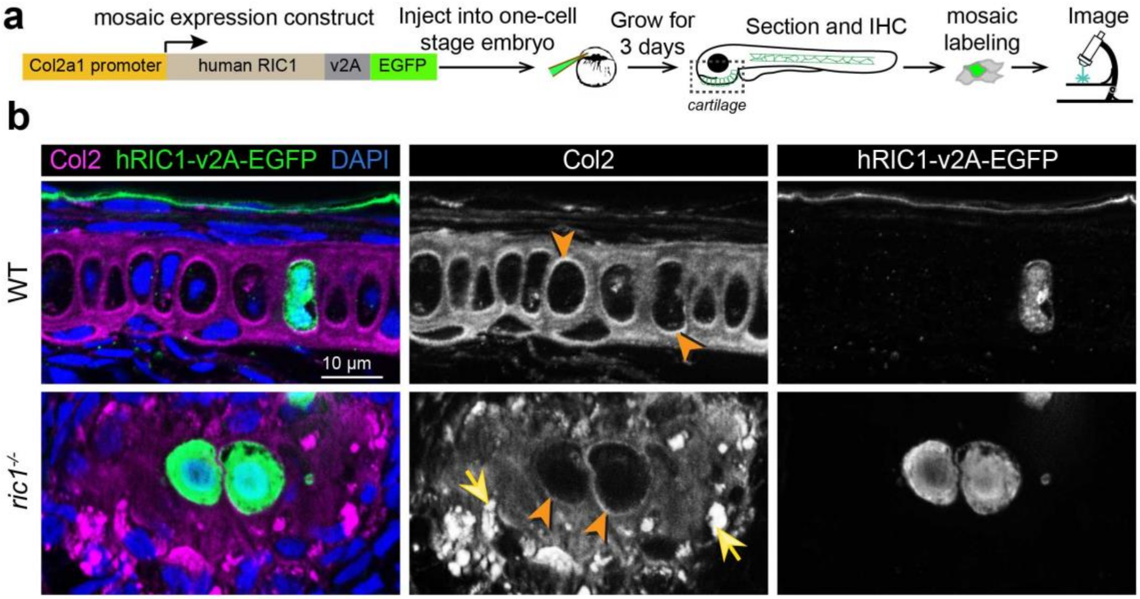


Figure 4.16. Overexpression of human RIC1 restores collagen secretion in *ric1*-deficient zebrafish chondrocytes

(a) Experimental design to mosaically overexpress human *RIC1* (hRIC1) gene in zebrafish. EGFP is linked to hRIC1 via self-cleavable viral 2A (v2A) peptide. **(b)** Co-immunostaining of Col2 and EGFP in cartilage. Green cells express the construct and neighboring cells act as an endogenous WT or mutant controls. Arrows point to collagen accumulation; arrowheads show secreted, extracellular collagen.

construct (Fig. 4.14g). Taken together, *RIC1* depletion likely leads to *RIC1*-associated CATIFA syndrome through disruption of procollagen secretion.

Discussion

Genotype-phenotype relationships are hardly linear in patients; thus, to describe the full phenome associated with a single gene remains a standing problem in genetic medicine. In addition, genomic modifiers and environmental factors produce a range of phenotypic severity in individual patients. Hence, an approach that integrates linear genotype-phenotype relationships and population-level variation is required to fully assess the phenome linked to a single gene.

While each of taken here approaches has led to successes in a variety of cases, each has strengths and limitations. For example, model systems are the preferred way to investigate molecular mechanisms and to perform experimental perturbations to track phenotypes, but it is hard to assess whether what one sees is physiologically relevant in vivo in humans. In omics-driven association studies, one can assess polygenic traits, and generate hypotheses, but it is hard to know what is physiologically important in patients (Editorial, 2018) (other than what one surmise from already known data). We asked how best to integrate the three approaches to capitalize on advantages of each and counter their limitations. The challenge of getting to pathomechanism is particularly acute in the common disease space, but remains prominent in Mendelian diseases as well.

In this study, we describe a novel approach based on integrated use of an animal model discovery tool, a human common disease biobank, and rare Mendelian disease to accelerate and refine the discovery process of disease mechanisms (Fig. 4.17). The integrated use of

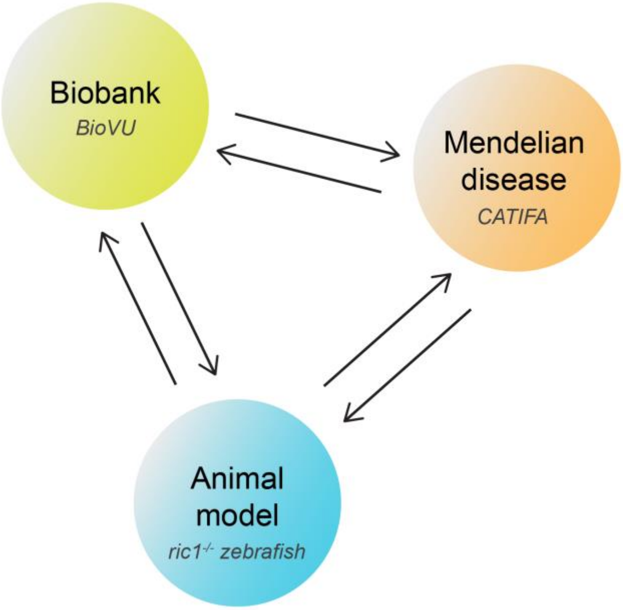


Figure 4.17. Iterative investigations in three systems.
Integrated approach of disease mechanism discovery using the animal model system, phenome ascertainment of common disease in a biobank and analysis of monogenic, rare disease phenome of RIC1/CATIFA syndrome.

the three groups of data is centered on the gene basic biological function and disease mechanisms. Large-effect coding mutations lead to monogenic disorders in which multiple phenotypic manifestations of disease are present from birth. Reduced expression of the same gene may be statistically associated in large samples with similar phenotypes, but subjects rarely have multiple phenotypic associations, and the phenotypes may be milder than in patients with the Mendelian disease. By characterizing how the null mutation in zebrafish or reduced expression in patients of a gene such as *RIC1* leads to particular associated phenome, such as skeletal dysmorphology, we are able to contribute essential pieces of information towards understanding the common disease phenome. The integration of these different sources of data allows us to move from proximal mechanisms such as reduction of gene expression, to cellular mechanisms such as the trafficking of procollagen, to phenotypic mechanisms related to the consequences of a depletion of fibrillar collagen in the ECM.

Collagen constitutes one third of the dry body mass in vertebrates and is essential for the structure and function of a wide-range of organ systems(Ishikawa and Bachinger, 2013; Mienaltowski and Birk, 2014). In disease states such as fibrosis and wound healing, collagen needs to be efficiently secreted to extracellular matrix to enclose the injury site. Despite such abundance and clinical importance, the exact mechanism by which procollagen moves through a cell has not been well understood. Here we, for the first time, contribute evidence that Ric1-Rgp1-Rab6a module in the trans-Golgi network is an essential transit point for the procollagen traffic *en route* to ECM. Although alternative pathways might exist to traffic large ECM cargos, the TGN Ric1-Rgp1-Rab6a regulatory module appears to be essential for high volume, rapid collagen secretion, and when disrupted creates a bottleneck resulting in multi-organ pathologies (e.g. craniofacial dysmorphology, developmental disturbance of the tooth eruption etc.). Impaired ECM secretion also affects the quality of matrices in zebrafish organs, best modeled by the notochord basement membrane (BM) that is thinned and contains fewer protein inclusions. Since all epithelia depend on functional BM, one may speculate that the phenome associated with hernia, gastrointestinal motility, and asthma in CATIFA syndrome, and the RIC1-associated common disease phenome including diverticulosis, non-celiac intestinal malabsorption, and asthma could as well be attributed to BM-related deficits. Despite the strong agreement between the common phenotypes we observed in relation to RIC1 reduced expression and the CATIFA phenotype, we acknowledge that the overlap is incomplete. This might be attributed to the limited number of patients or, perhaps more likely, to their young age relative to some of the phenotypes, which are known to be age-related e.g. intestinal diverticulosis. Furthermore, neurological defects in CATIFA patients, such as ADHD and intellectual disability, might be due to compromised neural ECM since matrix composition and remodeling have been implicated in synaptic plasticity, neuronal guidance, and maintenance of normal neurological functions(Elliott et al., 2018; Senkov et al., 2014; Soleman et al., 2013). Thus, detailed *RIC1*-linked phenome offers a rich dataset of phenotypes associated with post-Golgi transport deficits of procollagen, potentially serving as a reference panel for clinicians.

Better ability to diagnose the complete disease phenome improves the likelihood of success in fully understanding the mechanism of disease. However, clinical evaluation is often unguided and limited to the most severe and obvious symptoms patients manifest, thereby hindering the discovery of the full phenotypic spectrum. Initial phenotyping of LoF animal models and re-iteration of gene-associated phenotypes in a biobank led us to do “guided clinical re-evaluation” of monogenic patients. In our example, craniofacial dysmorphology in zebrafish *ric1*$^{-/-}$ mutants and in the BioVU dataset guided examination of skeletal signs in CATIFA patients. In return, developmental tooth deficits discovered in the biobank prompted teeth re-examination in zebrafish, which in turn facilitated clinical evaluation of dentation as part of the novel syndrome characteristics.

Detailed knowledge of the disease pathomechanism and spectrum of the associated phenome will enhance opportunities for improved patient care. For example, preferential use of non-invasive diagnostic methods due to predicted fragility of BM in various organ systems in CATIFA patients might be important to consider. Similarly, rational therapy designs guided by disease-specific pathomechanisms will improve overall patient care and well-being. The discovery of the genetic defects in genes controlling ECM production, and specifically intracellular trafficking, opens new possibilities for molecular diagnosis in pre- and postnatal care, as well as later in life. The gene-associated phenome in biobank patients can also give a preview of potential later-onset phenotypes and better prepare clinicians and families for planning future care of the patients. As genotype-linked health records grow and quantitative genetics tools (Bastarache et al., 2018; Khera et al., 2018) advance in the upcoming years, this approach will be even more accessible for personalized medicine practices. For example, patients with predicted reduced expression of *RIC1* could receive genetic counseling for expected phenotypes such as teeth conditions stress incontinence or visual disturbances. Hence, integrated use of a biobank with animal models and monogenic disease phenome will enrich the use of biobanks for better diagnostic and patient care-oriented applications.

Methods

Fish maintenance and breeding

Zebrafish were raised under standard laboratory conditions at 28.5°C as previously described(Barrallo-Gimeno et al., 2004; Montero-Balaguer et al., 2006). All experiments were conducted following the guidelines established by the IACUC at Vanderbilt University Medical Center under the protocol number ID: #M1700020-00. *round* alleles [m641, m713, m715 as isolated and described in the MGH zebrafish mutant screen (Driever et al., 1996; Neuhauss et al., 1996)] were kept in AB genetic background for phenotypic analysis. Embryos were staged and fixed at indicated stages, i.e. hours post-fertilization (hpf) and days post-fertilization (dpf) as described previously(Kimmel et al., 1995).

Live imaging of embryos & head size index analysis

5 dpf embryos were anesthetized in Tricaine (Sigma), mounted in 3% methylcellulose (Sigma) and imaged using Zeiss Stemi 2000-C with HRc camera. Using dorsal views in ImageJ, body and head lengths were measured from jaw protrusion to caudal fin fold and to the end of parachordals, respectively. Head size index was calculated by the following formula: head length / body length; statistical analysis was performed using Mann-Whitney test (two-tailed) in order to compare WT and *ric1*$^{-/-}$.

Cartilage and bone staining

5 or 7 dpf embryos were stained for bone (Alizarin red) and cartilage (Alcian blue) as described previously(Walker and Kimmel, 2007). 0.02% Alcian blue and 60 mM $MgCl_2$ in 70% ethanol, 0.005% alizarin red was used as double staining solution. Stained embryos were imaged with Zeiss Axioimager Z1 scope equipped with Axiocam HRc camera.

3D rendering of live hyosymplectic cartilage shape and chondrocyte volume

3 dpf Tg(Col2a1a:caax-EGFP; Col2a1a:H2A-mCherry) double transgenic embryos were lightly anesthetized and mounted in 1.2% low-melting agarose in glass-bottom dishes. Imaging was conducted via LSM880 confocal microscope with 20X/0.80 Plan Apochromat and 40X/1.1 LD C-Apochromat water-immersion objective. Confocal z-stacks of entire hyosymplectic cartilage (~80 µm-thick) and individual chondrocytes (~25µm-thick) were analyzed in Imaris 8.0 (Bitplane) for 3D rendering using the surface tool (shape analysis) and statistics tab (volume analysis).

Genetic mapping and positional cloning

The *round* locus was mapped in an F2 intercross using bulked segregate analysis. DNA samples were genotyped by PCR using SSLP markers evenly spaced across the zebrafish genome(Knapik et al., 1998). The mapped rnd^{m641}, rnd^{m713} and rnd^{m715} mutations were confirmed by Sanger sequencing of genomic DNA flanking the mutation sites from homozygous wild-type F2 animals, heterozygous F2 animals and homozygous mutant F2 animals(Melville et al., 2011).

Comprehensive PrediXcan Analyses

To evaluate the role *RIC1* may play in the etiology of a disease trait, we utilized the PrediXcan (Gamazon et al., 2015) approach. We estimated the genetic component of gene expression in the ten thousand BioVU subjects (Denny et al., 2013), to identify associated phenotypes (despite the lack of directly measured gene expression data). Specifically, from the weights $\hat{\beta}_j$derived from the gene expression model and the number of effect alleles X_{ij} at the variant *j*, we estimated the genetically determined component of gene expression:

$$\hat{G}_i = \sum_j X_{ij}\hat{\beta}_j$$

A significant association between the estimated genetic component of gene expression and a trait suggests a causal direction of effect (because the germline genetic profile is unlikely to be altered by the trait).

Generation of transgenic constructs

Transgenic constructs were generated with Gateway cloning-based Tol2kit method(Kwan et al., 2007). Zebrafish *rab6a* was cloned from 4 dpf wild type (AB) cDNA into vector pEGFP-C1. Next, EGFP-Rab6a fusion was subcloned into the pDONR221 vector via BP cloning, after adding attB1 and attB2 sites with PCR amplification. Obtained clone was utilized as middle entry vector in LR cloning step of Multistep Gateway Recombineering system, in combination with p5E-1.7kb Col2a1a promoter(Dale and Topczewski, 2011), p3E-ployA, pDestTol2pA2 and incubated with Gateway® LR clonase® II enzyme mix. Resulting destination clone was used for mosaic overexpression experiments. The Q72L mutation(Martinez et al., 1994)(constitutively active Rab6a mutant) was generated on destination vector via Q5® site-directed mutagenesis kit (New England Biolabs).

The human *RIC1* (hRIC1) cDNA clone was obtained from Dharmacon GE Lifesciences (Clone ID 9020606). The missing 5' region (including start codon) of *RIC1* was cloned from human BJ skin fibroblast (ATCC® CRL-2522™) cDNA. Two fragments were ligated using common AflII restriction site to obtain full length *RIC1* coding sequence. Full length *RIC1* was PCR-amplified with attB1(forward) and attB2 (reverse) recombineering sites. Similar Gateway® cloning approach was performed as described above, except that the p3E-v2A-EGFP vector (gifted by Dr. J.T. Gamse) was used as 3' element in order to label transgenic cells with cytoplasmic EGFP signal that forms upon self-cleavage of v2A-EGFP from the fusion polypeptide(Kim et al., 2011b; Provost et al., 2007). The R1265P *RIC1* mutation identified in CATIFA patients was introduced to destination vector via Q5® site-directed mutagenesis kit. Primers used are listed in Table 4.2.

Microinjection of Tol2kit clones for mosaic overexpression

Zebrafish embryos at 1-cell stage were injected with a combination of 50 pg medaka transposase mRNA and 10-20 pg of destination vector DNA and grown to indicated stages in embryo medium.

Immunofluorescence (IF) and fluorescence microscopy

IF was performed as previously described(Cox et al., 2018). Primary antibodies used were anti-GFP (Vanderbilt Antibody and Protein Resource), anti-collagen type-II (Rockland, #600-401-104), anti-Fibronectin (Sigma, #F3648), anti-β-catenin (Sigma, #C7207), anti-Matrilin (gifted by Dr. E. Kremmer). 4',6-diamidino-2-phenylindole (DAPI, Molecular Probes) was used as a nuclear counterstain. Fluorescently labeled secondary antibodies used were either Alexa Fluor 488- & 555- (Life Technologies) or DyLight 550-conjugates (Thermo Scientific). Proteoglycan staining was performed using Alexa Fluor-555 conjugated wheat germ agglutinin (WGA, Life Technologies). Fluorescence imaging was carried out using AxioImager Z1 equipped with an Apotome (Zeiss) and an EC Plan Neofluar 100x/1.30 Oil objective.

Analysis of collagen accumulation in zebrafish cells

Images of collagen IF were analyzed in ImageJ. For collagen accumulation in chondrocytes, intracellular area of each cell (demarcated by caax-EGFP signal), collagen area (Col2 signal) and nuclear area (DAPI) were measured. Percent collagen cytoplasmic area per cell was calculated by the following formula:

*[Collagen area / (Intracellular area – Nuclear area)] * 100*

For collagen accumulation in notochord sheath cells, maximum intensity projections were utilized to span entire cell boundaries. The region around visible full cells in maximum intensity projection views were drawn by tracing caax-EGFP signal and sub-selection images were produced. Total area of sub-selection was measured as total intracellular area.

Total collagen area was measured by tracing Col2 signal in individual cells and adding them up. Percent collagen area per cell was calculated by the following formula:

*[(Total collagen area / Total intracellular area) / number of cells in the region] * 100.*

Collagen content in WT and *ric1*$^{-/-}$ cells was statistically compared using Mann-Whitney test (two-tailed).

Electron microscopy & quantification of vesicles in zebrafish tissues

Samples were processed as previously described(Melville et al., 2011). Briefly, 3 and 4 dpf embryos were fixed in 2.5% glutaraldehyde in 0.1 M sodium cacodylate by incubating at room temperature for 1 hour first, then overnight at 4°C, followed by TEM sample processing. 70-nm sections were collected on a Leica Ultracut Microtome and analyzed on a Phillips CM-12 Transmission Electron Microscope provided by the VUMC Cell Imaging Shared Resource. Intracellular vesicles were classified according to electron density and morphology as described in ‘Results’ section. Width and height of vesicles were measured using ImageJ’s ‘Measure’ tool.

CRISPR/Cas9 genome editing

CRISPR/Cas9 target sites within zebrafish *rgp1* gene (GRCz10 assembly) were identified using the CHOPCHOP(Montague et al., 2014) web tool. The following site was targeted in this study; 5’-GGTGGTGGCATCTATGGCACGGG-3’. A cloning-free method to generate sgRNA templates was performed as previously described(Varshney et al., 2015).

Guide RNAs were synthesized with MEGAshortscript™ T7 transcription kit (ThermoFisher Scientific).

To generate mutations with CRISPR/Cas9 system, a mixture of 500pg purified Cas9 protein (PNA Bio Inc, # CP01) and 180 pg gRNA was injected into one-cell stage embryos. Injected embryos were grown to 3-5 dpf stage for phenotypic analysis. Mutations generated in injected embryos were detected via direct sequencing of the region surrounding the target site. PCR-amplified products were cloned into pGEM-T Easy vector (Promega) by T-A cloning and sequenced using SP6 primers in order to individually detect various mutations created by CRISPR/Cas9 system.

A similar strategy was followed for CRISPR-mediated genome editing of *ric1* gene (ENSDART00000082377 transcript of *KIAA1432* gene, *ric1*, on Zv8 genome assembly was used as target template). The following site within exon 25 (out 34 total exons) was targeted: 5'- GGTGCGTAATCTGGGTGAGCAGG-3' in ENSDART00000082377 transcript (Zv8).

Participants and Genetic Analysis

All patients, siblings and parents were clinically evaluated and enrolled in an IRB-approved research protocol with written informed consent (KFSHRC RAC# 2070023). Written photo consent was obtained from the parents for all individuals shown in Fig. 3. Venous blood was collected from all participants in EDTA tubes for DNA extraction.

DNA samples were genotyped on the Axiom SNP chip platform according to manufacturer's instructions (Affymetrix, Santa Clara, CA, USA). SNP genotypes were utilized to perform multi-point linkage analysis with the Genehunter program (version 2.1r5) within the EasyLinkage plus (V5.08) software package. Linkage analysis of both families confirmed a single significant linkage peak (LOD 3.44) on Chr9 spanning *RIC1*.

Human fibroblast culture, TEM imaging and IF

Dermal fibroblasts from one of the CATIFA patients (15DG2428) was isolated from skin biopsy and cultured in MEM (GIBCO) supplied with bovine serum. BJ skin fibroblasts (ATCC® CRL-2522™) were used as control for all experiments described further. EM was performed as described above. Cells grown in 10 cm^2 dishes to confluency were fixed in 2.5% glutaraldehyde in 0.1 M sodium cacodylate by incubating at room temperature for 1 hour first, then overnight at 4°C, which were then further processed for TEM imaging.

Prior to collagen-I IF, cells were treated with ascorbate for 1 hour at 37ºC with 5% CO_2 as described previously(Gorur et al., 2017). Primary antibodies used were anti-Collagen-I (abcam, ab34710), anti-p230 (BD Biosciences, 61120) and anti-Golgin-97 (ThermoFisher, A-21270). Anti-mouse and anti-rabbit secondary antibodies conjugated to Alexa Fluor 488- & 555 (Life Technologies) were used for fluorescence imaging. DAPI (Molecular Probes) was used as nuclear counterstain. Slides were imaged with LSM880 confocal microscope using 20X/0.80 Plan Apochromat and 63x/1.40 Oil objective. Confocal images were acquired under the same settings for control and patient's fibroblasts.

Image analysis of cultured fibroblasts

Fluorescence intensity of collagen-I signal per cell and total cellular area were measured in ImageJ using ‘Measure’ tool. Saturated pixels (255 value in dynamic range spectrum) and background-level low pixels (1-10 values) were eliminated from intensity datasets. Then, total intensity per cell was divided by cellular area and used as a factor of intracellular collagen-I content. Control and patient’s datasets were statistically compared using Student’s t-test.

Colocalization analysis of Collagen-I and p230 (TGN marker) was performed using Coloc2 tool in ImageJ. Pearson’s R-values were recorded from multiple cells and compared using Student’s t-test.

Table 4.2. Primers used in this study

Primer name	**Primer Sequence (5’→ 3’)**
zRab6a-XhoI-F	CCCCTCGAGCCATGTCTGCAGCAGGAGATTT
zRab6a-HindIII-R	CCCAAGCTTTCAGCATGAACAGCCGCCTT
EGFP-Rab6a-attB1F	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGTGAGCAAGGGCGAG
EGFP-Rab6a-attB2R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGCATGAACAGCCGCCTT
Rab6a-Q72L-F	ACAGCAGGACTGGAGCGTTTC
Rab6a-Q72L-R	GTCCCAAAGCTGCAGCCG
hRIC1-1stFrg-F	CTGAGTGTGACGGACGCAA
hRIC1-1stFrg-AflII-R	ATCTTAAGGGGATCTTTTTTGGTGCCAT
hRIC1-2ndFrg-F	CAAATGAAGGGGACACCCCA
hRIC1-2ndFrg-R	TACTGCCCCTTTGTGATGGA
hRIC1-attB1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTATTTTCTGAGCGGCTG
hRIC1-attB2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTTGGACACAGAACAGTCGTAAGTC
hRIC1-R1265P-F	GTCCAGCTTCCGTATTTGCTACAC
hRIC1-R1265P-R	CTGGGATTTATGAGGCCC
Rgp1-genotyp-F1	CGTCTCCGGTTGTTTTGTTT
Rgp1-genotyp-R1	GGTGTGACATTGGGTTTGTG
m713-genotyp-F	CCGGCTGTTGCTATGTTCTT
m713-genotyp-R	GCTACGGCAATCATGGAGTC
m641-genotyp-F	CTGCTGGTGCGTAATCTG
m641-genotyp-R	CTGCAGTATGATCAGGTAT
m715-genotyp-F	TTCGGCTTCTATCAGCAGGT
m715-genotyp-R	AGCCAATCACAGCCATTTCT

CHAPTER V

GRIK5 GENETICALLY REGULATED EXPRESSION ASSOCIATED WITHEYE AND VASCULAR PHENOME: DISCOVERY THROUGH ITERATION AMONG BIOBANKS, ELECTRONIC HEALTH RECORDS & ZEBRAFISH

Gokhan Unlu[1,2,3] *, Eric R. Gamazon[1,2,4,5] *, Xinzi Qi[1,2], Daniel S. Levic[1,3,]† Lisa Bastarache[6], Joshua C. Denny[2,6], Dan M. Roden[2,6,7], Ilya Mayzus[8], Max Breyer[1,2], Xue Zhong[1,2], Anuar I. Konkashbaev[1,2], Andrey Rzhetsky[8], Ela W. Knapik[1,2,3] #, Nancy J. Cox[1,2,4 #]

* Equal contribution, # Co-corresponding authors.

[1]Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN 37232. [2]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37232, USA. [3]Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN 37232. [4]Data Science Institute, Vanderbilt University, Nashville, TN 37232, USA. [5]Clare Hall, University of Cambridge, Cambridge, CB3 9AL United Kingdom. [6]Departments of Medicine and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37232. [7]Department of Pharmacology, Vanderbilt University, Nashville, TN 37232, USA. [8]Departments of Medicine and Human Genetics, The University of Chicago, Chicago, IL 60637

This chapter was published under this title in *The American Journal of Human Genetics* (Unlu G., Gamazon E.R. et al., 2019).

Abstract

While the use of model systems to study the mechanism of large effect mutations is common, we highlight here the ways that zebrafish model system studies of a gene, *GRIK5*, contributing to the polygenic liability to eye diseases have helped to illuminate mechanism implicating vascular biology in eye disease. Prediction of gene expression derived from a reference transcriptome panel applied to BioVU, a large electronic health record-linked biobank at Vanderbilt University Medical Center, implicated reduced *GRIK5* expression in diverse eye diseases. We tested the function of *GRIK5* by depletion / knockout of its ortholog in zebrafish, observing reduced blood vessel numbers and integrity in the eye, and increased vascular permeability. Analyses of electronic health records in > 2.6 million Vanderbilt subjects revealed significant comorbidity of eye and vascular diseases (relative risks 2-15), which was confirmed in 150 million individuals from a large insurance claims dataset. Subsequent studies in >60,000 genotyped BioVU participants confirmed the association of reduced genetically predicted expression of *GRIK5* with comorbid vascular and eye diseases. Our studies pioneer an approach for discovery of gene-phenotype relationships that allows a rapid iteration to primary genetic mechanism contributing to the pathophysiology of human disease. Our findings also add dimension to the understanding of the biology driven by glutamate receptors such as *GRIK5* and to mechanisms contributing to human eye diseases.

Introduction

Eye diseases are a diverse set of conditions that are both common and largely consistent with polygenic genetic architecture. Eye disease pathophysiologies arise in different cell types and appear to be distinct; however, understanding of the biological basis of eye diseases has been derived primarily from single-gene approaches or small-scale genetic studies. Unbiased, at scale big-data approaches using Electronic Health Records (EHR) may generate disease mechanism hypotheses that can be validated using animal models.

Gene-based tests, as opposed to tests of individual SNPs, offer an alternative approach for the identification of genes for common diseases. PrediXcan (for predicted expression scanning) uses a reference panel to create SNP-based prediction models for transcript levels for each gene in each tissue (Gamazon et al., 2015). Testing for associations with predicted expression improves power relative to individual SNP associations because we are testing a specific feature (in this case imputed genetically determined gene expression) of genome function that has been shown to drive a substantial fraction of the common variant heritability to common diseases (Gamazon et al., 2014; Gusev et al., 2014). Using predicted gene expression provides a clearer route to model organism validation because the association result is at the level of the gene rather than the SNP (as is the case with GWAS results), and has an easy-to-interpret direction of effect.

Recent advances in CRISPR/Cas9 technology have revolutionized the fidelity and speed of generation of animal models and now allow for precise genome editing in animals and cell lines with high efficiency and minimal off-target effects (Cong et al., 2013; Hwang et al., 2013; Iyer et al., 2018; Jao et al., 2013). The rapid development of transparent embryos and genetically modified adults facilitate phenotypic analysis of variants in intact zebrafish over long periods of time, enabling 4D imaging.

Studies over the last decade have shown EHR data linked to DNA biobanks are an efficient route to human genetic discovery (Gottesman et al., 2013; Kho et al., 2011; Klarin et al., 2017; McCarty et al., 2011; Wei and Denny, 2015; Wolford et al., 2018). Use of phenotype algorithms in EHR data allows both replication of existing genetic associations and discovery of associations (Denny et al., 2011) and allows for both genome-wide and phenome-wide association studies with results that replicate those from studies conducted with research quality diagnoses (Denny et al., 2013). The much larger sample sizes and broad spectrum of the medical phenome that are available (and affordable) through the use of EHR phenomics create opportunities for studies that would otherwise not be feasible and lead logically to research questions and discoveries that cannot be considered in research datasets comprised of cases and controls for a single disease.

We summarize here results of a set of studies beginning with discovery of the association of reduced genetically determined expression of *GRIK5*, which encodes glutamate ionotropic receptor kainate type subunit 5, with a diverse set of eye diseases in BioVU, the EHR-linked DNA biobank at Vanderbilt University Medical Center (Roden et al., 2008), using related publicly available large-scale GWAS, and validation studies in zebrafish. We show that the reduced *GRIK5* expression compromises vascular integrity in an animal model and increases genetic risk for diverse eye diseases in BioVU participants. Our study pioneers a model for discovery and rapid progression to an understanding of physiological mechanisms linking genome variation to human disease by iterating between methods for big-data integration (genome variation by transcriptome measurement) applied to biobanks and large-scale EHRs with modern approaches to gene editing in animal model systems.

Methods

DNA Biobank and Electronic Health Records Database

BioVU, the Vanderbilt University biobank linked to electronic health records (EHR), aggregates ICD-9 codes to represent diseases and other phenotypes for use in clinical or genetic studies (Denny et al., 2013; Denny et al., 2010). The phenotype codes ("phecodes") combine related ICD-9 codes to represent clinical traits. These phecodes, first implemented in 2010 (Denny et al., 2010), are supported by clinical co-occurrence data that identify related phenotypes across the ICD-9 code system. For our study, we used version 1.2, which includes 1,965 hierarchical phecodes and 20,203 ICD-9 codes (Denny et al., 2013). The phecode hierarchy includes clinical traits that are not found in the ICD-9 system, including "inflammatory bowel disease" as the parent node for "ulcerative colitis" and "Crohn's disease." For example, the phecode "other retinal disorders" includes ICD9 codes for retinal exudates and deposits, and retinal nerve fiber bundle defects as well as other retinal disorders and retinal disorders NEC. The phecode for other disorders of eye includes ICD9 codes for acute dacryoadenitis, other disorders of sclera, and other disorders of eye.

These phecodes have been employed in Phenome-Wide Association Studies (PheWAS) to validate known trait associated genetic variants as well as to discover associations, including those that have been subsequently validated (Denny et al., 2013; Denny et al., 2010; Hall et al., 2014; Pendergrass et al., 2013). However, these PheWAS studies have focused strictly on single nucleotide polymorphisms rather than on genes, which are natural, biologically relevant units for evaluating associations with disease and for proposing disease mechanisms. Here we applied PrediXcan (Gamazon et al., 2015) to perform gene-level PheWAS of a broad array of traits, using the unprecedented collection of human tissue data (n = 44) made available by the Genotype-Tissue Expression (GTEx) project (GTEx_Consortium., 2015). We used the GTEx reference resource to impute tissue-level gene expression in the BioVU subjects using the PrediXcan method. We refer to the imputed gene expression level as the *Genetically Regulated eXpression* (GReX) (Gamazon et al., 2015).

The Synthetic Derivative (SD) is a de-identified and continuously updated image of the EHR including data on > 2.6 million subjects (Roden et al., 2008). We conducted analyses on data from the SD to test the phenome level comorbidity between vascular traits and eye diseases. Included in these analyses were several vascular phenotypes and all of the original eye phenotypes showing association to the GReX (Gamazon et al., 2015) for *GRIK5*, including retinal detachment and defects, other retinal disorders, cataract, senile cataract, glaucoma, primary open-angle glaucoma, open-angle glaucoma, and other disorders of the eye. To control for potential confounding due to age and sex, we calculated the adjusted relative risk (RR) by using logistic regression (Kleinbaum et al., 1998) with age and sex as covariates. In addition, we performed stratified analysis, estimated the RR for each stratum (as defined by sex and age group), and pooled the resulting estimates into an adjusted RR. We performed replication of the RR estimates from a stratified analysis (by age and sex) of the MarketScan insurance claims dataset consisting of >150 million subjects. We used the Cochran-Mantel-Haenszel estimator, which takes into account the stratification (due to age and sex):

$$\widehat{RR_{CMH}} = {}^{\sum a_i(c_i + d_i)\,/n_i}\Big/_{\sum c_i(a_i + b_i)\,/n_i}$$

where the index *i* refers to each stratum and $n_i = a_i + b_i + c_i + d_i$. For each stratum *i*, *a*, *b*, *c*, and *d* indexed by *i* are defined by the following two-by-two table:

		Disease 2	
		Case	Control
Disease 1	Case	a	b
	Control	c	d

PrediXcan analysis

To evaluate the role a gene may play in the etiology of a disease trait, we utilized the PrediXcan approach (Gamazon et al., 2015). We estimated the genetic component of gene expression in the BioVU samples (Denny et al., 2013), to identify genes associated with a phenotype (despite the lack of directly measured gene expression data). Specifically, from the weights $\hat{\beta}_j$derived from the gene expression model based on local genetic variation (+/- 1 Mb of the gene) (Gamazon et al., 2015) and the number of effect alleles X_{ij} at the variant j, we estimated the genetically determined component of gene expression:

$$\hat{G}_i = \sum_j X_{ij}\hat{\beta}_j$$

A significant association between the estimated genetic component of gene expression and a trait proposes a causal direction of effect (because the germline genetic profile is unlikely to be altered by the trait). We also evaluated whether a significant gene-level association may be "contaminated" by associations to nearby genes; such contamination may result from linkage disequilibrium between SNP predictors of gene expression and trait causal variants (Wainberg et al., 2018). We note that in general the top SNP for a gene is not optimal for prediction of the expression of the gene and multi-SNP models tend to outperform single-SNP models (Gamazon et al., 2015). For example, the number of SNPs in the optimal imputation model for *GRIK5* generated from Depression Gene Network (DGN) whole blood expression data is 15.

eQTL-based heritability estimation and polygenic modeling

Polygenic modeling is an approach aimed at relating phenotypic variation to multiple genetic variants simultaneously. As a method for genome-wide analysis, it differs from conventional single-variant tests of association by testing large numbers of loci (potentially in the thousands) for their contribution to the genetic architecture of phenotype. We sought to estimate SNP-based heritability, defined as the proportion of the phenotypic variance captured by the genetic variants in aggregate, for a broad spectrum of complex human traits. We define the following mixed-effects model

$$Y = Xb + \sum_T g_T + C + e, \ \mathrm{var}(Y) = \sum_T A_T \sigma_T^2 + A_C \sigma_C^2 + I\sigma_e^2$$

Here, Y is a phenotype vector and b a vector of fixed effects (e.g., principal components [representing ancestry], sex, PEER factors [representing hidden or unmeasured variables] (Stegle et al., 2012)). A_T is the genetic relationship matrix (GRM) estimated from the SNP set T and g_T denotes the polygenic component attributable to the set T with mean of of g_T

equal to zero and variance equal to $A_T\sigma_T^2$. C is the aggregate genetic effect of the complement set of variants in the genome. The phenotype is modeled as the sum of these genetic effects (random effects) and the relevant covariates (fixed effects) and a residual. In the case of a single tissue, g_T and C are assumed to have possibly different distributions of effect sizes. Variances were estimated using restricted maximum likelihood (REML). The heritability attributable to the SNP set T is then calculated as the fraction of the phenotypic variance σ_Y^2:

$$h_T^2 = \sigma_T^2 / \sigma_Y^2$$

We utilized both BioVU data and publicly available (independent) GWAS data to test for the existence of a shared polygenic component underlying the eye and vascular phenome comorbidity results. We tested the top associations ($p<0.05$) from a large-scale GWAS of age-related macular degeneration (AMD) (Fritsche et al., 2016) for their association with vascular traits in BioVU. We generated a Q-Q plot from each (vascular trait association) analysis of the top AMD variants; a leftward shift from the diagonal “null” line in a Q-Q plot would indicate enrichment. To show that the observed enrichment was not driven by LD among the tested variants, we applied a polygenic test that takes into account the LD. We used the sum of all the chi-squared 1df statistics as the overall test statistic and simulated 1000 multivariate normal distributed n-vectors v (with n the number of tested variants) with mean zero and covariance matrix Σ given by the nxn pairwise LD matrix:

$$v \sim N(0, \Sigma)$$

$$v \epsilon R^n$$

The proportion of simulated datasets with test statistics that match or exceed the observed test statistic in the actual dataset yields an empirical enrichment p-value. More generally, this approach provides a test of polygenic enrichment using only summary statistics.

Initial PrediXcan studies

Initial PrediXcan studies were conducted in 5,240 BioVU subjects of European descent using prediction equations built from whole blood using data on > 900 subjects from the DGN data set with RNA-Seq and 922 genotype array data (Illumina Omni1-Quad) (Battle et al., 2014) and heart left ventricle on 190 subjects from GTEx. The BioVU subjects had been genotyped using the Omni1-Quad array, and because the data from these studies had not yet been imputed, we focused on the 125 genes in whole blood and the 298 genes in heart left ventricle for which all SNPs in the prediction equations were directly genotyped on the Omni1-Quad array.

The Institutional Review Board reviewed the research study and determined the study does not qualify as "human subject" research per §46.102(f)(2)

Comprehensive PrediXcan analyses

We used transcriptome data in 44 tissues and whole-genome genotype data (imputed to sequence data from the 1000 Genomes reference panel) derived from the same donors across tissues from the GTEx Consortium (GTEx_Consortium, 2013; GTEx_Consortium., 2015) as a reference resource to build gene expression prediction models, i.e., SNP predictors and corresponding weights for the gene expression level. For downstream analysis, we are interested primarily in GReX (i.e. imputed gene expression level) rather than the trait-altered component of gene expression or other factors (such as environmental regulators or technical confounders). GReX is then tested for association with a clinical trait as represented by a phecode in an electronic health records [EHR] database such as linked to BioVU (Denny et al., 2013; Flintoft, 2014). An observed association for GReX with a clinical trait proposes a causal, easy-to-interpret direction of effect because the clinical trait is not likely to alter the germline genetic profile. This analytic workflow allows us to generate a comprehensive medical phenome catalog, consisting of all associations, including direction of effect, between GReX for each tested gene and each clinical trait in the EHR.

We performed transcriptome-wide association (TWAS) analysis on BioVU using PrediXcan. This analysis generated a list of genes associated with clinical traits; in particular, we obtained the effect size and the level of significance (p-value) for each gene-trait association. As a framework for identifying genes associated with disease traits, we would apply Bonferroni adjustment (adjusted p-value<0.05), based on the number of tissues (n=44) and the number of phecodes (n=1,965).

The clinical traits used in this analysis are represented by phecodes (Denny et al., 2013), which are algorithmically defined using ICD-9 codes in the EHR database. One of the top findings was glutamate ionotropic kainate receptor KA2 encoded by *GRIK5*, based on level of significance and direction of effect for a set of related clinical traits affecting eye function. The gene is expressed in a variety of tissues, including the brain regions sampled within GTEx.

Assessing significance for a set of related clinical traits

We conducted permutation analysis that preserves both linkage disequilibrium for genetic variants and correlations among phecodes. Let Y denote the NxQ matrix of phecodes, where N is the number of individuals and Q is the number of phecodes. Let G be the NxP genotype matrix (encoded as the number of effect alleles), where P is the number of genetic variants in the imputation model of a gene. Let $\hat{\beta}$ be the P-dimensional effect size vector from the imputation model. The rows for the phecode matrix Y were randomly shuffled to generate permuted datasets (n=10^6) while the genotype matrix G was left untouched. For each permuted dataset, logistic regression was performed, as in the actual data, between each phecode (column Y_{*j}) and the N-dimensional vector $G * \hat{\beta}$ of genetically determined expression.

We calculated a statistic T, namely the number of nominally significant associations (p<0.05) of reduced genetically determined expression with the phecodes for eye phenotypes, from each permuted dataset, yielding an empirical distribution. We compared the value T_0 of the statistic for the actual dataset with this empirical distribution and assessed the significance of the multi-trait association as the proportion $\hat{P}(T \geq T_0)$ of permuted datasets with statistic T that matches or exceeds T_0.

We also calculated a combined test statistic using Fisher's method ($X_{2k}^2 \sim -2\sum_{i=1}^{k} \ln(p_i)$) applied to the p-values (p_i) for the *GRIK5* associations and evaluated its significance based on a comparison with an empirical frequency distribution of the maximum combined test statistic across all genes tested.

Zebrafish Studies: Fish husbandry and breeding

Zebrafish were raised at 28.5ºC, in standard laboratory conditions as previously described (Montero-Balaguer et al., 2006). All experiments were performed according to an approved animal protocol and guidelines established by IACUC at Vanderbilt University Medical Center.

Quantitative PCR analysis

Quantitative real-time PCR (qRT-PCR) was performed as described previously (Melville et al., 2011; Sarmah et al., 2010). Total RNA was extracted from 10-15 embryos (per sample) at different developmental stages using the TRIzol reagent (ThermoFisher Scientific). 500 ng of total RNA was used as template for reverse transcription to make cDNA using M-MLV reverse transcriptase (Promega) and poly-T primer. Each PCR reaction was performed with 20 ng of cDNA, SYBR Green Real-Time PCR Master Mix and 2 µM of each primer. Primer sequences used in this study are as follows:

β-actin: 5’-GACTCAGGATGCGGAAACTG-3’
5’-GAAGTCCTGCAAGATCTTCAC-3’,
grik5_set#1: 5’-CCACCAGCCTGGACATCAAT-3’
5’-AGCTACGGCCAAATCAGCTT-3’,
grik5_set#2: 5’- CCGTACGGATGGCTGCTATT-3’
5’-GACCACGCCCTTTGGTAGAA-3’.

qRT-PCR reactions were run on CFX96 (Biorad) system. Data were analyzed with –ΔΔCt method.

Morpholino knockdown & mRNA rescue

An antisense morpholino oligonucleotide (MO) (Gene Tools) was designed to target 5’UTR of *grik5* (5'-GAGATGCCTTCTGCTGCCTATAGCA-3') as described previously (Levic et al., 2015). 1 nl of MO was injected into one-cell stage zebrafish embryos at varying concentrations (1 ng, 2 ng, 4 ng, 6 ng) to determine effective dose and 6 ng MO was used throughout this study. As negative control, a 25-base random sequence mixture control oligo (Gene Tools) was injected at the same concentration (6 ng) throughout this study. Phenotypes were evaluated by two screeners, independently, in a double-blinded manner.

The human *GRIK5* cDNA clone in pENTR223.1 was purchased from Plasmid ID, Harvard Medical School (Clone ID: HsCD00295568) and subcloned into pCS2+ using ClaI and XbaI (NEB) restriction enzyme sites for ligation. A ligated construct was then linearized with NotI (NEB) and used for in vitro transcription reaction to synthesize mRNA with mMESSAGE mMACHINE SP6 Transcription Kit (ThermoFisher Scientific). 250 pg human *GRIK5* mRNA was injected in combination with MO in rescue experiments.

CRISPR/Cas9 genome editing

CRISPR/Cas9 target sites within zebrafish *grik5* (GRCz10 assembly) were identified using the CHOPCHOP web tool. Off-targets recognized by gRNAs with all potential 0, 1, 2 and 3 mismatches were investigated using CHOPCHOP web tool to avoid non-specific mutations. gRNAs with *in silico* efficiency scores greater than 0.55 were prioritized. The following site that shows no off-targets were selected in this study; g03:

GGTGGACGATGGTCTGTACGGGG. A cloning-free method to generate gRNA templates was used as previously described (Varshney et al., 2015). RNA template was obtained by annealing a 80-nt chimeric gRNA oligo (5’-TTTTGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC-3’) with a gene-specific oligo (5’-AATTAATACGACTCACTATAGG[N18]GTTTTAGAGCTAGAAATAGC-3’) that contains T7 promoter sequence, 18 nt *grik5* gRNA target (TGGACGATGGTCTGTACG) and overlapping sequence to the chimeric sgRNA oligo. Upon annealing two oligos, gaps were filled with T4 DNA polymerase (NEB), yielding a double stranded linear template for in vitro transcription. Guide RNAs were synthesized with MEGAshortscript T7 transcription kit (ThermoFisher).

To generate mutations with CRISPR/Cas9 system, a mixture of 500 pg purified Cas9 protein (PNA Bio Inc, Cat No # CP01) and 100 pg – 500 pg gRNA was injected into one-cell stage embryos. Prior to microinjection, gRNA and Cas9 protein were mixed and incubated on ice for 10 minutes to form ribonucleoprotein complex. Cas9 protein-only and gRNA-only injections were included in the experimental design to control for potential non-specific phenotypes. Embryos injected with Cas9 protein and *grik5*-targeting gRNA were raised to adulthood to generate founders (G_0).

Direct sequencing for mutation detection

PCR products of CRISPR targets were cloned into pGEM-T easy (Promega) following the instruction manual. Plasmids isolated from pGEM-T clones were sequenced with SP6 primer (Genewiz, NJ). Mutant sequences were aligned with WT *grik5* sequence with NCBI/NLM's BLAST® tool to detect deletions, insertions and substitutions.

Whole mount imaging of transgenic zebrafish embryos

Transgenic, Tg(flk1:eGFP), embryos at 2 days post fertilization (dpf) or 3 dpf stage were mounted in low melt agarose (Sigma A-9414, 1.2% in embryo medium) in glass bottom dishes (MatTek Corporation). Images were acquired with Spinning Disk Confocal Microscope (Nikon) using either Plan Apo λ 10X or 20X objectives and Andor DU-897 EMCCD camera. Z-stacks were acquired for 3D analysis and volume views were generated with Nikon NIS-Elements software.

Mutant *grik5*^{g03} embryos, their WT and heterozygous siblings were imaged and analyzed in a blinded manner, before genotype was confirmed by sequencing.

Live imaging of zebrafish embryos

Live embryos were anesthetized in tricaine (Sigma) and mounted in 3% methylcellulose (Sigma) on a bridge slide for imaging. Images were acquired with Stemi 2000-C Stereomicroscope (Zeiss) and Axiocam HRc camera under transmitted light illumination. For magnified views of hemorrhage phenotypes and detailed localization analysis, embedded embryos were imaged under double spot illumination with Axioimager.Z1 (EC Plan NEOFLUAR 5X and 10X objectives) equipped with Axiocam HRc camera.

Nanobead (microsphere) injection and microangiography

Carboxylate-modified microspheres of 0.02 µm size (FluoSpheres®, ThermoFisher Scientific F8786), emitting red fluorescence (580/605) were diluted 1:5 in 0.3X Danieau buffer to 0.4% final concentration. Nanobeads were sonicated with Heat systems sonicator three times with 5 minute-cycles to prevent clumping. Microspheres were, then, injected into common cardinal veins of 3 dpf transgenic Tg(flk1:eGFP) zebrafish embryos to mark blood plasma in red/ magenta. Injected, live embryos were lightly anesthetized in tricaine and mounted sagittally in 1.2% low-melting agarose (Sigma, A9414) in coverslip-bottom dishes (MatTek Corp., P35G-1.5-14C) for confocal imaging. Multichannel images were acquired with Spinning Disk Confocal Microscope (Nikon) using Plan Apo λ 20X objective and Andor DU-897 EMCCD camera. Z-stacks were acquired for 3D analysis and volume views were generated with Nikon NIS-Elements software.

Results

Construction of PredixVU, Gene-by-Medical Phenome Electronic Catalog

We devised a strategy to link genome variation to EHR phenotypes that resulted in the construction of a large-scale medical phenome catalog (Fig. 5.1). We used reference data including measurement of both genome variation and gene expression levels, such as GTEx (GTEx_Consortium, 2013) or Depression Genes and Networks (DGN) (Battle et al., 2014), to build SNP-based prediction or imputation models for measured gene expression levels in up to 44 tissues using the PrediXcan (Gamazon et al., 2015) approach (Fig. 5.1A).

The prediction models were then applied to genotyped individuals from BioVU to estimate the level of genetically regulated expression, GReX (see Methods), for each gene in each tissue (Fig. 5.1B). We then tested the associations of the GReX values calculated with DNA variation from subjects in a biobank with 1145 "phecodes" (phenotype codes, see Methods) associated with these subjects (Fig 5.1C). The results of these association studies comprise a comprehensive gene-by-medical phenome electronic catalog we have termed PredixVU (Fig 5.1D).

Here, we present one example of an association we discovered between reduced *GRIK5* expression and the eye phenome, and test it by *in vivo* loss-of-function approaches in a model organism, zebrafish and with analysis of phenome relationships in large-scale EHR data.

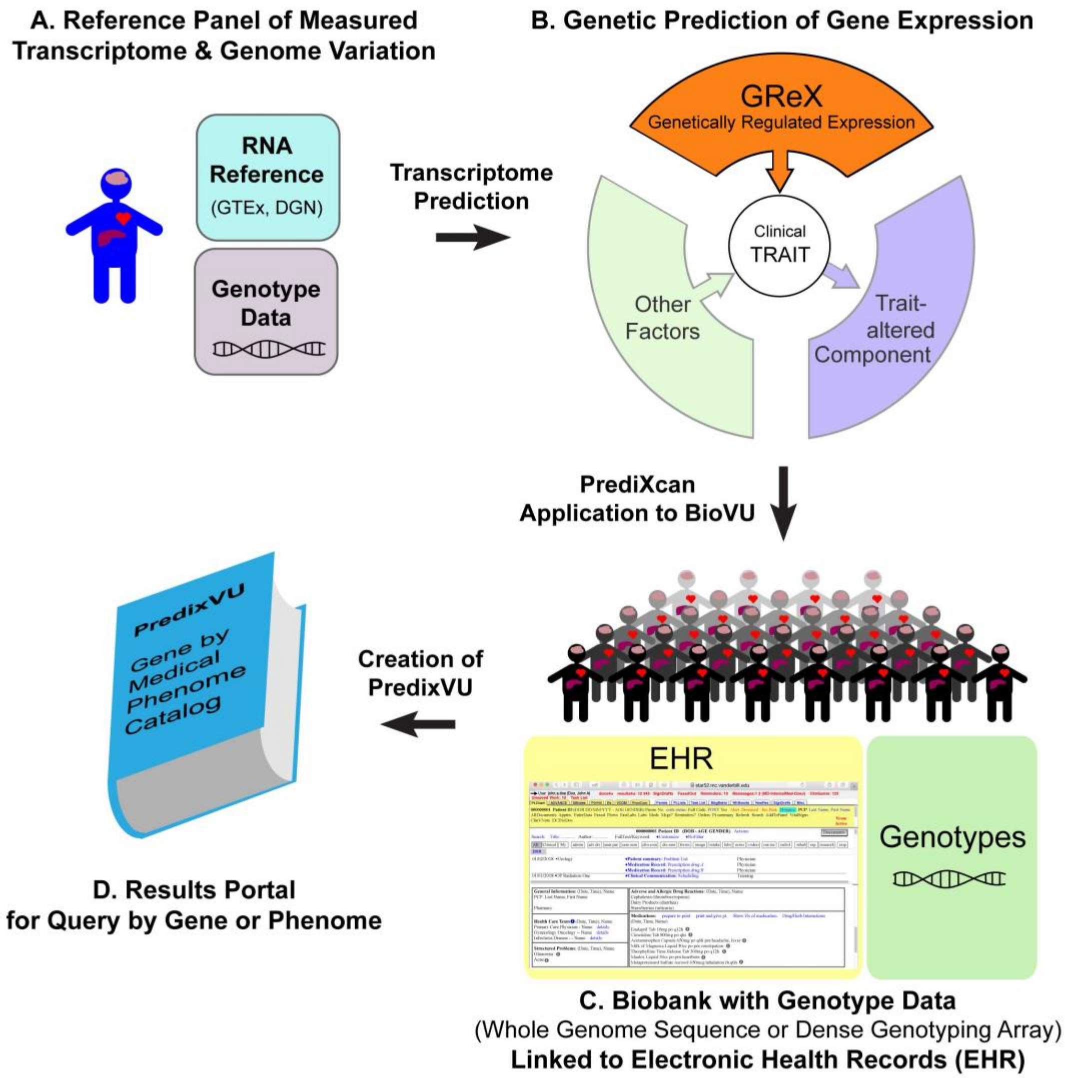

A. Reference Panel of Measured Transcriptome & Genome Variation
RNA Reference
(GTEx, DGN)
Genotype Data
Transcriptome Prediction
B. Genetic Prediction of Gene Expression
GReX
Genetically Regulated Expression
Clinical TRAIT
Other Factors
Trait-altered Component
PrediXcan Application to BioVU
EHR
Genotypes
C. Biobank with Genotype Data
(Whole Genome Sequence or Dense Genotyping Array)
Linked to Electronic Health Records (EHR)
Creation of PredixVU
PredixVU
Gene by Medical Phenome Catalog
D. Results Portal for Query by Gene or Phenome

Figure 5.1. Experimental Pipeline for the Construction of Medical Phenome Catalog. **(A)** We utilized transcriptome data in 44 tissues and whole-genome genotype data derived from the same donors from the GTEx Consortium as a reference resource to build gene expression imputation models. (**B)** We used the PrediXcan method and these prediction models to impute tissue-level gene expression in the BioVU subjects. This imputed expression level is referred to as the Genetically Regulated eXpression (GReX) to distinguish it from measured expression levels. For downstream analysis, we are interested primarily in GReX rather than the trait-altered component of gene expression or other factors (such as environmental regulators). (**C)** Biobank with Genotype Data Linked to EHR. GReX is tested for association with a clinical trait as represented by a phecode. (**D)** This analytic workflow allows us to generate a comprehensive gene-by-medical phenome catalog, consisting of all associations, including direction of effect, between the GReX for each tested gene and each clinical trait in the EHR.

Discovery of Association between GRIK5 GReX and Eye Diseases

Summary results for the initial analysis of GReX (whole blood transcriptome using DGN as reference panel) associations to eye disease phecodes ($p<0.05$) in 5,240 BioVU subjects of European descent (determined using principal components analysis (Price et al., 2006) of genotype data) are illustrated in Fig 5.2A. Reduced GReX of *GRIK5* showed nominally significant associations with 18 of 55 eye phenotypes having at least 20 cases, and with 16 of 39 eye phenotypes having at least 50 cases. We considered the possibility of LD-contaminated associations in the locus. None of the 10 most adjacent genes in each direction showed a similar pattern of eye phenome association; indeed, none of the 4 phenotypes associated at $p<0.001$ at *GRIK5* that was also associated with one of the 20 adjacent genes was an eye phenotype. Using permutation analysis (see Methods) that preserves the correlation among the phecodes and the correlations among genome variants in linkage disequilibrium, we found that the observed number of eye disease associations was significantly greater than expected by chance (permutation $p<10^{-6}$, Fig 5.2A, in 100,000 permutations, 1 gene had significant associations with a maximum of 6 eye phenotypes; no genes showed significant associations with 7 or more eye phenotypes). A combined test statistic ($X_{2k}^2 \sim -2\sum_{i=1}^{k}\ln(p_i)$) using Fisher's method applied to the p-values for *GRIK5* was highly significant (empirical adjusted $p<10^{-6}$) based on a comparison with an empirical frequency distribution of the maximum combined test statistic across all genes tested.

In analyses of the largest sample sizes currently available to us for studies on the association of the reduced GReX of *GRIK5* (~60,000 participants), among those with the lowest *GRIK5* expression (bottom 2.5% of the distribution, Fig 5.2B), there is significantly increased risk of peripheral vascular disease and comorbid eye traits (Fig 5.2C), including “other retinal disorders” (p=5.57e-08), cataract (p=2.41e-06), glaucoma (p=6.26e-06) (Fig 5.2C), consistent with a vascular component to the pathophysiology of these ocular diseases. Notably, we find no associations with infection of the eye (p=0.61) or strabismus and other disorders of binocular eye movements (p=0.22), suggesting a distinct pathophysiology for such eye disorders. These findings led to the fundamental question of the physiological function of *GRIK5*, and how that might relate to potential comorbidity of eye and vascular diseases.

Reduced Expression of grik5 in Zebrafish Leads to Eye Phenotypes and Bleeding

GRIK5 encodes glutamate-binding receptor that has been implicated in regulation of neurotransmission in the brain and sensory organs by animal model research and association studies (Haumann et al., 2017; Pin and Duvoisin, 1995; Young et al., 2011). However, significantly less is known about its other physiological functions. To examine the role of *grik5* during development, we generated loss-of-function models using CRISPR/Cas9 genome editing strategies (knockout, KO) and morpholino oligonucleotide-based (MO) protein knockdown (KD) (Fig. 5.3A and 5.3B). The genetic KO lines of *grik5* offer independent assessment of the phenotypes observed in morphants. We used MO approach because its dose can be titrated (Fig. 5.4), thus testing effects of partially reduced protein levels, which in turn better resemble reduced gene expression in humans. We used the dose of 6 ng of the 5’UTR targeting MO per embryo. Effectiveness of the *grik5* MO to block mRNA translation was confirmed by *in vivo* tests (Fig. 5.5).

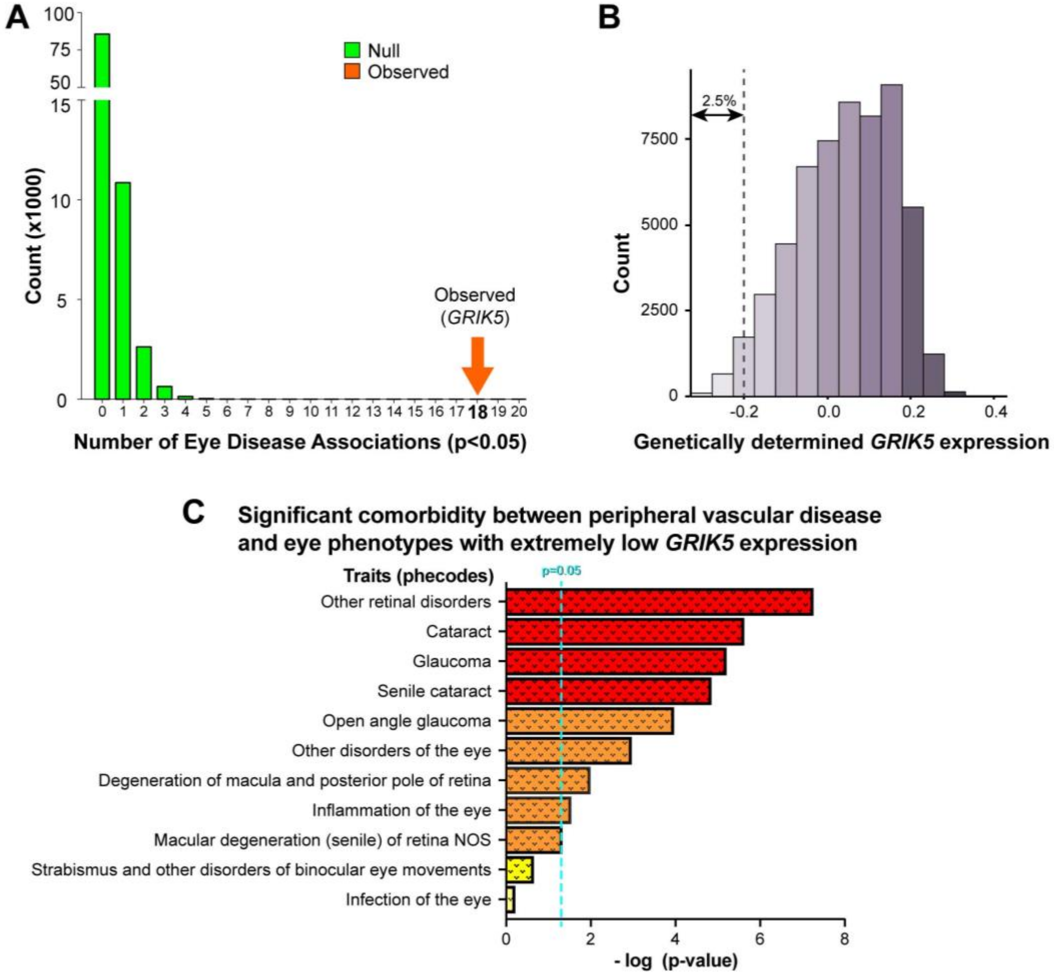


Figure 5.2. *GRIK5* genetically determined expression and association with disease phenome.

(A) The number of associations (p<0.05) of the genetically determined expression of *GRIK5* with disorders of the eye is significantly greater than expected by chance (empirical p<0.001) based on 100,000 permuted datasets that preserve the pairwise SNP-SNP correlations (i.e., linkage disequilibrium) and the pairwise trait-trait correlations. The orange arrow shows the observed number in the actual data while green bars show the null distribution from the permuted datasets. **(B)** Histogram displaying distribution of genetically determined *GRIK5* expression in 60K BioVU subjects. Dotted line indicates the cutoff for the bottom 2.5% (with extreme reduced genetically determined *GRIK5* expression) of the population. **(C)** Comorbidity analysis between peripheral vascular disease and eye phenotypes (traits) in bottom 2.5% of the population shown in B. Dotted line shows p=0.05 cutoff. Yellow: non-significant, orange: nominally significant (p<0.05), red: significant (Bonferroni-adjusted p<0.05) traits. Only traits with >20 cases were included.

Zebrafish *grik5* knockdown with morpholino oligonucleotides and CRISPR induced mutations presented with two visible structural defects: smaller eyes (Fig. 5.6) and defects in blood vessel integrity (Fig. 5.3C and 3E). The small eye phenotype observed in zebrafish is more likely to be related to the expression of *grik5* in the photoreceptor cells and the outer plexiform layer of the retina as shown in mouse embryos(Haumann et al., 2017). Our *in situ* hybridization study also confirmed abundant *grik5* transcripts in the developing zebrafish eye further corroborating the initial findings (Fig. 5.6). Since *grik5* expression might be directly or indirectly linked to metabotropic activity of the glutamate receptor, it would not be surprising that it results in small eye phenotype. The deficits in blood vessel integrity included blood extravasation in various organs at 3 days post fertilization (dpf) (Fig. 5.3C, 5.3D, 5.4, 5.7).

Human and zebrafish *GRIK5/grik5,* respectively, gene products are 85% identical, 94% similar at primary sequence level. Taking advantage of this high homology, we tested the specificity of *grik5* MO knockdown by designing a rescue experiment for the bleeding phenotype with human *GRIK5* (*hGRIK5*) mRNA (non-targetable by *grik5* MO). Overexpression of human *GRIK5* in morphants rescued bleeding to levels comparable with control MO and non-injected siblings (Fig 5.3D and 5.4B) demonstrating that bleeding defects are specific to *grik5* depletion. Hemorrhage in *grik5* morphants was most often visible in the fourth ventricle of the brain and/or retro-pericardial space.

Concurrently, we generated genetic mutations in *grik5* targeting the glutamate binding site by the guide RNA (Fig. 5.3A, 5.3B, 5.3E and 5.8). The founder fish carrying germline *grik5* mutations were sequenced and analyzed in trans-heterozygous F1 progeny

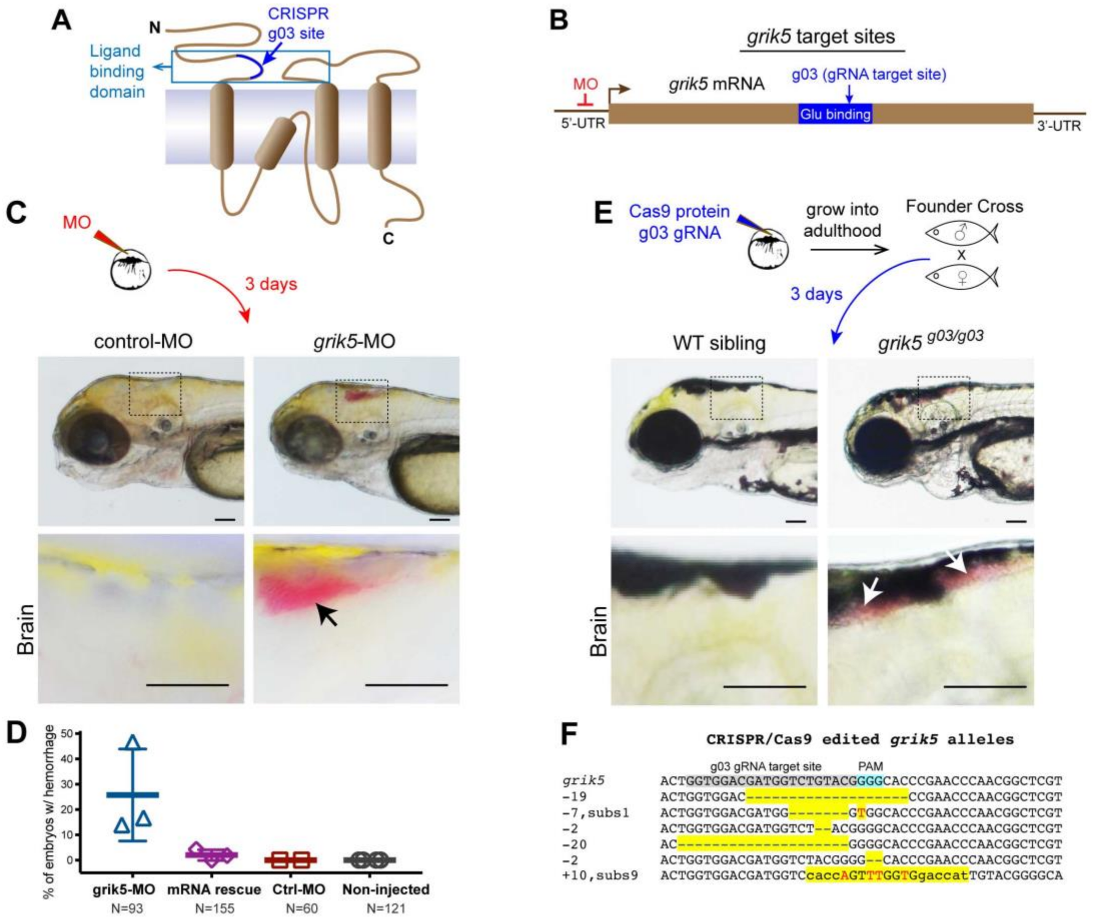

A
N
C
CRISPR
g03 site
Ligand
binding
domain
B
grik5 target sites
MO
grik5 mRNA
g03 (gRNA target site)
Glu binding
5'-UTR
3'-UTR
C
MO
3 days
control-MO
grik5-MO
Brain
D
% of embryos w/ hemorrhage
50
40
30
20
10
0
grik5-MO
N=93
mRNA rescue
N=155
Ctrl-MO
N=60
Non-injected
N=121
E
Cas9 protein
g03 gRNA
grow into
adulthood
Founder Cross
♂
x
♀
3 days
WT sibling
grik5 g03/g03
Brain
F
CRISPR/Cas9 edited grik5 alleles
g03 gRNA target site
PAM
grik5 ACTGGTGGACGATGGTCTGTACGGGGCACCCGAACCCAACGGCTCGT
-19 ACTGGTGGAC-------------------CCGAACCCAACGGCTCGT
-7,subs1 ACTGGTGGACGATGG-------GTGGCACCCGAACCCAACGGCTCGT
-2 ACTGGTGGACGATGGTCT--ACGGGGCACCCGAACCCAACGGCTCGT
-20 AC--------------------GGGGCACCCGAACCCAACGGCTCGT
-2 ACTGGTGGACGATGGTCTACGGGG--CACCCGAACCCAACGGCTCGT
+10,subs9 ACTGGTGGACGATGGTCcaccAGTTTGGTGgaccatTGTACGGGGCA

Figure 5.3. Loss-of-function of *grik5* in zebrafish leads to bleeding phenotype. **(A)** Model of *GRIK5*/GLUK5 protein structure shows predicted glutamate binding pocket targeted by the gRNA g03. **(B)** Position of the morpholino (MO) and guide RNA (gRNA) CRISPR/Cas9 target sites used to generate *grik5*-depletion models. **(C)** Experimental design and lateral views of head regions in live embryos. Brain bleeding is boxed, and zoomed in below, arrow points to the 4th ventricle. **(D)** Summary of percentage of embryos with hemorrhage in MO, h*GRIK5* mRNA rescue and control groups. Human *GRIK5* mRNA non-targetable by zf *grik5*-MO was used in rescue experiments. N: total number of tested animals. Results of independent experiments are indicated with shapes as data points. **(E)** Live images of WT and CRISPR/Cas9-edited *grik5*$^{g03/g03}$ mutants. Zoomed-in images of boxed brain regions are shown below. Arrows mark blood accumulation. **(F)** Sequences of *grik5* alleles. Highlighted regions mark detected mutations in *grik5*^{g03}; dashes (-): deletions; red: substitutions, lowercase: insertions. Scale bars = 0.1 mm in panels **C** & **E**. Examples of pigment blocker (PTU) treated (**C**) and untreated (**E**) embryos.

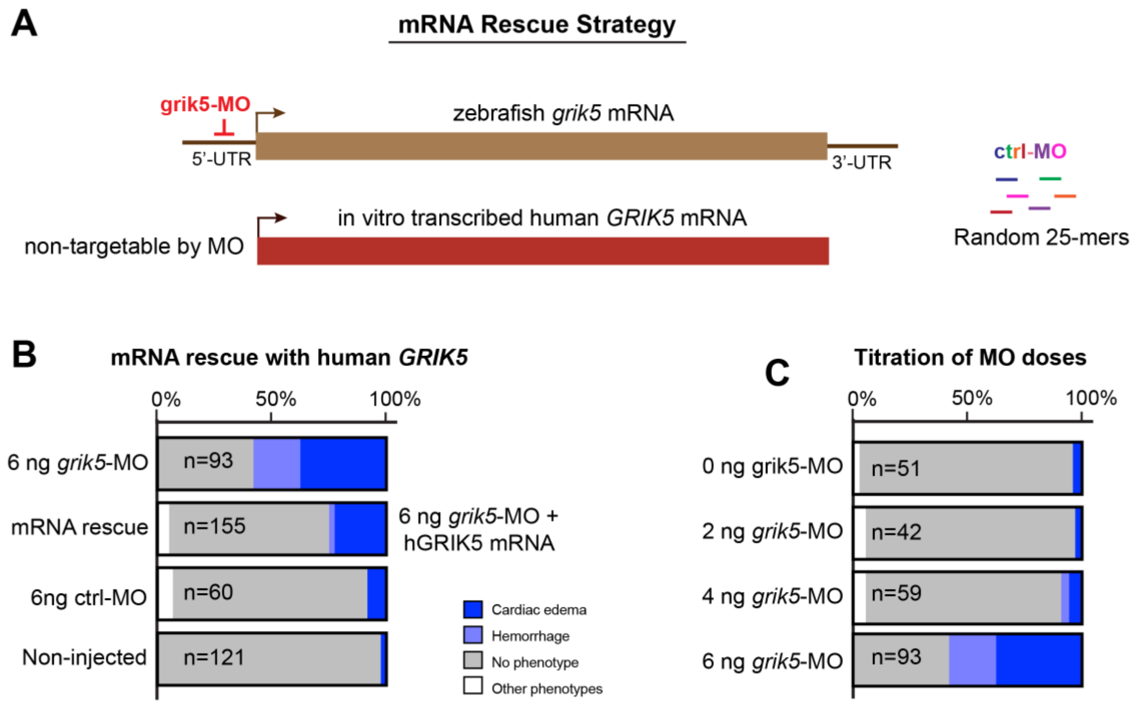


Figure 5.4. mRNA rescue of MO: depletion of zebrafish *grik5* by MO specifically leads to bleeding defects.

(A) Schematic for *grik5* MO knockdown and human *GRIK5* rescue construct. **(B)** Phenotype categories and quantification of bleeding defects in *grik5*-morphants and controls. Replacement of KD zebrafish gene with human *GRIK5* mRNA reduced the number of embryos with bleeding phenotypes to approximately 2.5%. NIC: Non-injected control, ctrl-MO: random 25-mer oligomers. **(C)** Graph showing percentage of embryos with vascular phenotypes under varying *grik5*-MO doses.

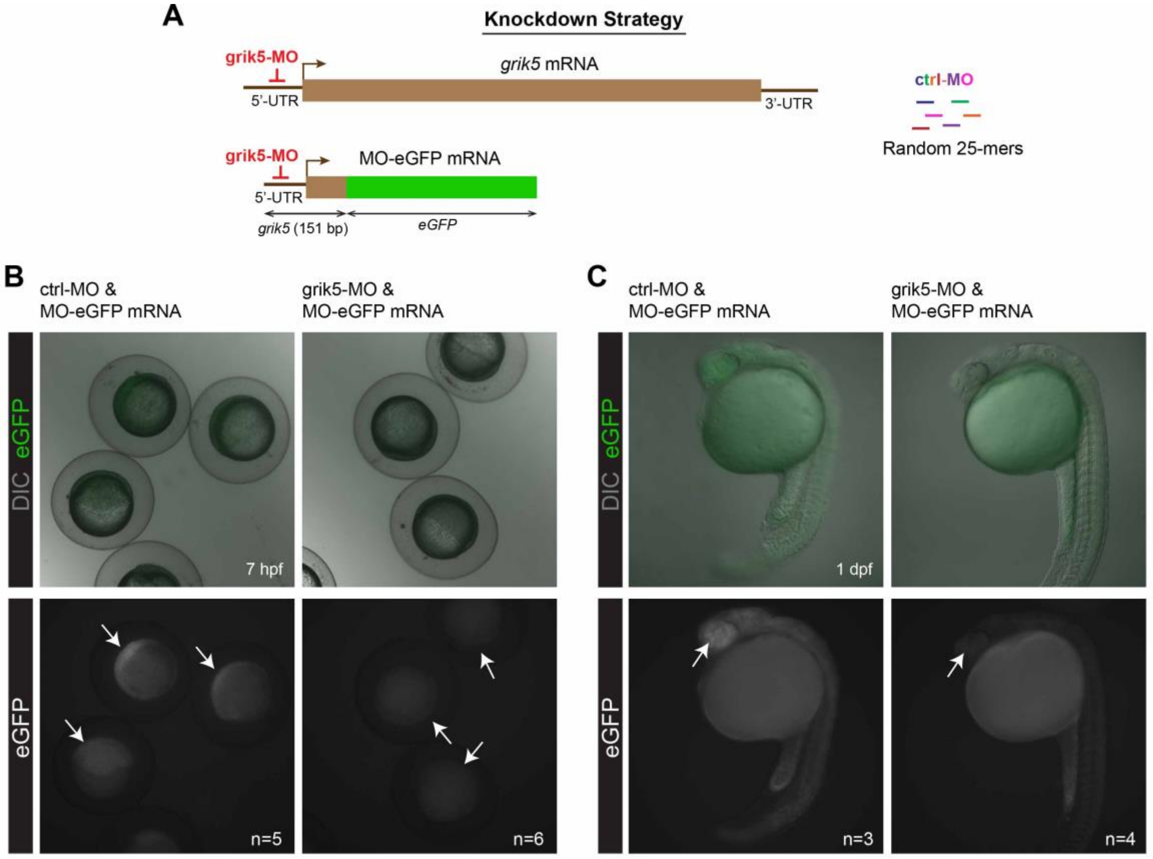


Figure 5.5. MO KD effectiveness.

grik5-MO is effective to block mRNA translation. **(A)** Strategy for MO knockdown and test of its effectiveness. **(B)** DIC and green channel images of 7 hpf zebrafish embryos injected with MO-eGFP mRNA and either ctrl- or *grik5*-MO. Arrows point to animal pole. Note reduced eGFP signal in *grik5*-MO injected group. **(C)** DIC and green channel images of 1 dpf embryos. Reduced eGFP signal is noted in the eye and CNS of *grik5*-MO injected group (arrows).

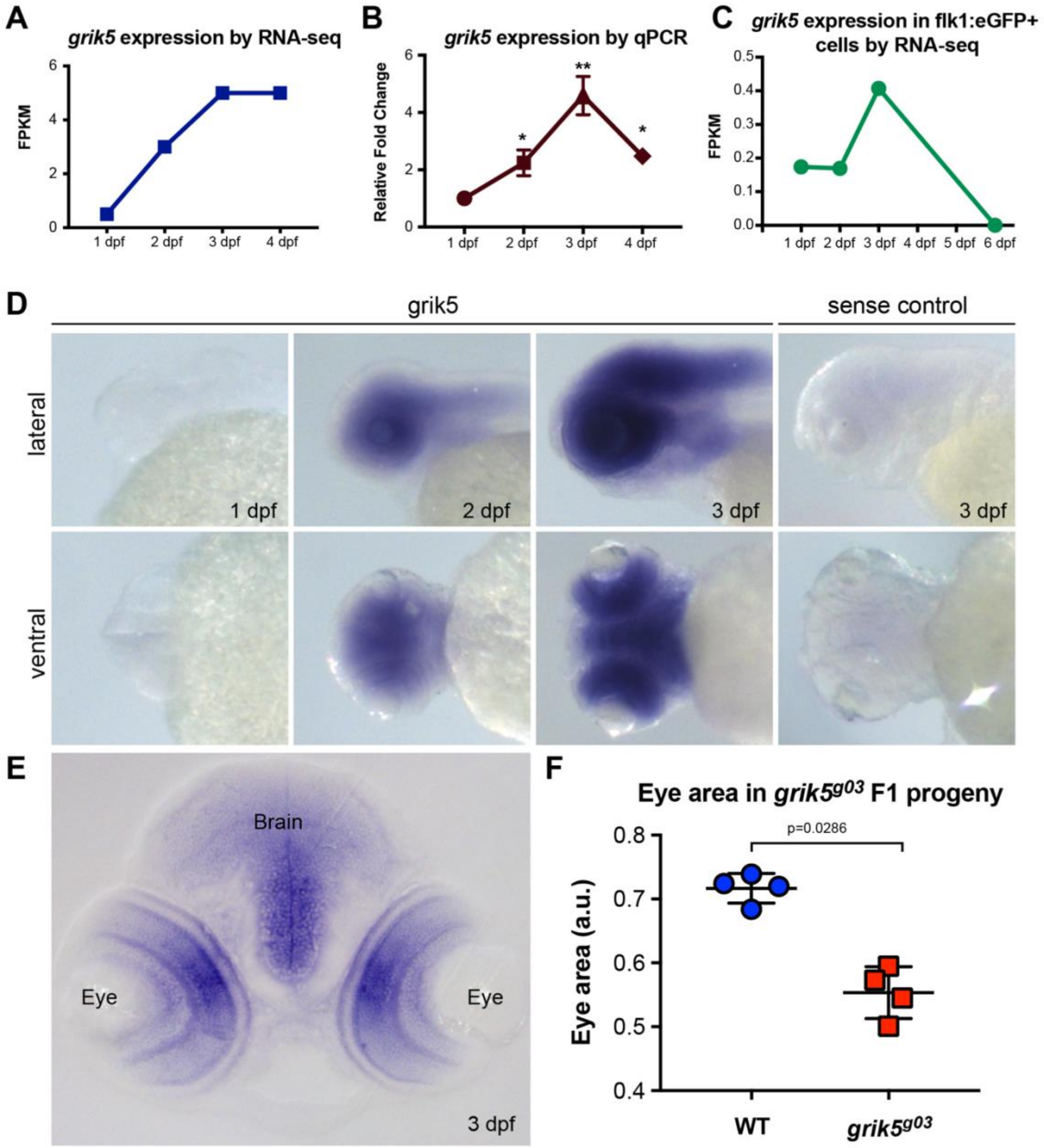


Figure 5.6 Expression analysis of *grik5*.

Expression analysis of *grik5* zebrafish embryos at various developmental stages **(A)** by RNA-seq (data from White RJ et al., 2017) **(B)** by qPCR (Student's t-test, CI = 95%, standard error bars indicated, *p<0.05, **p<0.01) **(C)** in sorted flk1:eGFP+ endothelial cells by RNA-seq (data from Kasper DM et al., 2017) **(D)** *in situ* hybridization with *grik5* riboprobe at 1, 2 and 3 dpf; and sense probe (negative) control. **(E)** Transverse section of 3 dpf embryo showing *grik5* expression in the brain; and the photoreceptor and plexiform layers in the eyes. **(F)** Graph displaying eye area measurements in WT and *grik5*^{g03} mutant embryos, indicating smaller eye areas in mutants (Student's t-test, CI = 95%, standard deviation bars indicated).

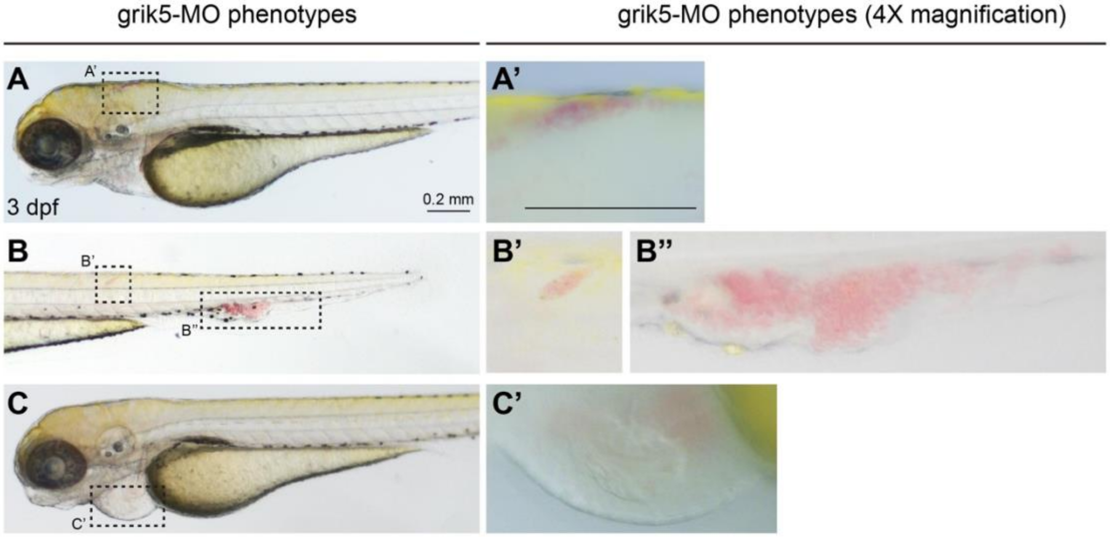


Figure 5.7. Various hemorrhagic foci and cardiac edema detected in *grik5*-MO embryos.

(A) Hemorrhage in the brain. (**A'**) 4X magnified view of the boxed region of in A. **(B)** Hemorrhage in trunk region. 4X zoomed view of hemorrhage in **(B')** somite area and **(B'')** caudal vein. **(C)** Cardiac edema with no noticeable hemorrhage. **(C')** 4X zoomed view of cardiac chamber.

to rule out off-target effects (Fig. 5.3E and 5.3F). F1 animals displayed local hemorrhage in brain and/or in retro-pericardial space in 57 out of 309 analyzed fish (~18%). To ensure that observed bleeding phenotype was specific to mutants, we sequenced individual F1 embryos in a blinded, unbiased manner. We detected deleterious mutations in embryos with hemorrhage, whereas siblings with normal morphology were either heterozygotes or homozygotes wild type for *grik5* (Fig. 5.8).

We reasoned that expression of *grik5* would correlate with the onset of the phenotype. We used publicly available data for whole fish RNA-Seq expression (White et al., 2017), which we confirmed by qPCR (Fig. 5.6A and 5.6B) at 1-4 dpf, and by in situ hybridization at 3 dpf (Fig. 5.6D and 5.6E). Quantitative PCR (qPCR) analysis of whole embryos indicated that *grik5* expression peaked at 3 dpf and localized to the brain and the retina (Fig. 5.6B), as confirmed by an independent set of primers and riboprobes (data not shown). These findings were corroborated by independent RNA-seq data from FACS sorted endothelial cells in zebrafish Tg(flk1:GFP+ cells) at the same developmental stages (Fig. 5.6C) (Kasper et al., 2017). Collectively, *grik5* is expressed at the time and location that could explain the bleeding (Fig. 5.3C and 5.3E) and smaller eye phenotypes (Fig. 5.6F and 5.8) in *grik5* morphants and mutants.

The Comorbidity of Vascular and Eye Disease

The initial and most recent PrediXcan results in BioVU and the zebrafish validation studies combined suggest that defects of early and potentially lifelong vascular conditions (reduced blood vessel number and integrity, increased vascular permeability)

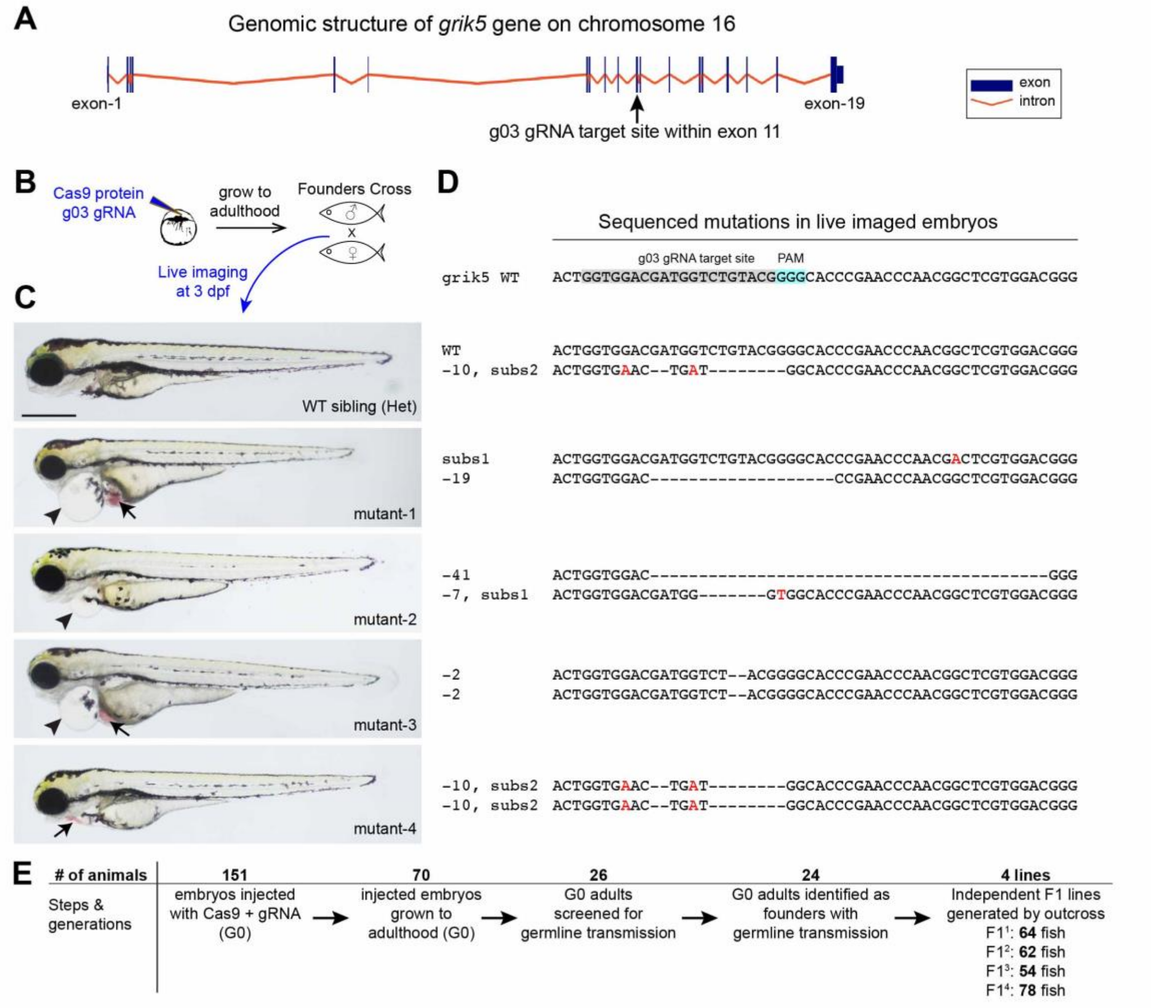
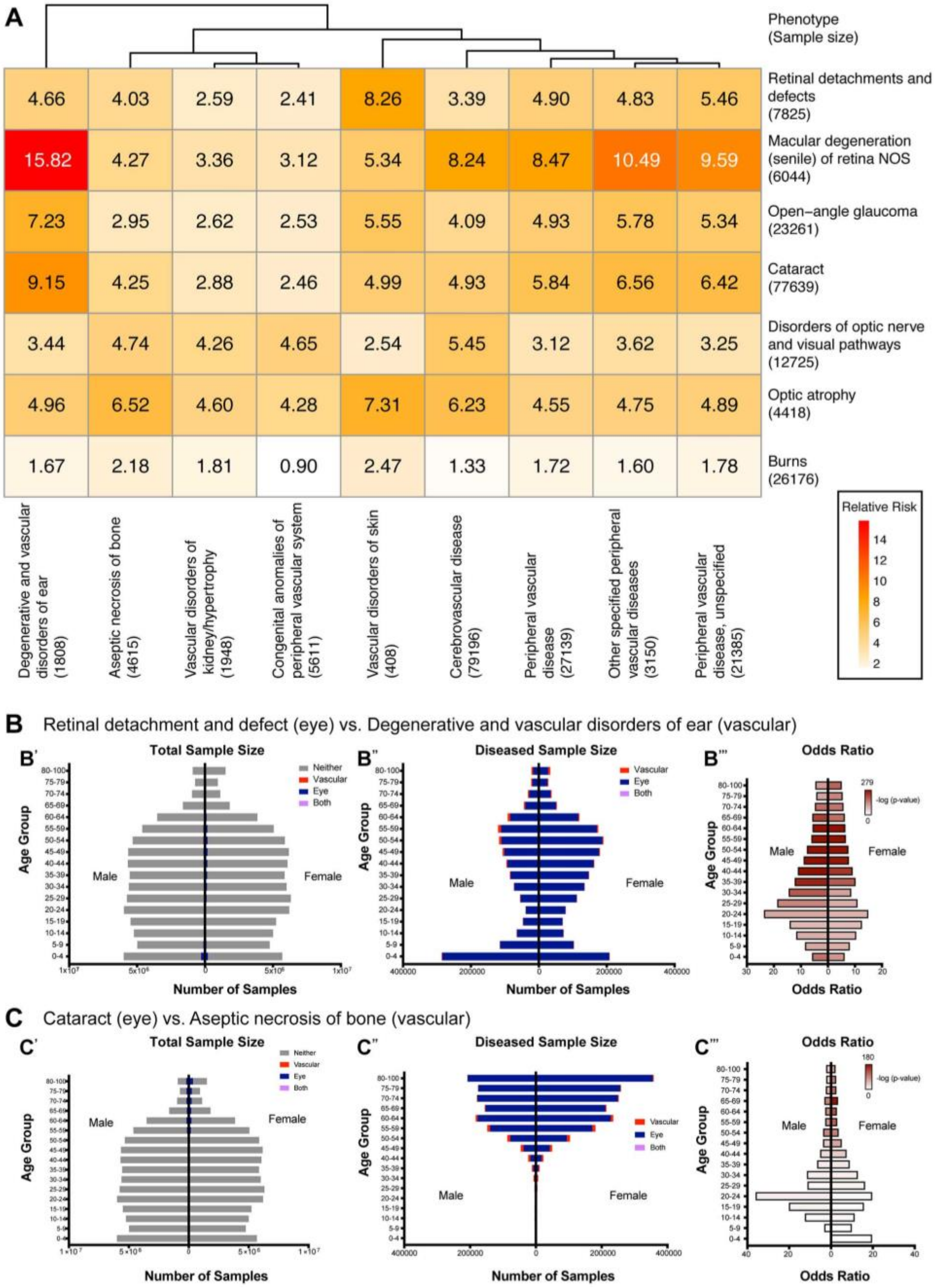

A
Genomic structure of grik5 gene on chromosome 16
exon-1
exon-19
g03 gRNA target site within exon 11
exon
intron
B
Cas9 protein
g03 gRNA
grow to adulthood
Founders Cross
♂
x
♀
Live imaging at 3 dpf
C
WT sibling (Het)
mutant-1
mutant-2
mutant-3
mutant-4
D
Sequenced mutations in live imaged embryos
g03 gRNA target site
PAM
grik5 WT ACTGGTGGACGATGGTCTGTACGGGGCACCCGAACCCAACGGCTCGTGGACGGG
WT ACTGGTGGACGATGGTCTGTACGGGGCACCCGAACCCAACGGCTCGTGGACGGG
-10, subs2 ACTGGTGAAC--TGAT--------GGCACCCGAACCCAACGGCTCGTGGACGGG
subs1 ACTGGTGGACGATGGTCTGTACGGGGCACCCGAACCCAACGACTCGTGGACGGG
-19 ACTGGTGGAC-------------------CCGAACCCAACGGCTCGTGGACGGG
-41 ACTGGTGGAC---GGG
-7, subs1 ACTGGTGGACGATGG-------GTGGCACCCGAACCCAACGGCTCGTGGACGGG
-2 ACTGGTGGACGATGGTCT--ACGGGGCACCCGAACCCAACGGCTCGTGGACGGG
-2 ACTGGTGGACGATGGTCT--ACGGGGCACCCGAACCCAACGGCTCGTGGACGGG
-10, subs2 ACTGGTGAAC--TGAT--------GGCACCCGAACCCAACGGCTCGTGGACGGG
-10, subs2 ACTGGTGAAC--TGAT--------GGCACCCGAACCCAACGGCTCGTGGACGGG
E
# of animals
Steps & generations
151
embryos injected with Cas9 + gRNA (G0)
70
injected embryos grown to adulthood (G0)
26
G0 adults screened for germline transmission
24
G0 adults identified as founders with germline transmission
4 lines
Independent F1 lines generated by outcross
F1¹: 64 fish
F1²: 62 fish
F1³: 54 fish
F1⁴: 78 fish

Figure 5.8. CRISPR/Cas9-mediated generation of *grik5* mutant lines. **(A)** Genomic structure of *grik5* gene (graphics adapted from http://chopchop.cbu.uib.no/). Guide RNA target site is marked by arrow. **(B)** Experimental design showing production of mutant F1 embryos from founder cross. **(C)** Live images of 3 dpf zebrafish embryos reveal hemorrhage (arrow) and cardiac edema (arrowhead) phenotypes in *grik5* mutants. **(D)** Sequence of mutations detected in imaged embryos. -: deletion, subs: substitution, red: substituted nucleotide. **(E)** Number of animals, experimental steps and fish generations produced for *grik5* mutant lines via CRISPR/Cas9 system.

could increase the risk of late-onset eye disease in humans. We therefore hypothesized that eye diseases would be comorbid with vascular diseases. To test this hypothesis, we calculated the relative risk (RR) for eye diseases initially found to be associated with reduced GReX of *GRIK5* in individuals with a set of vascular disease phenotypes that were the most strongly associated with reduced GReX of *GRIK5* in studies on larger numbers of BioVU subjects using Vanderbilt's Synthetic Derivative, the de-identified image of the EHR in > 2.6 million subjects. There are substantial increases in the risk for eye disease across the entire spectrum of vascular phenotypes (2 to 15-fold; Fig. 5.9A). The "Burns" phenotype is included as a control phenotype that has a broad range of age at diagnosis and, while prior vascular disease may complicate healing from a burn, is unlikely to have a major impact on seeking medical treatment for a burn, or vice versa. 72% (39/54) of the eye / vascular phenotype pairs have RR > 4.0. Indeed, the two RR > 4.0 that we did observe had already been noted and reported in the literature (Yu et al., 2014). We similarly estimated the RR (Kleinbaum et al., 1998) for these phenotypes in > 150 million subjects from the MarketScan insurance claims dataset (Blair et al., 2013) (Fig. 5.9B, 5.9C) by age and sex, confirming the magnitude of the comorbidity for eye and vascular disease (see Methods). Age was not significantly associated with the extent of comorbidity in males and females (Kruskal-Wallis test p=0.38 and p=0.70, respectively) and no significant difference was observed between males and females (Mann-Whitney U test p=0.43), demonstrating the robustness of the observation to potential confounding from these demographic variables.



Phenotype (Sample size)	Degenerative and vascular disorders of ear (1808)	Aseptic necrosis of bone (4615)	Vascular disorders of kidney/hypertrophy (1948)	Congenital anomalies of peripheral vascular system (5611)	Vascular disorders of skin (408)	Cerebrovascular disease (79196)	Peripheral vascular disease (27139)	Other specified peripheral vascular diseases (3150)	Peripheral vascular disease, unspecified (21385)
Retinal detachments and defects (7825)	4.66	4.03	2.59	2.41	8.26	3.39	4.90	4.83	5.46
Macular degeneration (senile) of retina NOS (6044)	15.82	4.27	3.36	3.12	5.34	8.24	8.47	10.49	9.59
Open−angle glaucoma (23261)	7.23	2.95	2.62	2.53	5.55	4.09	4.93	5.78	5.34
Cataract (77639)	9.15	4.25	2.88	2.46	4.99	4.93	5.84	6.56	6.42
Disorders of optic nerve and visual pathways (12725)	3.44	4.74	4.26	4.65	2.54	5.45	3.12	3.62	3.25
Optic atrophy (4418)	4.96	6.52	4.60	4.28	7.31	6.23	4.55	4.75	4.89
Burns (26176)	1.67	2.18	1.81	0.90	2.47	1.33	1.72	1.60	1.78

Figure 5.9. Comorbidity Analysis of Eye and Vascular Disorders.
(A) Heatmap shows the relative risk for eye diseases in individuals with a set of vascular disorders. The comorbidity estimates were calculated using Vanderbilt's Synthetic Derivative, the de-identified image of the electronic health records in > 2.6 million participants. There is a two to fifteen-fold increase in risk for eye disease across the set of vascular phenotypes. The "burn" is included as a control phenotype. Notably, 72% of the eye-vascular disease pairings have Relative Risk (RR)>4.0. **(B-C)** Relative risk analysis of

eye and vascular diseases in MarketScan dataset (Blair et al., 2013). **B.** Retinal detachment (eye) and defect vs. Degenerative and vascular disorders of ear (vascular). **(C)** Cataract (eye) vs. Aseptic necrosis of bone (vascular). **(B’-C’)** Total sample sizes represented in each gender and age group. **(B’’-C’’)** Sample sizes include only diseased subjects (excluding subjects with neither eye nor vascular diseases analyzed). **(B’’’-C’’’)** Odds ratios of eye vs. vascular disorders displayed in gender and age groups.

Shared Gene Mechanisms Underlying the Vascular and Eye Phenome

We sought additional support for the hypothesis that vascular and eye diseases share genetic architecture by integrating publicly available GWAS summary data to further investigate pleiotropic genetic effects for eye diseases and vascular traits. Using data from the recent GWAS meta-analysis (n=375,000) of migraine risk (a neurological disorder with a vascular etiology) (Gormley et al., 2016), we found that reduced *GRIK5* GReX (whole blood transcriptome using DGN) was significantly associated with this vascular trait ($p=9x10^{-8}$). Interestingly, we also found a significant comorbidity (RR=2.4, Bonferroni-corrected $p=4.3x10^{-36}$) between migraine and sensory retina dystrophies in the MarketScan dataset. Conversely, we utilized the recent GWAS of age-related macular degeneration, the largest GWAS meta-analyses available for eye disease (Fritsche et al., 2016) (AMD, n>33,000), to test for associations with vascular traits in BioVU (n=28,358). We found significant PrediXcan associations (Benjamini-Hochberg adjusted p<0.05) with "peripheral vascular disease" in BioVU among the top 100 AMD-associated genes. This extends similar results in the most recent 60,000 BioVU sample in which we observed co-morbidity of a vascular trait, cerebrovascular disease, with "macular degeneration (senile) of retina NOS" ($p=9.73x10^{-4}$) among the participants with the lowest *GRIK5* expression (in the bottom 2.5%; Fig. 5.10).

We hypothesized the presence of a shared polygenic component between eye diseases and a wide spectrum of vascular traits. We evaluated the top AMD-associated genetic variants from the meta-analysis (p<0.05) for their association with a variety of vascular traits in BioVU, testing the hypothesis using independent base and target studies. Notably, we found among the AMD-associated variants a highly significant enrichment for low p-values with vascular traits (as illustrated by the significant departure from the null distribution in the Q-Q plots (Fig. 5.11A-D, Kolmogorov Smirnov p<2.2x10^{-16}), including congenital anomalies of peripheral vascular system, aseptic necrosis of bone, and degenerative and vascular disorders of ear, though not with peripheral vascular disease. Using a polygenic empirical test for enrichment (see Methods) that takes into account the linkage disequilibrium (LD) among variants, we continued to observe a highly significant enrichment (empirical p<0.001) for the same set of vascular traits.

We developed a genetic risk score using the effect alleles in the AMD dataset and found a significant association with disease status for each vascular trait in BioVU, including, notably, peripheral vascular disease (p<2.2x10^{-16} for each phenotype). Although the AMD-associated variants were not enriched for the most significant associations with peripheral vascular disease, the significant (Spearman) correlation (p=7.1x10^{-9}) in per-SNP heritability (which is proportional to β^2*p*(1-p) where β is the estimated effect size and *p* is the minor allele frequency in controls) between the traits for the top AMD variants contributes to the predictive performance of the genetic risk score on the vascular trait. Furthermore, the significant comorbidity in the BioVU participants with the lowest genetically determined *GRIK5* expression (in the bottom 2.5%) of “macular degeneration (senile) of retina NOS” with vascular conditions, including cerebrovascular disease (Fig. 5.10), as noted above, and peripheral vascular disease (Fig. 5.2C), suggests that genetic regulation of *GRIK5* contributes to the overall effect of the genetic risk score on these vascular traits.

The availability of the large-scale AMD dataset also allowed us to evaluate the relevance of the imputation models in non-eye tissues for ocular traits. Using a recently

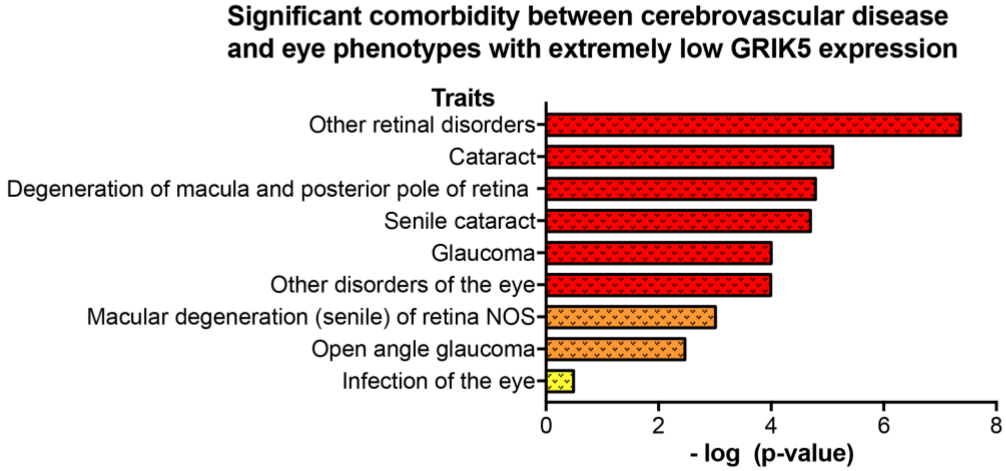


Figure 5.10. Comorbidity analysis between cerebrovascular and eye phenotypes (traits)

Comorbidity analysis between cerebrovascular disease and eye phenotypes (traits) in bottom 2.5% of the population shown in Fig. 5.2B. Yellow: non-significant, orange: nominally significant ($p<0.05$), red: significant (Bonferroni-adjusted $p<0.05$) traits. Only traits with >20 cases were included.

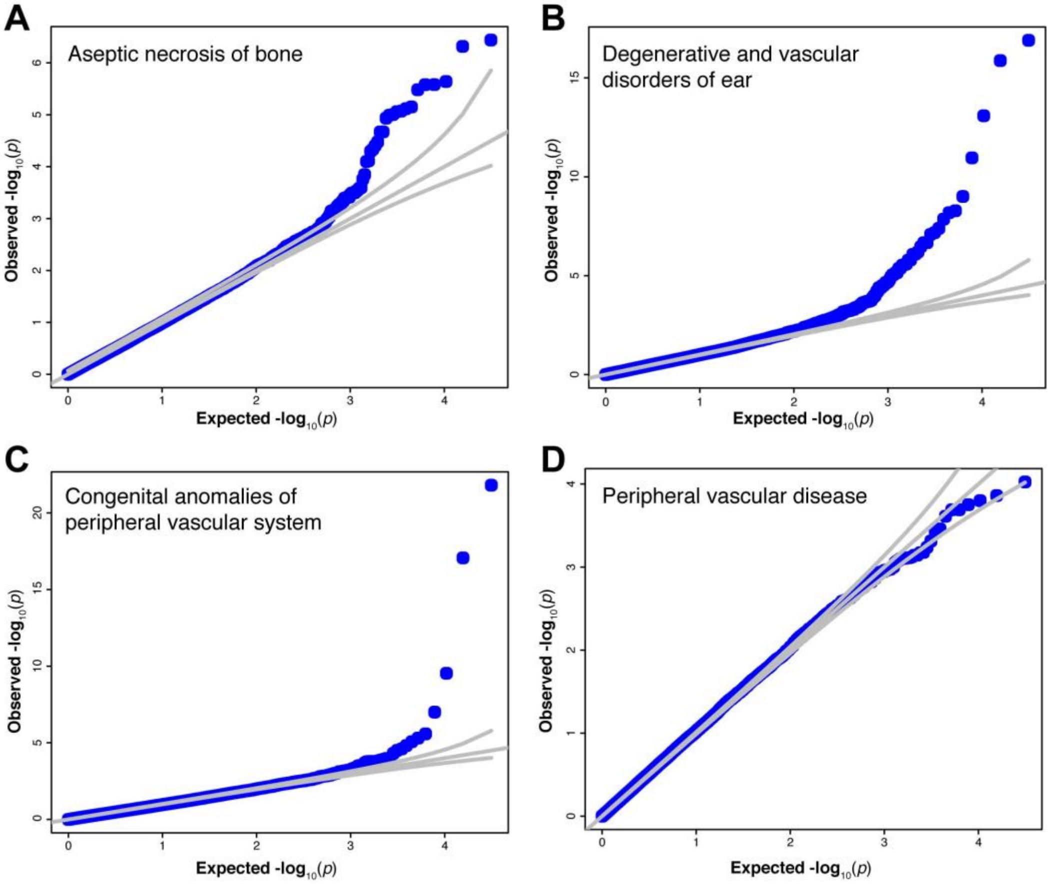


Figure 5.11. Shared Genetic Architecture Underlying Eye Disease Phenome.
Q-Q plots show the distribution of association p-values with the vascular traits for the top age-related macular degeneration (AMD) variants. Using publicly available GWAS of AMD, we find that top AMD variants ($p<0.05$) are significantly enriched for a variety of vascular disorders, including, **(A)** aseptic necrosis of bone, **(B)** degenerative and vascular disorders of ear, and **(C)** congenital anomalies of peripheral vascular system, though not with **(D)** the peripheral vascular disease. Enrichment for association with vascular traits persisted in comparison with null sets (n=1000) of SNPs with matching minor allele frequency (generated from bins of length 5% using the 1000 Genomes EUR samples), distance to nearest gene (using GENCODE gene annotation data), and number of LD partners (r^2>0.50).

developed methodology (Finucane et al., 2015; Gamazon et al., 2018), we found that tissue-shared eQTLs (defined as eCAVIAR posterior probability ≥0.90 in over 90% of the GTEx tissues, i.e., 40 or more) explain a greater proportion (2-fold) of the heritability (h^2) and capture a greater proportion of estimated true positive rate (0.11 vs 0.05) of associations for AMD than tissue-specific eQTLs (posterior probability ≥0.90 in at most 10% of the tissues, i.e., at most 5 tissues).

Vascular Quality and Number of Vessels are Compromised in grik5 Depleted Animals

Statistical analysis in BioVU and MarketScan EHR data revealed strong comorbidity between eye and vascular traits. Thus, to investigate the effect of *grik5* depletion in vasculature, we examined *grik5* morphant zebrafish embryos displaying hemorrhage, which typically results from disruption of a blood vessel and leakage into interstitial spaces. Diminished integrity of vascular endothelium is the common cause of hemorrhage (Butler et al., 2011; Montero-Balaguer et al., 2009). Thus, we reasoned that *GRIK5* expression is required to support vascular patterning and integrity. To model reduced *grik5* expression we injected MO into transgenic Tg(flk1:EGFP) embryos, which expresses EGFP in vascular endothelial cells (Choi et al., 2007), and analyzed *grik5* morphants' and control embryos' vasculature by confocal imaging and 3D reconstruction (Fig. 5.12A). We observed failure of sprouting in central brain arteries (CtA) and posterior mesencephalic central arteries (PMCtA) in severe *grik5* morphants (Fig. 5.12B and 5.12C). Although eye vasculature was patterned, the vessels of the inner optic circle (IOC) were dilated and/or constricted throughout (Fig. 5.12B', 5.12B'', 5.12C', 5.12C''), suggesting that architectural defects in vasculature were not due to general developmental delay, but rather they indicate diminished vascular quality and endothelial integrity. Similar and consistent vascular defects were observed in *grik5* genetic mutants and, to some extent, in heterozygous siblings (Fig. 5.13), corroborating the link between reduced *grik5* expression and compromised vascular quality.

To directly test functional vascular integrity in *grik5*-depleted embryos, we injected 0.02 µm-size nanobeads into circulation through the common cardinal vein (CCV) at 3 dpf (Fig. 5.12D). Fluorescent nanobeads were used to monitor blood flow within green-labeled vasculature of the Tg(flk1:EGFP) transgenic line. Live embryos were imaged using spinning disk confocal microscopy of the head and the trunk. In control embryos, nanobeads were confined to GFP+ vessels, whereas in *grik5*-morphants beads extravasated from vessels that appear intact (Fig. 5.12E). In the head, we detected nanobead extravasation in the forebrain, eye and ear of *grik5*-morphants (Fig. 5.12E), whereas none of the controls showed extravasation. Similarly, in the trunk, nanobeads extravasated from intersegmental vessels in both morphants and genetic mutants (Fig. 5.12F-I).

In summary, both MO knockdowns and CRISPR/Cas9-mediated knockout strategies produced similar vascular defects in zebrafish embryos supporting the hypothesis that *grik5* is required for vascular integrity in zebrafish (Fig. 5.12J).

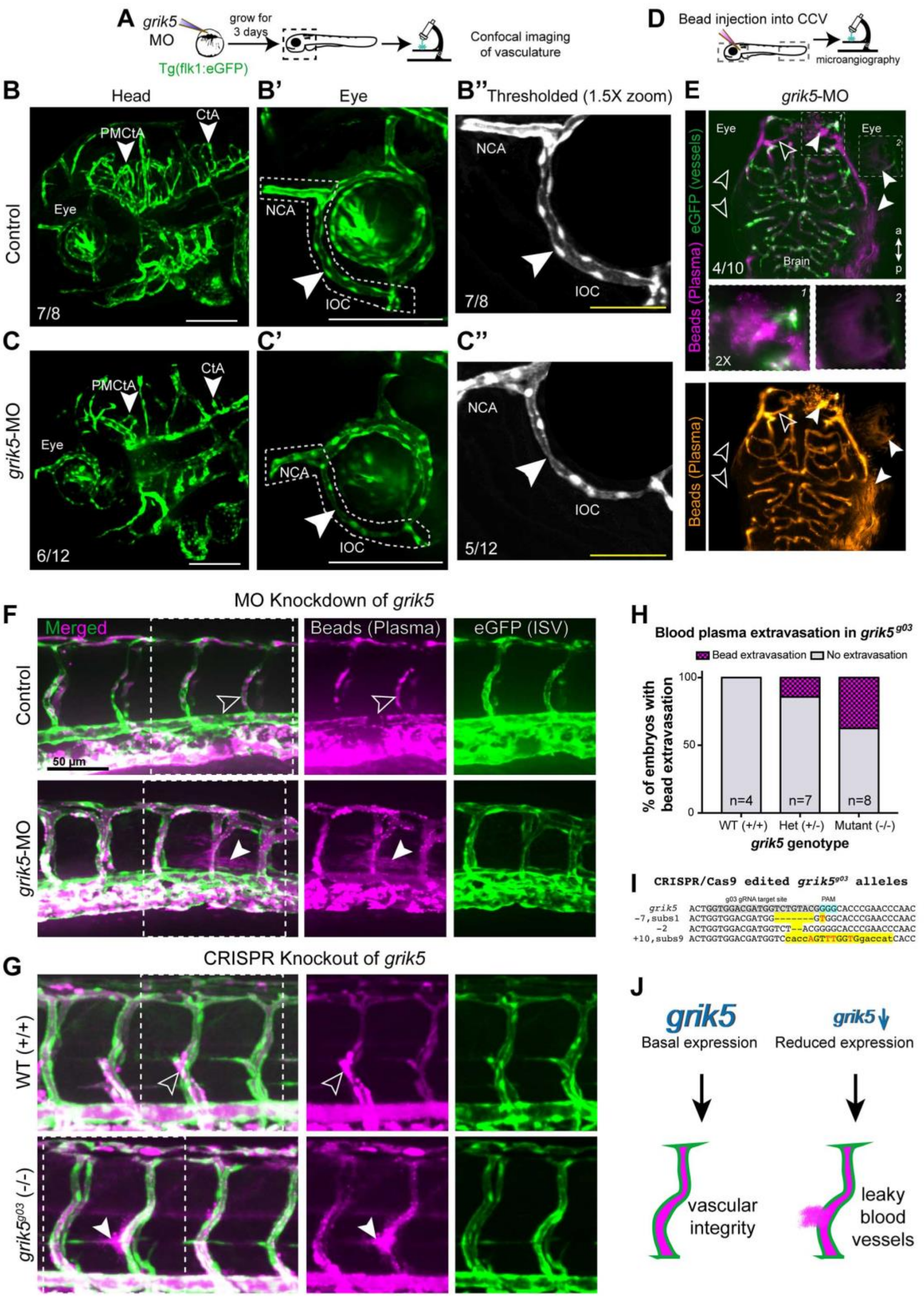


Figure 5.12. Blood plasma extravasation in *grik5*-depleted embryos revealed by nanobeads.

(A) Experimental design and **(B-C)** live, spinning disk confocal images of the lateral head in Tg(flk1:eGFP) show defects in central arteries (CtA) and posterior mesencephalic central arteries (PMCtA) of the brain in **(B)** control and **(C)** *grik5*-KD (3D reconstruction of confocal stacks). **(B'-C')** Confocal images of eye vessels. Thinning of the inner optic circle (IOC) in *grik5*-KD as compared to control (see arrowheads). Digitized (by ImageJ) depiction of comparable IOC vessel thinning area (arrowheads) demarcated in **B''-C''**. NCA: Nasal ciliary artery. **(D)** Experimental design for fluorescent nanobead injection (0.02 µm) and microangiography. **(E)** Maximum intensity projection of the *grik5*-KD's (head, dorsal view) shows plasma leakage around the right anterior cerebral vein (ACeV) (arrowhead), and no leakage on the contralateral site (open arrowhead). Leakage is also detected in the right eye and ear (arrowheads), and absent in contralateral, control site (open arrowheads). 2X zoomed views of boxed regions 1 and 2 are below; beads (plasma) channel shows sites of leakage. a: anterior, p: posterior. **(F, G)** Maximum intensity projection images of trunk intersegmental vessels (ISV) in Tg(flk1:eGFP) embryos (green) injected with fluorescent nanobeads (magenta) in morphants (**F**) and genetic mutants (**G**). Microangiography reveals leakage of blood plasma (nanobeads) from ISVs into interstitial space (arrow). Open arrowhead marks no extravasation in WT (+/+) siblings. (**H**) Graph shows percentage of embryos with nanobeads extravasation, **(I)** the genotype of CRISPR/Cas9 edited *grik5*^{g03} alleles analyzed with microangiography. **(J)** Schematic summarizing the effect of reduced *grik5* expression on vascular integrity. Insets in panels B, C and E indicate the number of embryos exhibiting the represented phenotype out of the total analyzed.

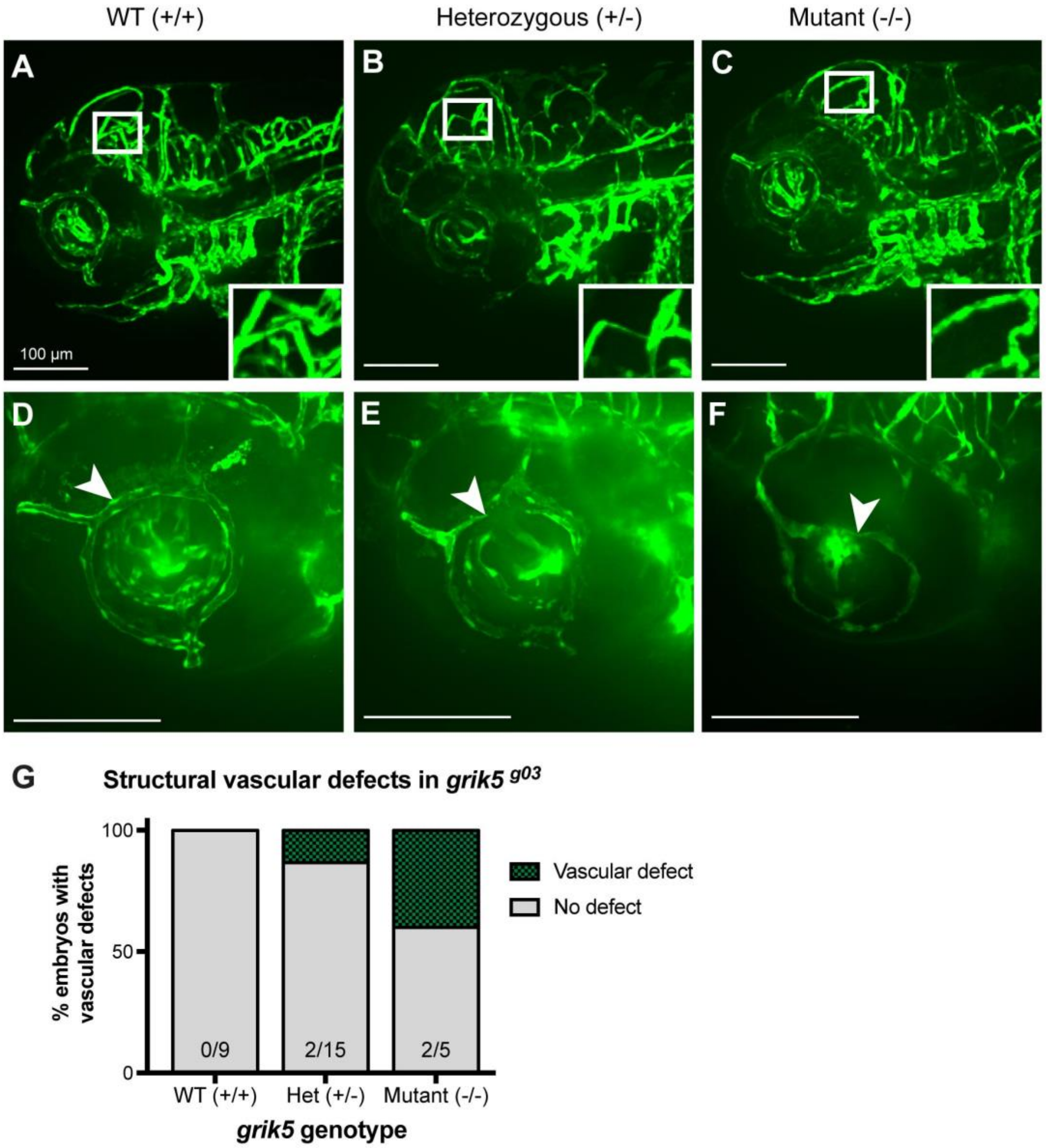


Figure 5.13. Structural vascular defects in *grik5* ^{g03} genetic line.

Live spinning disk confocal images of Tg(flk1:eGFP) transgenic **(A)** homozygous wild-type, +/+; **(B)** heterozygous, +/- and **(C)** mutant, -/-, *grik5*^{g03} embryos. Maximum intensity projection images of the lateral views are displayed. Insets show 2.5X magnified (digital) views of brain vasculature. Eye vasculature of **(D)** homozygous wild-type, +/+; **(E)** heterozygous, +/- and **(F)** mutant, -/-, *grik5*^{g03} embryos imaged at a higher magnification. Maximum intensity projections are displayed. Arrows point to corresponding regions across genotypes. Note thinning of vessels in **E** and **F,** marked by arrows. Panels **B** and **E** represent same embryo, imaged at different magnifications. All others are images of different embryos. **(G)** Graph displaying percentage of embryos with structural vascular defects in each corresponding *grik5*^{g03} phenotype. Number of embryos with vascular defects per total analyzed in each group is indicated within the bars. Het: heterozygous.

Discussion

Figure 5.14 summarizes the paradigm for discovery that we describe here. The overall pattern of results from PrediXcan investigations conducted in BioVU were notable and unusual. Rather than a single significant association of the genetically predicted expression of a gene to a phenotype, we observed the reduced genetically determined expression of a single gene being associated with many different eye diseases not typically considered to have a shared genetic mechanism. The use of a biobank was key to the observation, in that it would otherwise not have been possible to detect the association to so many different eye diseases. It was the unusual nature of this observation that created the impetus to conduct model system validation studies. Zebrafish was a particularly attractive model system in this case because the zebrafish eye is easily studied throughout early development, and modern gene editing technologies facilitate rapid validation studies (Luderman et al., 2017; Unlu et al., 2014; Vacaru et al., 2014). Initial results of studies in zebrafish *GRIK5* knockdown and knockout models highlighted the contribution of this gene to normal vascularization of the eye (both during development and over a lifetime in vascular permeability). The evidence for high expression of *GRIK5* in the developing eye in zebrafish here, and in mouse and squirrel as reported by others (Haumann et al., 2017; Lindstrom et al., 2014), coupled with our data on the function of *grik5* in vascular patterning suggested possible contributing mechanisms to the diverse eye diseases we observed to be associated with the GReX of *GRIK5*. We sought support for this hypothesis by looking for shared genetic architecture between eye and vascular disease, replication of the association of reduced GReX of *GRIK5* with vascular and eye disease in increased

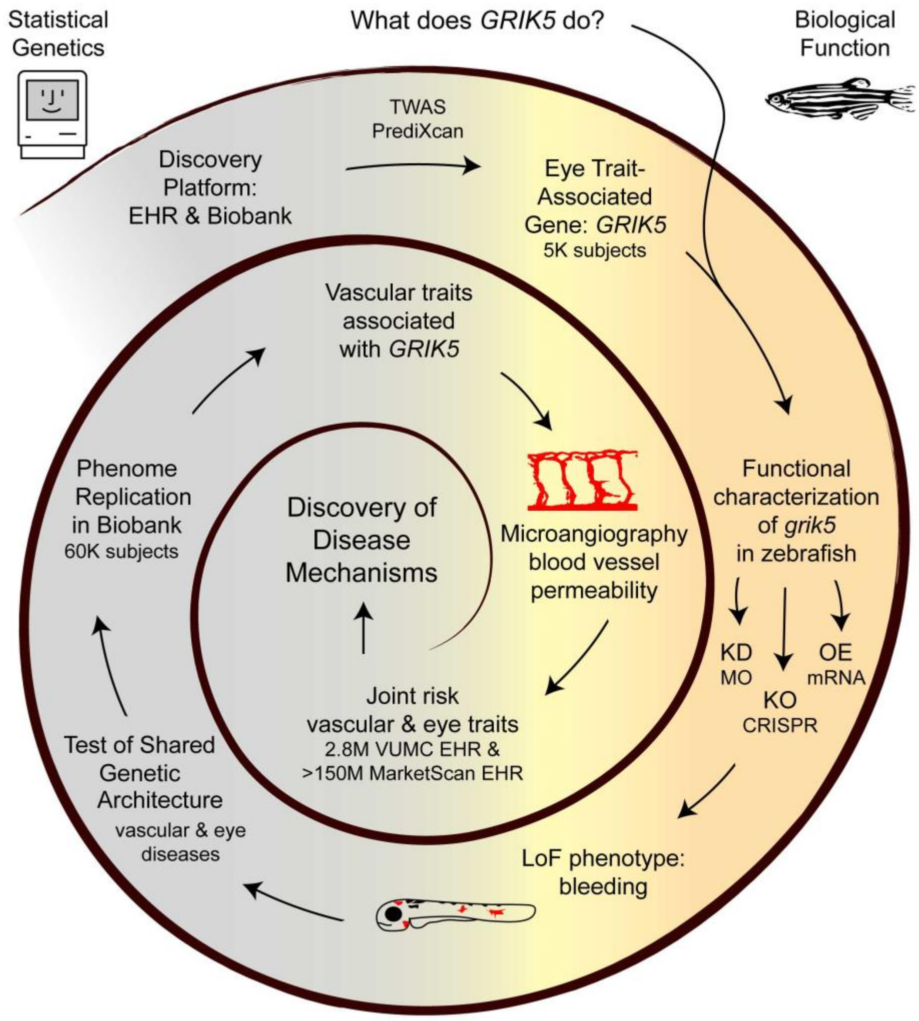

Statistical Genetics
What does *GRIK5* do?
Biological Function
TWAS
PrediXcan
Discovery Platform: EHR & Biobank
Eye Trait-Associated Gene: *GRIK5*
5K subjects
Vascular traits associated with *GRIK5*
Phenome Replication in Biobank
60K subjects
Discovery of Disease Mechanisms
Microangiography blood vessel permeability
Functional characterization of *grik5* in zebrafish
KD
MO
KO
CRISPR
OE
mRNA
Joint risk vascular & eye traits
2.8M VUMC EHR &
>150M MarketScan EHR
Test of Shared Genetic Architecture
vascular & eye diseases
LoF phenotype: bleeding

Figure 5.14. Strategy for the Identification of Disease Mechanisms using EHR & Biobank-based Discovery Platforms and an Animal Model as a Validation Tool.

BioVU was used as an EHR and biobank-based discovery platform. Through the transcriptome-wide association method PrediXcan, trait-associated genes were identified from BioVU. *GRIK5* displayed significant association with numerous diverse eye disease traits and was thus selected for functional validation in animal model, zebrafish. Gain- and loss-of-function approaches such as CRISPR/Cas9-mediated genome editing (KO) and morpholino oligonucleotide-mediated knockdown (KD) led to discovery of phenotypes in zebrafish. These phenotypes further informed statistical analyses using independent phenome datasets (acquired from further biobank genotyping, and publicly available GWAS) to replicated eye and vascular disease phenotypes implicated in zebrafish. System-level assays in zebrafish provided evidence for disease mechanisms, i.e. vascular traits, associated with *GRIK5* expression in participants from BioVU. Subsequent comorbidity studies in all Vanderbilt EHR data and in more than 150 million subjects (Blair et al., 2013) extend the zebrafish findings that eye and vascular diseases are comorbid, and have shared genetic architecture. OE: overexpression sample sizes within BioVU, and for significant comorbidity between vascular and eye disease in large-scale EHR studies.

In these studies, we found that a polygenic risk score built from GWAS meta-analysis results for eye disease (macular degeneration) is significantly associated with vascular disease in BioVU, and further, that vascular and eye diseases are significantly comorbid in not only the > 2.6 million subjects in the Vanderbilt SD, but also in 150 million subjects from a large-scale insurance database. The ability to iterate between the discoveries in the biobank, to the validation and further discoveries in the zebrafish model, and then back to the biobank and the larger EHR datasets for comorbidity studies, and studies of shared genetic architecture between eye and vascular phenotypes confirmative of the zebrafish mechanistic discoveries, is a key feature of our paradigm for discovery.

Our data support the hypothesis that sub-optimal vascularization of the developing human eye might indeed be expected to increase risk of a number of different eye diseases and pathologies in adulthood as a part of a polygenic genetic architecture for such diseases. Maintenance of normal vasculature is also an obvious additional mechanism to investigate as contributory to late onset human eye disease (Yanagi et al., 2011), based on results of these studies. In this regard, the role of the complement system, which has been implicated through genetic studies as contributing to the risk of macular degeneration (Fritsche et al., 2016), in vascular permeability has intriguing parallels to the leaky blood vessels characterized in our *grik5* zebrafish model. As shown through the comorbidity studies on EHR data from very large numbers of individuals, macular degeneration is particularly likely to be comorbid with vascular disease across the board. Such observations may increase the likelihood that the contribution of the complement system to vascular permeability (auf dem Keller et al., 2013) may indeed part of the mechanism by which genetic variation in complement system genes increase risk of macular degeneration. The specificity of the eye and vascular disease comorbidity in electronic health records and confirmation of the eye / vascular disease comorbidity results in the > 150 million individuals from the MarketScan insurance claims database lends further support to the shared genetic component between eye pathologies and vascular traits.

The zebrafish studies provided key information that the zebrafish *grik5* ortholog was highly expressed during early embryonic development (day 3) of the eye, ear and brain, results that have also been observed in the mouse eye (Fukushima et al., 2011; Haumann et al., 2017; Matthaei et al., 2013). Complete or substantial reduction of *grik5* by morpholino or CRISPR-based editing at this time results in notable reductions in blood vessel numbers and in increased vascular permeability. The fact that many of the comorbid vascular phenotypes are very early onset (congenital anomalies of the vasculature or Mendelian diseases present from birth) and that top SNP signals from the large meta-analysis of macular degeneration had such strong enrichment in vascular diseases that substantially predate macular degeneration is consistent with the hypotheses that early developmental and/or subtle lifelong deficits in vascular function can increase the risk of late onset eye diseases.

The variability in expression of *GRIK5* observed in humans is likely to be more modest than the changes we characterized here in zebrafish, and the consequences to the normal vasculature of the eye of reduced expression of *GRIK5* in the ranges predicted for BioVU subjects are likely to be more modest as well. In humans, genetic variation affecting the expression of *GRIK5* would unquestionably be part of the **polygenic architecture** of vascular biology and diseases, and also to eye diseases, as opposed to being a “major gene” capable of causing, by itself, a Mendelian eye or vascular disease, or even a “core gene” in the nomenclature of Boyle et al. (Boyle et al., 2017). Consequently, we did not see, or even expect to see, the equivalent of late-onset human eye disease in zebrafish with reduced expression or knockout of *grik5*. Given the polygenic nature of late onset eye and vascular disease, it was unclear whether we would observe any phenotype at all, and part of the initial focus was more on learning the extent and distribution of *GRIK5* expression in zebrafish.

The observed association of vascular phenotypes with expression of *grik5* in zebrafish ascribes important function to this glutamate receptor that is congruent with what is being learned about multifunctionality of other glutamate receptors. Our findings suggest *grik5* to be linked to patterning and maintenance of the vasculature. In our study of both knockdown and knockout animals, we have discovered that the *grik5*-deficient embryos fail to pattern blood vessels in the brain, resulting in fewer, smaller branches. We also found that the diameter of the blood vessels, as observed in the live embryos in transgenic background, varied and resulted in dilations and constrictions in some regions. Microangiography directly confirmed increased blood vessel permeability in *grik5*-depleted embryos. The observed vascular functions attributed to *GRIK5* expression generated the hypothesis that vascular biology may contribute to the development of late onset eye disease. We showed evidence supporting this hypothesis through both shared genetic architecture and comorbidity studies in large-scale EHR, consistent with the hypothesis that vascular disease is among the contributing factors to the development of late onset human eye disease.

While we are confident that the vascular consequences of reduced expression of *GRIK5* contribute to increased risk for eye disease, we cannot be sure that this is the only mechanism by which reduced expression of *GRIK5* might contribute to eye disease. *GRIK5* is expressed in endothelial cells (Fig. 5.6C) (Fukushima et al., 2011; Kasper et al., 2017; Matthaei et al., 2013), but has also been shown to be expressed in mice in photoreceptor cells of the retina (Karunakaran et al., 2016). We observed *GRIK5* to be highly expressed in the eye of 3-day-old zebrafish embryo (Fig. 5.6). Thus, additional early or later biological consequences of reduced expression of *GRIK5* might also contribute to development of eye disease.

Methodologically, our study presents an efficient framework for discovering gene-disease associations. Indeed, by exploiting the large number of disease phenotypes available in the biobank (1145 phecodes tested), we find, among those with the lowest *GRIK5* expression (i.e., those closest to exhibiting *GRIK5* knockout), significantly increased risk of vascular traits and eye diseases. Thus, conditioning on genetically determined expression and conducting a phenome-wide pleiotropic scan can discover genes associated with the medical phenome. Our study illustrates the methodology, focusing on the comorbidity of the vascular conditions with eye phenotypes because of the number of eye diseases implicated, but the methodology is broadly applicable to the discovery of associations and corresponding mechanisms.

The research questions asked in biobank-based studies are more fundamentally similar to the questions asked in model-system knockout studies – “What does this gene do?”, rather than “What genes and variants contribute to this disease?”; and therefore, has a more natural parallel to the model system studies, leading to more obvious ways of iterating between human and model system studies. While discovery of genes contributing to common human diseases and of the mechanisms by which this genetic variation affects risk of disease will remain a challenging endeavor for the foreseeable future, the combination of modern gene editing technologies coupled to large-scale data integration approaches applied to biobanks offers a model for pursuing this goal.

Web Resources

GTEx, https://gtexportal.org

CHOPCHOP, http://chopchop.cbu.uib.no/

NCBI Gene Portal, https://www.ncbi.nlm.nih.gov/gene/

Ensembl, http://useast.ensembl.org/index.html

Phecodes, https://phewascatalog.org/phecodes

CHAPTER VI

CONCLUDING REMARKS & FUTURE DIRECTIONS

Summary

The research presented here identified novel factors regulating ECM components and their transport during tissue homeostasis and disease conditions. A combination of powerful genetic and imaging approaches led to the discovery of novel cellular and molecular mechanisms regulating ECM deposition during embryonic development in zebrafish. Collaboration with clinical researchers and statistical geneticists paved the way to translate our findings from zebrafish to human, presenting a new way of conducting translational research at the junction of animal models, statistical genetics using big data and clinical investigation.

Focusing on embryonic zebrafish chondrocytes, a highly secretory cell-type, provided a great opportunity to maximize available experimental tools to study ECM secretory mechanisms. Capitalizing on transgenic labeling of chondrocytes, my studies established the first *in vivo* zebrafish chondrocyte transcriptome (Chapter II), which showed enriched RNA levels encoding ECM components and the trafficking machinery required for their secretion. Analyzing genes with enriched expression levels in zebrafish chondrocytes, we found that transcription factor *creb3L2* and COPII vesicle components *sec23a* and *sec24d* are overrepresented in chondrocytes and have similar expression profiles throughout zebrafish developmental stages. This observation established the basis of studies in Chapter III, which provide evidence that Creb3L2, indeed, activates Sec24d expression by directly binding to its promoter in zebrafish chondrocytes. Follow-up studies demonstrated that the “Creb3L2-Sec23a-Sec24d” secretory module acts cell-autonomously to modulate efficient collagen transport in mature chondrocytes.

I took advantage of powerful, unbiased forward genetics screens to identify a novel trafficking axis regulating post-Golgi transport of collagen, presented in Chapter IV. Studying *round* mutant zebrafish, I identified that disrupted activation of Rab6a by the Ric1-Rgp1 GEF complex led to collagen secretion defects and compromised ECM; resulting in short stature, craniofacial and skeletal defects at the physiological level. In collaboration with the Vanderbilt Genetics Institute (use of PredixVU catalog described in Chapter V) and King Faisal Specialist Hospital and Research Center in Saudi Arabia, we discovered that *RIC1* gene function is also required for craniofacial and dental development in human; and identified a novel *RIC1*-linked congenital syndrome, which we named CATIFA.

Chondrocyte-enriched transcriptome as a powerful tool for cartilage biology

Chapter II of this dissertation research presents an original method to selectively sort zebrafish chondrocytes and identify their transcriptome. Bioinformatics analyses of zebrafish chondrocyte transcriptome and its comparison to human fetal cartilage RNA expression dataset revealed over 200 common, evolutionarily conserved chondrocyte-enriched gene transcripts. Even though the majority of these genes have been well-studied and connected to several skeletal disorders, my dataset suggests that dozens of other genes, whose functions are poorly understood may be involved in cartilage biology and possibly in skeletal disorders. Unbiased genetic screens and Genome Wide Association Studies

(GWAS) to discover causative genes for undiagnosed skeletal disorders may benefit from the chondrocyte-enriched gene list presented here to filter potential candidates.

The chondrocyte sorting strategy we present here will benefit researchers since it is readily adaptable to other organisms used to study chondrocyte biology. This FACS-based method could be modified for epigenetic studies as well as constructing non-coding RNA transcriptome of chondrocytes at various developmental stages and under different experimental conditions.

Transcriptional regulation of trafficking and its relevance to human diseases

Secretory cells like chondrocytes require to produce and secrete a greater load of proteins as they differentiate. During this critical differentiation period, differentiating cells need to ensure availability of secretory machinery required for the transport of tissue-specific cargos. Chapter III provides a great example of such a mechanism in chondrocytes. I found that Creb3L2 cell-autonomously regulates collagen-II secretion in differentiated chondrocytes through activation of the inner COPII components, *sec23a* and *sec24d*. Likewise, other CREB-family transcription factors have been reported to control secretory capacity. For instance, *Drosophila* Creb3A (orthologous to vertebrate Creb3L1 and Creb3L2) was reported to directly activate several secretory pathway components, and its loss caused reduction in protein secretion from salivary gland (Abrams and Andrew, 2005; Fox et al., 2010). CREB3L1, vertebrate ortholog of CrebA, was shown to act in osteoblasts to activate collagen-I (Col1) expression. *Creb3L1*$^{-/-}$ mouse osteoblasts displayed expanded ER morphology and decreased Col1 in bone matrix. Several mutations in human *CREB3L1* gene were linked to osteogenesis imperfecta (OI), a brittle bone disease (Keller et al., 2017;

Lindahl et al., 2018; Symoens et al., 2013). These patients were diagnosed with repeated fractures of bones, low collagen-I expression and low glycosaminoglycan secretion in bone tissue. Taken together, these observations raise the question of whether CREB3L1 might regulate trafficking machinery specifically required for bone matrix components during osteogenesis. Preliminary reports suggest that CREB3L1 may induce SEC24D expression, as tested by luciferase assays (Keller et al., 2018). However, direct binding assays using ChIP analysis and functional in vivo studies will be essential to test the link between CREB3L1 and its direct targets in the secretory pathway. Zebrafish model lends itself as a powerful tool for such studies as we have already shown this model system's utility for functional testing of Creb3L1's close paralog Creb3L2 in zebrafish chondrocytes.

Studies I described in chapter III and earlier work from our lab (Melville et al., 2011) identified mechanistic roles of Creb3L2 in cartilage development and homeostasis. Nonetheless, *CREB3L2* has not been directly associated with a human disease, yet. As discussed above, several mutations in its closest paralog, *CREB3L1*, were associated with OI. In addition, *CREB3L2*'s direct targets, *SEC23A* and *SEC24D*, were linked to skeletal syndromes with similar diagnostic features, including facial dysmorphology and broad ossification defects. All of these skeletal conditions were presented with intracellular collagen accumulation in patients' cells. Given these common clinical and cellular findings, I postulate that mutations affecting *CREB3L2* expression and/or function would likely be a causative factor in yet undiagnosed skeletal syndromes with very similar manifestations to OI. Future clinical genetic studies will test whether CREB3L2 is actually the missing link in yet unexplained etiology of an orphan disease.

BioVU and the use of Electronic Health Records in gene-phenome associations

Chapter IV and V present first uses of PredixVU catalog, a collection of phenome linked to altered 'genetically determined expression' of human genes. Our collaborators in Vanderbilt Genetics Institute applied PrediXcan algorithm (Gamazon et al., 2018; Gamazon et al., 2015) to Vanderbilt University Medical Center's (VUMC) DNA biobank, BioVU that is linked to EHR with detailed phenotype information. Utilizing reference datasets with matching genotype and gene expression data, they were able to impute 'genetically determined expression' of genes by using informative polymorphisms in the genomic sequence (see Chapter V). This algorithm has revolutionized the way EHRs have been used since we, only now, have gained the ability to link a gene, through its imputed altered expression, to a collection of clinical traits using already genotyped and phenotyped BioVU dataset. Currently, BioVU sample size is >60,000. With only ~10,000 BioVU samples (our initial scan), we were able to assign *RIC1*-linked phenotypes, which were corroborated in monogenic CATIFA subjects carrying biallelic *RIC1* mutation (chapter IV). In future, increased sample size will greatly increase predictive power of PredixVU.

Most traits categorized by PredixVU, which are derived from phecodes (Denny et al., 2010), are by nature polygenic, i.e. resulted from changes in multiple genes. In addition, several genes have pleiotropic effects, thus contributing to occurrence of several traits. Due to these complicated gene-phenome interactions, animal model-driven studies, like the one presented in Chapter IV, will be instrumental to filter and prioritize PredixVU-generated profiles to select for simple gene-phenotype relationships. As an example, we took advantage of zebrafish loss-of-*ric1* phenotypes that included lack of pharyngeal teeth to sort RIC1-linked phenome and prioritize corresponding traits such as disturbances of tooth eruption in BioVU. This observation was verified in monogenic CATIFA patients who exhibit tooth eruption problems. Hence, we constructed a rather linear relationship between *RIC1* and tooth development. However, *RIC1*-linked phenome contained many more multi-system traits such as ADHD, asthma and pervasive developmental disorders. These traits are recapitulated in CATIFA subjects; however, are more complex to model and study in model organisms due to their polygenic nature. Overall, combinatory approach utilizing animal model systems and PredixVU provide otherwise inaccessible entry points into explaining polygenic traits, and maybe in future, will help obtain a list of associated genes for polygenic traits such as asthma or ADHD.

PredixVU catalog provides bidirectional information in that it is possible inquire (1) phenotypes (i.e. traits) associated with a single gene, and (2) genes associated with a set of various phenotypes. Chapter IV presents an example of the utility of option (1). In chapter V, we embarked on a project that tested option (2), specifically, we explored whether a plethora of eye phenotypes correlating with reduced genetically regulated expression of a single gene. This approach led to discover biologically relevant function of a known gene, *GRIK5*, in the context of comorbid eye and vascular diseases, through combined use of modeling in zebrafish, statistical analysis of electronic health records and PredixVU.

Future implications and remaining questions

In this dissertation, I described collagen trafficking pathways I identified throughout my thesis studies. My projects led me to both ER-to-Golgi (i.e. Creb3L2-Sec23a-Sec24d) and post-Golgi (i.e. Ric1-Rgp1) routes of collagen transport. Even though blockage of collagen secretion at either site leads to overall similar phenotypes such as craniofacial malformations and shortened body length at the organismal level, cellular phenotypes differ significantly. Particularly, *ric1* or *rgp1*-depletion leads to constricted cell shape, while *creb3l2, sec23a* or *sec24d* depleted chondrocytes show no sign of constriction. This observation suggests to me that *ric1-rgp1* complex may regulate cargo trafficking pathways essential for cell growth and cell shape maintenance, in addition to collagen transport. However, those cargos regulating cell shape could be transported out of ER in a Sec23a-Sec24d independent manner. Identification of specific cargos and trafficking machineries involved in cell growth and shape maintenance during cartilage development will help advance our knowledge on how cells acquire their characteristic shape. More specifically, these studies could reveal versatile roles of the secretory machinery in not only transporting cargos to specific locations, but also shaping cellular architecture.

Both zebrafish and patient-derived dermal fibroblast studies suggested that Ric1-Rgp1 complex functions at TGN to facilitate procollagen transport. This finding will provide entry points to studies of post-Golgi procollagen trafficking and potentially may lead to discoveries of further steps in the secretory pathway, e.g. exocytic pathways. We still do not know whether collagen is transported in vesicular carriers out of Golgi? If so, what are the vesicle coats, small GTPases, tethers, membrane docking components used for procollagen trafficking? I am interested to see future studies answering these questions. Robust, *in vivo* models such as zebrafish chondrocytes and notochord sheath cells could provide new entry points into discovery of collagen secretion.

List of other contributed studies

Cox NJ, **Unlu G**, Bisnett BJ, Meister TR, Condon B, Luo PM, Smith TJ, Hanna M, Chhetri A, Soderblom E, Audhya A, Knapik EW, Boyce M (2017) Dynamic glycosylation of COPII proteins governs vertebrate vesicle trafficking,

Biochemistry, 10.1021/acs.biochem.7b00870

Abstract: The COPII coat complex, which mediates secretory cargo trafficking from the endoplasmic reticulum, is a key control point for subcellular protein targeting. Because misdirected proteins cannot function, protein sorting by COPII is critical for establishing and maintaining normal cell and tissue homeostasis. Indeed, mutations in COPII genes cause a range of human pathologies, including cranio-lenticulo-sutural dysplasia (CLSD), which is characterized by collagen trafficking defects, craniofacial abnormalities, and skeletal dysmorphology. Detailed knowledge of the COPII pathway is required to understand its role in normal cell physiology and to devise new treatments for disorders in which it is disrupted. However, little is known about how vertebrates dynamically regulate COPII activity in response to developmental, metabolic, or pathological cues. Several COPII proteins are modified by O-linked β-N-acetylglucosamine (O-GlcNAc), a dynamic form of intracellular protein glycosylation, but the biochemical and functional effects of these modifications remain unclear. Here, we use a combination of chemical, biochemical, cellular, and genetic approaches to demonstrate that site- specific O-GlcNAcylation of COPII proteins mediates their protein−protein interactions and modulates cargo secretion. In particular, we show that individual O-GlcNAcylation sites of SEC23A, an essential COPII component, are required for its function in human cells and vertebrate development, because mutation of these sites impairs SEC23A-dependent in vivo collagen trafficking and skeletogenesis in a zebrafish model of CLSD. Our results indicate that O-GlcNAc is a conserved and critical regulatory modification in the vertebrate COPII-dependent trafficking pathway.

Hockman D, Burns AJ, Fisher S, Schlosser G, Gates KP, Jevans B, Mongera A, Fisher S, **Unlu G**, Knapik EW, Kaufman CK, Mosimann C, Zon LI, Luncman JJ, Dong PDS, Lickert H, Tucker AS, Baker CVH (2017) Evolution of the hypoxia-sensitive cells involved in amniote respiratory reflexes,

eLife, 10.7554/eLife.21231

Abstract: The evolutionary origins of the hypoxia-sensitive cells that trigger amniote respiratory reflexes–carotid body glomus cells, and ‘pulmonary neuroendocrine cells’(PNECs)-are obscure. Homology has been proposed between glomus cells, which are neural crest-derived, and the hypoxia-sensitive ‘neuroepithelial cells’(NECs) of fish gills, whose embryonic origin is unknown. NECs have also been likened to PNECs, which differentiate in situ within lung airway epithelia. Using genetic lineage-tracing and neural crest-deficient mutants in zebrafish, and physical fatemapping in frog and lamprey, we find that NECs are not neural crest-derived, but endodermderived, like PNECs, whose endodermal origin we confirm. We discover neural crest-derived catecholaminergic cells associated with zebrafish pharyngeal arch blood vessels, and propose a new model for amniote hypoxia-sensitive cell evolution: endoderm-derived NECs were retained as PNECs, while the carotid body evolved via the aggregation of neural crest-derived catecholaminergic (chromaffin) cells already associated with blood vessels in anamniote pharyngeal arches.

Luderman LN*, **Unlu G***, Knapik EW (2017) ‘Chapter Three: Zebrafish Developmental Models of Skeletal Diseases’ in Zebrafish at the Interface of Development and Disease Research.

Current Topics in Developmental Biology, 124: 81-124, Edited by Kirsten Sadler Edepli, ISBN: 9780128033081, Academic Press, 10.1016/bs.ctdb.2016.11.004

Abstract: Zebrafish skeleton shares many similarities with human and other vertebrate skeletons. Over the past years the zebrafish model has contributed to understanding basic developmental mechanisms and cellular pathways directing skeletal development and homeostasis. This review will focus on the cell biology of cartilage and bone and how the basic cellular processes within chondrocytes and osteocytes function to assemble the structural frame of a vertebrate body. We will discuss fundamental functions of skeletal cells in production of extracellular matrix (ECM) and cellular activities leading to differentiation of progenitors to mature cells within the organ. We will aim to compare and contrast clinical findings in human skeletal syndromes and zebrafish developmental models to show the utility of zebrafish in unraveling molecular mechanisms of cellular functions necessary to maintain healthy human skeleton.

*co-first author.

Unlu G, Levic DS, Melville DB, Knapik EW (2014) Trafficking mechanisms of extracellular matrix macromolecules: insights from vertebrate development and human diseases.

Int J Biochem Cell Biol **47:** 57-67, 10.1016/j.biocel.2013.11.005

Abstract: Cellular life depends on protein transport and membrane traffic. In multicellular organisms, membrane traffic is required for extracellular matrix deposition, cell adhesion, growth factor release, and receptor signaling, which are collectively required to integrate the development and physiology of tissues and organs. Understanding the regulatory mechanisms that govern cargo and membrane flow presents a prime challenge in cell biology. Extracellular matrix (ECM) secretion remains poorly understood, although given its essential roles in the regulation of cell migration, differentiation, and survival, ECM secretion mechanisms are likely to be tightly controlled. Recent studies in vertebrate model systems, from fishes to mammals and in human patients, have revealed complex and diverse loss-of-function phenotypes associated with mutations in components of the secretory machinery. A broad spectrum of diseases from skeletal and cardiovascular to neurological deficits have been linked to ECM trafficking. These discoveries have directly challenged the prevailing view of secretion as an essential but monolithic process. Here, we will discuss the latest findings on mechanisms of ECM trafficking in vertebrates.

Vacaru AM, **Unlu G**, Spitzner M, Mione M, Knapik EW, Sadler KC (2014) In vivo cell biology in zebrafish - providing insights into vertebrate development and disease.

J Cell Sci **127:** 485-495, 10.1242/jcs.140194

Abstract: Over the past decades, studies using zebrafish have significantly advanced our understanding of the cellular basis for development and human diseases. Zebrafish have rapidly developing transparent embryos that allow comprehensive imaging of embryogenesis combined with powerful genetic approaches. However, forward genetic screens in zebrafish have generated unanticipated findings that are mirrored by human genetic studies: disruption of genes implicated in basic cellular processes, such as protein secretion or cytoskeletal dynamics, causes discrete developmental or disease phenotypes. This is surprising because many processes that were assumed to be fundamental to the function and survival of all cell types appear instead to be regulated by cell-specific mechanisms. Such discoveries are facilitated by experiments in whole animals, where zebrafish provides an ideal model for visualization and manipulation of organelles and cellular processes in a live vertebrate. Here, we review well-characterized mutants and newly developed tools that underscore this notion. We focus on the secretory pathway and microtubule-based trafficking as illustrative examples of how studying cell biology in vivo using zebrafish has broadened our understanding of the role fundamental cellular processes play in embryogenesis and disease.

BIBLIOGRAPHY

Abrams, E.W., Andrew, D.J., 2005. CrebA regulates secretory activity in the Drosophila salivary gland and epidermis. Development 132, 2743-2758.

Albagha, O.M., Wani, S.E., Visconti, M.R., Alonso, N., Goodman, K., Brandi, M.L., Cundy, T., Chung, P.Y., Dargie, R., Devogelaer, J.P., Falchetti, A., Fraser, W.D., Gennari, L., Gianfrancesco, F., Hooper, M.J., Van Hul, W., Isaia, G., Nicholson, G.C., Nuti, R., Papapoulos, S., Montes Jdel, P., Ratajczak, T., Rea, S.L., Rendina, D., Gonzalez-Sarmiento, R., Di Stefano, M., Ward, L.C., Walsh, J.P., Ralston, S.H., 2011. Genome-wide association identifies three new susceptibility loci for Paget's disease of bone. Nat Genet 43, 685-689.

Andreeva, V., Connolly, M.H., Stewart-Swift, C., Fraher, D., Burt, J., Cardarelli, J., Yelick, P.C., 2011. Identification of adult mineralized tissue zebrafish mutants. Genesis 49, 360-366.

Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T., Zheng, N., 2006. Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. Nature 443, 590-593.

Aridor, M., Fish, K.N., Bannykh, S., Weissman, J., Roberts, T.H., Lippincott-Schwartz, J., Balch, W.E., 2001. The Sar1 Gtpase Coordinates Biosynthetic Cargo Selection with Endoplasmic Reticulum Export Site Assembly. J Cell Biol 152, 213-230.

Arnold, W.V., Fertala, A., 2013. Skeletal diseases caused by mutations that affect collagen structure and function. Int J Biochem Cell Biol 45, 1556-1567.

Auer, T.O., Del Bene, F., 2014. CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. Methods.

auf dem Keller, U., Prudova, A., Eckhard, U., Fingleton, B., Overall, C.M., 2013. Systems-level analysis of proteolytic events in increased vascular permeability and complement activation in skin inflammation. Science signaling 6, rs2.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., Fisher, A.G., 2006. Chromatin signatures of pluripotent cell lines. Nat Cell Biol 8, 532-538.

Baines, A.C., Adams, E.J., Zhang, B., Ginsburg, D., 2013. Disruption of the sec24d gene results in early embryonic lethality in the mouse. PLoS One 8, e61114.

Bard, F., Casano, L., Mallabiabarrena, A., Wallace, E., Saito, K., Kitayama, H., Guizzunti, G., Hu, Y., Wendler, F., Dasgupta, R., Perrimon, N., Malhotra, V., 2006. Functional genomics reveals genes involved in protein secretion and Golgi organization. Nature 439, 604-607.

Barlowe, C., 2003. Molecular recognition of cargo by the COPII complex: a most accommodating coat. Cell 114, 395-397.

Barlowe, C., d'Enfert, C., Schekman, R., 1993. Purification and characterization of SAR1p, a small GTP-binding protein required for transport vesicle formation from the endoplasmic reticulum. J Biol Chem 268, 873-879.

Barlowe, C., Orci, L., Yeung, T., Hosobuchi, M., Hamamoto, S., Salama, N., Rexach, M.F., Ravazzola, M., Amherdt, M., Schekman, R., 1994. COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. Cell 77, 895-907.

Barlowe, C., Schekman, R., 1993. SEC12 encodes a guanine-nucleotide-exchange factor essential for transport vesicle budding from the ER. Nature 365, 347-349.

Barrallo-Gimeno, A., Holzschuh, J., Driever, W., Knapik, E.W., 2004. Neural crest survival and differentiation in zebrafish depends on mont blanc/tfap2a gene function. Development 131, 1463-1477.

Bastarache, L., Hughey, J.J., Hebbring, S., Marlo, J., Zhao, W., Ho, W.T., Van Driest, S.L., McGregor, T.L., Mosley, J.D., Wells, Q.S., Temple, M., Ramirez, A.H., Carroll, R., Osterman, T., Edwards, T., Ruderfer, D., Velez Edwards, D.R., Hamid, R., Cogan, J., Glazer, A., Wei, W.Q., Feng, Q., Brilliant, M., Zhao, Z.J., Cox, N.J., Roden, D.M., Denny, J.C., 2018. Phenotype risk scores identify patients with unrecognized Mendelian disease patterns. Science 359, 1233-1239.

Battle, A., Mostafavi, S., Zhu, X., Potash, J.B., Weissman, M.M., McCormick, C., Haudenschild, C.D., Beckman, K.B., Shi, J., Mei, R., Urban, A.E., Montgomery, S.B., Levinson, D.F., Koller, D., 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res 24, 14-24.

Bell, D.M., Leung, K.K., Wheatley, S.C., Ng, L.J., Zhou, S., Ling, K.W., Sham, M.H., Koopman, P., Tam, P.P., Cheah, K.S., 1997. SOX9 directly regulates the type-II collagen gene. Nat Genet 16, 174-178.

Berendsen, A.D., Olsen, B.R., 2015. Bone development. Bone 80, 14-18.

Bhattacharya, N., J, O.D., Stagg, S.M., 2012. The structure of the Sec13/31 COPII cage bound to Sec23. J Mol Biol 420, 324-334.

Bi, X., Corpina, R.A., Goldberg, J., 2002. Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. Nature 419, 271-277.

Bi, X., Mancias, J.D., Goldberg, J., 2007. Insights into COPII coat nucleation from the structure of Sec23.Sar1 complexed with the active fragment of Sec31. Dev Cell 13, 635-645.

Bianchi, P., Fermo, E., Vercellati, C., Boschetti, C., Barcellini, W., Iurlo, A., Marcello, A.P., Righetti, P.G., Zanella, A., 2009. Congenital dyserythropoietic anemia type II (CDAII) is caused by mutations in the SEC23B gene. Hum Mutat 30, 1292-1298.

Bielli, A., Haney, C.J., Gabreski, G., Watkins, S.C., Bannykh, S.I., Aridor, M., 2005. Regulation of Sar1 NH2 terminus by GTP binding and hydrolysis promotes membrane deformation to control COPII vesicle fission. J Cell Biol 171, 919-924.

Blair, D.R., Lyttle, C.S., Mortensen, J.M., Bearden, C.F., Jensen, A.B., Khiabanian, H., Melamed, R., Rabadan, R., Bernstam, E.V., Brunak, S., Jensen, L.J., Nicolae, D., Shah, N.H., Grossman, R.L., Cox, N.J., White, K.P., Rzhetsky, A., 2013. A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. Cell 155, 70-80.

Bogdanovic, O., Fernandez-Minan, A., Tena, J.J., de la Calle-Mustienes, E., Gomez-Skarmeta, J.L., 2013. The developmental epigenomics toolbox: ChIP-seq and MethylCap-seq profiling of early zebrafish embryos. Methods 62, 207-215.

Bonfanti, L., Mironov, A.A., Jr., Martinez-Menarguez, J.A., Martella, O., Fusella, A., Baldassarre, M., Buccione, R., Geuze, H.J., Mironov, A.A., Luini, A., 1998. Procollagen traverses the Golgi stack without leaving the lumen of cisternae: evidence for cisternal maturation. Cell 95, 993-1003.

Bosman, F.T., Stamenkovic, I., 2003. Functional structure and composition of the extracellular matrix. J Pathol 200, 423-428.

Bowen, M.E., Ayturk, U.M., Kurek, K.C., Yang, W., Warman, M.L., 2014. SHP2 regulates chondrocyte terminal differentiation, growth plate architecture and skeletal cell fates. PLoS Genet 10, e1004364.

Boyadjiev, S.A., Fromme, J.C., Ben, J., Chong, S.S., Nauta, C., Hur, D.J., Zhang, G., Hamamoto, S., Schekman, R., Ravazzola, M., Orci, L., Eyaid, W., 2006. Cranio-lenticulo-sutural dysplasia is caused by a SEC23A mutation leading to abnormal endoplasmic-reticulum-to-Golgi trafficking. Nat Genet 38, 1192-1197.

Boyadjiev, S.A., Kim, S.D., Hata, A., Haldeman-Englert, C., Zackai, E.H., Naydenov, C., Hamamoto, S., Schekman, R.W., Kim, J., 2011. Cranio-lenticulo-sutural dysplasia associated with defects in collagen secretion. Clin Genet 80, 169-176.

Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169, 1177-1186.

Braasch, I., Postlethwait, J.H., 2012. Polyploidy in Fish and the Teleost Genome Duplication, in: Soltis, P.S., Soltis, D.E. (Eds.), Polyploidy and Genome Evolution. Springer-Verlag, Berlin Heidelberg, pp. 341-383.

Bradley, K.M., Elmore, J.B., Breyer, J.P., Yaspan, B.L., Jessen, J.R., Knapik, E.W., Smith, J.R., 2007. A major zebrafish polymorphism resource for genetic mapping. Genome Biol 8, R55.

Brandizzi, F., Barlowe, C., 2013. Organization of the ER-Golgi interface for membrane traffic control. Nat Rev Mol Cell Biol 14, 382-392.

Butler, M.G., Gore, A.V., Weinstein, B.M., 2011. Zebrafish as a model for hemorrhagic stroke. Methods in cell biology 105, 137-161.

Cabral, W.A., Makareeva, E., Colige, A., Letocha, A.D., Ty, J.M., Yeowell, H.N., Pals, G., Leikin, S., Marini, J.C., 2005. Mutations near amino end of alpha1(I) collagen cause combined osteogenesis imperfecta/Ehlers-Danlos syndrome by interference with N-propeptide processing. J Biol Chem 280, 19259-19269.

Canty, E.G., Kadler, K.E., 2005. Procollagen trafficking, processing and fibrillogenesis. J Cell Sci 118, 1341-1353.

Charcosset, M., Sassolas, A., Peretti, N., Roy, C.C., Deslandres, C., Sinnett, D., Levy, E., Lachaux, A., 2008. Anderson or chylomicron retention disease: molecular impact of five mutations in the SAR1B gene on the structure and the functionality of Sar1b protein. Mol Genet Metab 93, 74-84.

Chevalier, X., Groult, N., Larget-Piet, B., Zardi, L., Hornebeck, W., 1994. Tenascin distribution in articular cartilage from normal subjects and from patients with osteoarthritis and rheumatoid arthritis. Arthritis Rheum 37, 1013-1022.

Choi, J., Dong, L., Ahn, J., Dao, D., Hammerschmidt, M., Chen, J.N., 2007. FoxH1 negatively modulates flk1 gene expression and vascular formation in zebrafish. Dev Biol 304, 735-744.

Cleutjens, J.P., Verluyten, M.J., Smiths, J.F., Daemen, M.J., 1995. Collagen remodeling after myocardial infarction in the rat heart. Am J Pathol 147, 325-338.

Clore, J.N., Cohen, I.K., Diegelmann, R.F., 1979. Quantitation of collagen types I and III during wound healing in rat skin. Proc Soc Exp Biol Med 161, 337-340.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, 819-823.

Connerly, P.L., Esaki, M., Montegna, E.A., Strongin, D.E., Levi, S., Soderholm, J., Glick, B.S., 2005. Sec16 is a determinant of transitional ER organization. Curr Biol 15, 1439-1447.

Copic, A., Latham, C.F., Horlbeck, M.A., D'Arcangelo, J.G., Miller, E.A., 2012. ER cargo properties specify a requirement for COPII coat rigidity mediated by Sec13p. Science 335, 1359-1362.

Coutinho, P., Parsons, M.J., Thomas, K.A., Hirst, E.M., Saude, L., Campos, I., Williams, P.H., Stemple, D.L., 2004. Differential requirements for COPI transport during vertebrate early development. Dev Cell 7, 547-558.

Cox, N.J., Unlu, G., Bisnett, B.J., Meister, T.R., Condon, B.M., Luo, P.M., Smith, T.J., Hanna, M., Chhetri, A., Soderblom, E.J., Audhya, A., Knapik, E.W., Boyce, M., 2018. Dynamic Glycosylation Governs the Vertebrate COPII Protein Trafficking Pathway. Biochemistry 57, 91-107.

Cutrona, M.B., Beznoussenko, G.V., Fusella, A., Martella, O., Moral, P., Mironov, A.A., 2013. Silencing of mammalian Sar1 isoforms reveals COPII-independent protein sorting and transport. Traffic 14, 691-708.

Dale, R.M., Topczewski, J., 2011. Identification of an evolutionarily conserved regulatory element of the zebrafish col2a1a gene. Dev Biol 357, 518-531.

Dalgleish, R., 1997. The human type I collagen mutation database, Nucleic Acids Res, pp. 181-187.

Davis, E.E., Savage, J.H., Willer, J.R., Jiang, Y.H., Angrist, M., Androutsopoulos, A., Katsanis, N., 2013. Whole exome sequencing and functional studies identify an intronic mutation in TRAPPC2 that causes spondyloepiphyseal dysplasia tarda (SEDT). Clin Genet.

Del Nery, E., Miserey-Lenkei, S., Falguieres, T., Nizak, C., Johannes, L., Perez, F., Goud, B., 2006. Rab6A and Rab6A' GTPases play non-overlapping roles in membrane trafficking. Traffic 7, 394-407.

Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., Basford, M.A., Carrell, D.S., Peissig, P.L., Kho, A.N., Pacheco, J.A., Rasmussen, L.V., Crosslin, D.R., Crane, P.K., Pathak, J., Bielinski, S.J., Pendergrass, S.A., Xu, H., Hindorff, L.A., Li, R., Manolio, T.A., Chute, C.G., Chisholm, R.L., Larson, E.B., Jarvik, G.P., Brilliant, M.H., McCarty, C.A., Kullo, I.J., Haines, J.L., Crawford, D.C., Masys, D.R., Roden, D.M., 2013. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nature biotechnology 31, 1102-1110.

Denny, J.C., Crawford, D.C., Ritchie, M.D., Bielinski, S.J., Basford, M.A., Bradford, Y., Chai, H.S., Bastarache, L., Zuvich, R., Peissig, P., Carrell, D., Ramirez, A.H., Pathak, J., Wilke, R.A., Rasmussen, L., Wang, X., Pacheco, J.A., Kho, A.N., Hayes, M.G., Weston, N., Matsumoto, M., Kopp, P.A., Newton, K.M., Jarvik, G.P., Li, R., Manolio, T.A., Kullo, I.J., Chute, C.G., Chisholm, R.L., Larson, E.B., McCarty, C.A., Masys, D.R., Roden, D.M., de Andrade, M., 2011. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. Am J Hum Genet 89, 529-542.

Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D.R., Roden, D.M., Crawford, D.C., 2010. PheWAS:

demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 26, 1205-1210.

Dephoure, N., Zhou, C., Villen, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., Gygi, S.P., 2008. A quantitative atlas of mitotic phosphorylation. Proc Natl Acad Sci U S A 105, 10762-10767.

Driever, W., Solnica-Krezel, L., Schier, A.F., Neuhauss, S.C., Malicki, J., Stemple, D.L., Stainier, D.Y., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J., Boggs, C., 1996. A genetic screen for mutations affecting embryogenesis in zebrafish. Development 123, 37-46.

Driever, W., Stemple, D., Schier, A., Solnica-Krezel, L., 1994. Zebrafish: genetic tools for studying vertebrate development. Trends Genet 10, 152-159.

Eames, B.F., Singer, A., Smith, G.A., Wood, Z.A., Yan, Y.L., He, X., Polizzi, S.J., Catchen, J.M., Rodriguez-Mari, A., Linbo, T., Raible, D.W., Postlethwait, J.H., 2010. UDP xylose synthase 1 is required for morphogenesis and histogenesis of the craniofacial skeleton. Dev Biol 341, 400-415.

Eames, B.F., Yan, Y.L., Swartz, M.E., Levic, D.S., Knapik, E.W., Postlethwait, J.H., Kimmel, C.B., 2011a. Mutations in fam20b and xylt1 Reveal That Cartilage Matrix Controls Timing of Endochondral Ossification by Inhibiting Chondrocyte Maturation. PLoS Genet 7.

Eames, B.F., Yan, Y.L., Swartz, M.E., Levic, D.S., Knapik, E.W., Postlethwait, J.H., Kimmel, C.B., 2011b. Mutations in fam20b and xylt1 reveal that cartilage matrix controls timing of endochondral ossification by inhibiting chondrocyte maturation. PLoS Genet 7, e1002246.

Editorial, 2018. GWAS to the people. Nature Medicine 24, 1483.

Elliott, L.T., Sharp, K., Alfaro-Almagro, F., Shi, S., L., M.K., Douaud, G., Marchini, J., Smith, S.M., 2018. Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature 562, 210-216.

Espenshade, P., Gimeno, R.E., Holzmacher, E., Teung, P., Kaiser, C.A., 1995. Yeast SEC16 gene encodes a multidomain vesicle coat protein that interacts with Sec23p. J Cell Biol 131, 311-324.

Fath, S., Mancias, J.D., Bi, X., Goldberg, J., 2007. Structure and organization of coat proteins in the COPII cage. Cell 129, 1325-1336.

Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F.R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J.R., Okada, Y., Raychaudhuri, S., Daly, M.J., Patterson, N., Neale, B.M., Price, A.L., 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47, 1228-1235.

Flintoft, L., 2014. Disease genetics: phenome-wide association studies go large. Nature reviews. Genetics 15, 2.

Fornzler, D., Her, H., Knapik, E.W., Clark, M., Lehrach, H., Postlethwait, J.H., Zon, L.I., Beier, D.R., 1998. Gene mapping in zebrafish using single-strand conformation polymorphism analysis. Genomics 51, 216-222.

Forster, D., Armbruster, K., Luschnig, S., 2010. Sec24-dependent secretion drives cell-autonomous expansion of tracheal tubes in Drosophila. Curr Biol 20, 62-68.

Fox, R.M., Hanlon, C.D., Andrew, D.J., 2010. The CrebA/Creb3-like transcription factors are major and direct regulators of secretory capacity. J Cell Biol 191, 479-492.

Francomano, C.A., Liberfarb, R.M., Hirose, T., Maumenee, I.H., Streeten, E.A., Meyers, D.A., Pyeritz, R.E., 1987. The Stickler syndrome: evidence for close linkage to the structural gene for type II collagen. Genomics 1, 293-296.

Franz-Wachtel, M., Eisler, S.A., Krug, K., Wahl, S., Carpy, A., Nordheim, A., Pfizenmaier, K., Hausser, A., Macek, B., 2012. Global detection of protein kinase D-dependent phosphorylation events in nocodazole-treated human cells. Mol Cell Proteomics 11, 160-170.

Fritsche, L.G., Igl, W., Bailey, J.N., Grassmann, F., Sengupta, S., Bragg-Gresham, J.L., Burdon, K.P., Hebbring, S.J., Wen, C., Gorski, M., Kim, I.K., Cho, D., Zack, D., Souied, E., Scholl, H.P., Bala, E., Lee, K.E., Hunter, D.J., Sardell, R.J., Mitchell, P., Merriam, J.E., Cipriani, V., Hoffman, J.D., Schick, T., Lechanteur, Y.T., Guymer, R.H., Johnson, M.P., Jiang, Y., Stanton, C.M., Buitendijk, G.H., Zhan, X., Kwong, A.M., Boleda, A., Brooks, M., Gieser, L., Ratnapriya, R., Branham, K.E., Foerster, J.R., Heckenlively, J.R., Othman, M.I., Vote, B.J., Liang, H.H., Souzeau, E., McAllister, I.L., Isaacs, T., Hall, J., Lake, S., Mackey, D.A., Constable, I.J., Craig, J.E., Kitchner, T.E., Yang, Z., Su, Z., Luo, H., Chen, D., Ouyang, H., Flagg, K., Lin, D., Mao, G., Ferreyra, H., Stark, K., von Strachwitz, C.N., Wolf, A., Brandl, C., Rudolph, G., Olden, M., Morrison, M.A., Morgan, D.J., Schu, M., Ahn, J., Silvestri, G., Tsironi, E.E., Park, K.H., Farrer, L.A., Orlin, A., Brucker, A., Li, M., Curcio, C.A., Mohand-Said, S., Sahel, J.A., Audo, I., Benchaboune, M., Cree, A.J., Rennie, C.A., Goverdhan, S.V., Grunin, M., Hagbi-Levi, S., Campochiaro, P., Katsanis, N., Holz, F.G., Blond, F., Blanche, H., Deleuze, J.F., Igo, R.P., Jr., Truitt, B., Peachey, N.S., Meuer, S.M., Myers, C.E., Moore, E.L., Klein, R., Hauser, M.A., Postel, E.A., Courtenay, M.D., Schwartz, S.G., Kovach, J.L., Scott, W.K., Liew, G., Tan, A.G., Gopinath, B., Merriam, J.C., Smith, R.T., Khan, J.C., Shahid, H., Moore, A.T., McGrath, J.A., Laux, R., Brantley, M.A., Jr., Agarwal, A., Ersoy, L., Caramoy, A., Langmann, T., Saksens, N.T., de Jong, E.K., Hoyng, C.B., Cain, M.S., Richardson, A.J., Martin, T.M., Blangero, J., Weeks, D.E., Dhillon, B., van Duijn, C.M., Doheny, K.F., Romm, J., Klaver, C.C., Hayward, C., Gorin, M.B., Klein, M.L., Baird, P.N., den Hollander, A.I., Fauser, S., Yates, J.R., Allikmets, R., Wang, J.J., Schaumberg, D.A., Klein, B.E., Hagstrom, S.A., Chowers, I., Lotery, A.J., Leveillard, T., Zhang, K., Brilliant, M.H., Hewitt, A.W., Swaroop, A., Chew, E.Y., Pericak-Vance, M.A., DeAngelis, M., Stambolian, D., Haines, J.L., Iyengar, S.K., Weber, B.H., Abecasis, G.R., Heid, I.M., 2016. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nature genetics 48, 134-143.

Fromme, J.C., Orci, L., Schekman, R., 2008. Coordination of COPII vesicle trafficking by Sec23. Trends Cell Biol 18, 330-336.

Fromme, J.C., Ravazzola, M., Hamamoto, S., Al-Balwi, M., Eyaid, W., Boyadjiev, S.A., Cosson, P., Schekman, R., Orci, L., 2007. The genetic basis of a craniofacial disease provides insight into COPII coat assembly. Dev Cell 13, 623-634.

Fukushima, Y., Okada, M., Kataoka, H., Hirashima, M., Yoshida, Y., Mann, F., Gomi, F., Nishida, K., Nishikawa, S., Uemura, A., 2011. Sema3E-PlexinD1 signaling selectively suppresses disoriented angiogenesis in ischemic retinopathy in mice. J Clin Invest 121, 1974-1985.

Funari, V.A., Day, A., Krakow, D., Cohn, Z.A., Chen, Z., Nelson, S.F., Cohn, D.H., 2007. Cartilage-selective genes identified in genome-scale analysis of non-cartilage and cartilage gene expression. BMC Genomics 8, 165.

Gamazon, E.R., Cox, N.J., Davis, L.K., 2014. Structural architecture of SNP effects on complex traits. American journal of human genetics 95, 477-489.

Gamazon, E.R., Segre, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., Ongen, H., Konkashbaev, A., Derks, E.M., Aguet, F., Quan, J., Consortium, G.T., Nicolae, D.L., Eskin, E., Kellis, M., Getz, G., McCarthy, M.I., Dermitzakis, E.T., Cox, N.J., Ardlie, K.G., 2018. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nature genetics 50, 956-967.

Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., GTEx-Consortium, Nicolae, D.L., Cox, N.J., Im, H.K., 2015. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 47, 1091-1098.

Garbes, L., Kim, K., Riess, A., Hoyer-Kuhn, H., Beleggia, F., Bevot, A., Kim, M.J., Huh, Y.H., Kweon, H.S., Savarirayan, R., Amor, D., Kakadia, P.M., Lindig, T., Kagan, K.O., Becker, J., Boyadjiev, S.A., Wollnik, B., Semler, O., Bohlander, S.K., Kim, J., Netzer, C., 2015. Mutations in SEC24D, encoding a component of the COPII machinery, cause a syndromic form of osteogenesis imperfecta. Am J Hum Genet 96, 432-439.

Gaur, T., Lengner, C.J., Hovhannisyan, H., Bhat, R.A., Bodine, P.V., Komm, B.S., Javed, A., van Wijnen, A.J., Stein, J.L., Stein, G.S., Lian, J.B., 2005. Canonical WNT signaling promotes osteogenesis by directly stimulating Runx2 gene expression. J Biol Chem 280, 33132-33140.

Gay, S., Fietzek, P.P., Remberger, K., Eder, M., Kuhn, K., 1975. Liver cirrhosis: immunofluorescence and biochemical studies demonstrate two types of collagen. Klin Wochenschr 53, 205-208.

Gay, S., Muller, P.K., Lemmen, C., Remberger, K., Matzen, K., Kuhn, K., 1976. Immunohistological study on collagen in cartilage-bone metamorphosis and degenerative osteoarthrosis. Klin Wochenschr 54, 969-976.

Gedeon, A.K., Colley, A., Jamieson, R., Thompson, E.M., Rogers, J., Sillence, D., Tiller, G.E., Mulley, J.C., Gecz, J., 1999. Identification of the gene (SEDL) causing X-linked spondyloepiphyseal dysplasia tarda. Nat Genet 22, 400-404.

Gedeon, A.K., Tiller, G.E., Le Merrer, M., Heuertz, S., Tranebjaerg, L., Chitayat, D., Robertson, S., Glass, I.A., Savarirayan, R., Cole, W.G., Rimoin, D.L., Kousseff, B.G., Ohashi, H., Zabel, B., Munnich, A., Gecz, J., Mulley, J.C., 2001. The Molecular Basis of X-Linked Spondyloepiphyseal Dysplasia Tarda. Am J Hum Genet 68, 1386-1397.

Gentili, C., Cancedda, R., 2009. Cartilage and bone extracellular matrix. Curr Pharm Des 15, 1334-1348.

Goldring, M.B., Tsuchimochi, K., Ijiri, K., 2006. The control of chondrogenesis. J Cell Biochem 97, 33-44.

Gormley, P., Anttila, V., Winsvold, B.S., Palta, P., Esko, T., Pers, T.H., Farh, K.H., Cuenca-Leon, E., Muona, M., Furlotte, N.A., Kurth, T., Ingason, A., McMahon, G., Ligthart, L., Terwindt, G.M., Kallela, M., Freilinger, T.M., Ran, C., Gordon, S.G., Stam, A.H., Steinberg, S., Borck, G., Koiranen, M., Quaye, L., Adams, H.H., Lehtimaki, T., Sarin, A.P., Wedenoja, J., Hinds, D.A., Buring, J.E., Schurks, M., Ridker, P.M., Hrafnsdottir, M.G., Stefansson, H., Ring, S.M., Hottenga, J.J., Penninx, B.W., Farkkila, M., Artto, V., Kaunisto, M., Vepsalainen, S., Malik, R., Heath, A.C., Madden, P.A., Martin, N.G., Montgomery, G.W., Kurki, M.I., Kals, M., Magi, R., Parn, K., Hamalainen, E., Huang, H., Byrnes, A.E., Franke, L., Huang, J., Stergiakouli, E., Lee, P.H., Sandor, C., Webber, C., Cader, Z., Muller-Myhsok, B., Schreiber, S., Meitinger, T., Eriksson, J.G., Salomaa, V., Heikkila, K., Loehrer, E., Uitterlinden, A.G., Hofman, A., van Duijn, C.M., Cherkas, L., Pedersen, L.M., Stubhaug, A., Nielsen, C.S., Mannikko, M., Mihailov, E., Milani, L., Gobel, H., Esserlind, A.L., Christensen, A.F., Hansen, T.F., Werge, T., Kaprio, J., Aromaa, A.J., Raitakari, O., Ikram, M.A., Spector, T., Jarvelin, M.R., Metspalu, A., Kubisch, C., Strachan, D.P., Ferrari, M.D., Belin, A.C., Dichgans, M., Wessman, M., van den Maagdenberg, A.M., Zwart, J.A., Boomsma, D.I., Smith, G.D., Stefansson, K., Eriksson, N., Daly, M.J., Neale, B.M., Olesen, J., Chasman, D.I., Nyholt, D.R., Palotie, A., 2016. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nature genetics 48, 856-866.

Gorur, A., Yuan, L., Kenny, S.J., Baba, S., Xu, K., Schekman, R., 2017. COPII-coated membranes function as transport carriers of intracellular procollagen I. J Cell Biol 216, 1745-1759.

Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W.A., Li, R., Manolio, T.A., Sanderson, S.C., Kannry, J., Zinberg, R., Basford, M.A., Brilliant, M., Carey, D.J., Chisholm, R.L., Chute, C.G., Connolly, J.J., Crosslin, D., Denny, J.C., Gallego, C.J., Haines, J.L., Hakonarson, H., Harley, J., Jarvik, G.P., Kohane, I., Kullo, I.J., Larson, E.B., McCarty, C., Ritchie, M.D., Roden, D.M., Smith, M.E., Bottinger, E.P., Williams, M.S., e, M.N., 2013. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genetics in medicine : official journal of the American College of Medical Genetics 15, 761-771.

Grynpas, M.D., Eyre, D.R., Kirschner, D.A., 1980. Collagen type II differs from type I in native molecular packing. Biochim Biophys Acta 626, 346-355.

GTEx_Consortium, 2013. The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580-585.

GTEx_Consortium., 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648-660.

Gusev, A., Lee, S.H., Trynka, G., Finucane, H., Vilhjalmsson, B.J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E., Schizophrenia Working Group of the Psychiatric Genomics, C., Consortium, S.-S., Kahler, A.K., Hultman, C.M., Purcell, S.M., McCarroll, S.A., Daly, M., Pasaniuc, B., Sullivan, P.F., Neale, B.M., Wray, N.R., Raychaudhuri, S., Price, A.L., Schizophrenia Working Group of the Psychiatric Genomics, C., Consortium, S.-S., 2014. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. American journal of human genetics 95, 535-552.

Hall, M.A., Verma, A., Brown-Gentry, K.D., Goodloe, R., Boston, J., Wilson, S., McClellan, B., Sutcliffe, C., Dilks, H.H., Gillani, N.B., Jin, H., Mayo, P., Allen, M., Schnetz-Boutaud, N., Crawford, D.C., Ritchie, M.D., Pendergrass, S.A., 2014. Detection of pleiotropy through a Phenome-wide association study (PheWAS) of epidemiologic data as part of the Environmental Architecture for Genes Linked to Environment (EAGLE) study. PLoS genetics 10, e1004678.

Hancock, J.F., Paterson, H., Marshall, C.J., 1990. A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21ras to the plasma membrane. Cell 63, 133-139.

Hardingham, T., Bayliss, M., 1990. Proteoglycans of articular cartilage: changes in aging and in joint disease. Semin Arthritis Rheum 20, 12-33.

Haumann, I., Junghans, D., Anstotz, M., Frotscher, M., 2017. Presynaptic localization of GluK5 in rod photoreceptors suggests a novel function of high affinity glutamate receptors in the mammalian retina. PLoS One 12, e0172967.

He, H., Dai, F., Yu, L., She, X., Zhao, Y., Jiang, J., Chen, X., Zhao, S., 2002. Identification and characterization of nine novel human small GTPases showing variable expressions in liver cancer tissues. Gene Expr 10, 231-242.

Heckel, D., Brass, N., Fischer, U., Blin, N., Steudel, I., Tureci, O., Fackler, O., Zang, K.D., Meese, E., 1997. cDNA cloning and chromosomal mapping of a predicted coiled-coil proline-rich protein immunogenic in meningioma patients. Hum Mol Genet 6, 2031-2041.

Holmborn, K., Habicher, J., Kasza, Z., Eriksson, A.S., Filipek-Gorniok, B., Gopal, S., Couchman, J.R., Ahlberg, P.E., Wiweger, M., Spillmann, D., Kreuger, J., Ledin, J., 2012. On the roles and regulation of chondroitin sulfate and heparan sulfate in zebrafish pharyngeal cartilage morphogenesis. J Biol Chem 287, 33905-33916.

Hsu, Y.H., Kiel, D.P., 2012. Genome-Wide Association Studies of Skeletal Phenotypes: What We Have Learned and Where We Are Headed. J Clin Endocrinol Metab 97, E1958-1977.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K.D., Irizarry, R.A., Lawrence, M., Love, M.I., MacDonald, J., Obenchain, V., Oles, A.K., Pages, H., Reyes, A., Shannon, P., Smyth, G.K., Tenenbaum, D., Waldron, L., Morgan, M., 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nature methods 12, 115-121.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., Joung, J.K., 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229.

Hynes, R.O., 2009. Extracellular matrix: not just pretty fibrils. Science 326, 1216-1219.

Iolascon, A., Esposito, M.R., Russo, R., 2012. Clinical aspects and pathogenesis of congenital dyserythropoietic anemias: from morphology to molecular approach. Haematologica 97, 1786-1794.

Iolascon, A., Russo, R., Esposito, M.R., Asci, R., Piscopo, C., Perrotta, S., Feneant-Thibault, M., Garcon, L., Delaunay, J., 2010. Molecular analysis of 42 patients with congenital dyserythropoietic anemia type II: new mutations in the SEC23B gene and a search for a genotype-phenotype relationship. Haematologica 95, 708-715.

Ishikawa, Y., Bachinger, H.P., 2013. A molecular ensemble in the rER for procollagen maturation. Biochim Biophys Acta 1833, 2479-2491.

Ishikura-Kinoshita, S., Saeki, H., Tsuji-Naito, K., 2012. BBF2H7-mediated Sec23A pathway is required for endoplasmic reticulum-to-Golgi trafficking in dermal fibroblasts to promote collagen synthesis. J Invest Dermatol 132, 2010-2018.

Iyer, V., Boroviak, K., Thomas, M., Doe, B., Riva, L., Ryder, E., Adams, D.J., 2018. No unexpected CRISPR-Cas9 off-target activity revealed by trio sequencing of gene-edited mice. PLoS Genet 14, e1007503.

Jaager, K., Islam, S., Zajac, P., Linnarsson, S., Neuman, T., 2012. RNA-seq analysis reveals different dynamics of differentiation of human dermis- and adipose-derived stromal stem cells. PLoS One 7, e38833.

Jao, L.E., Wente, S.R., Chen, W., 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proceedings of the National Academy of Sciences of the United States of America 110, 13904-13909.

Jin, L., Pahuja, K.B., Wickliffe, K.E., Gorur, A., Baumgartel, C., Schekman, R., Rape, M., 2012. Ubiquitin-dependent regulation of COPII coat size and function. Nature 482, 495-500.

Jones, B., Jones, E.L., Bonney, S.A., Patel, H.N., Mensenkamp, A.R., Eichenbaum-Voline, S., Rudling, M., Myrdal, U., Annesi, G., Naik, S., Meadows, N., Quattrone, A., Islam, S.A., Naoumova, R.P., Angelin, B., Infante, R., Levy, E., Roy, C.C., Freemont, P.S., Scott, J., Shoulders, C.C., 2003. Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders. Nat Genet 34, 29-31.

Kaiser, C.A., Schekman, R., 1990. Distinct sets of SEC genes govern transport vesicle formation and fusion early in the secretory pathway. Cell 61, 723-733.

Kar, K., Amin, P., Bryan, M.A., Persikov, A.V., Mohs, A., Wang, Y.H., Brodsky, B., 2006. Self-association of collagen triple helic peptides into higher order structures. J Biol Chem 281, 33283-33290.

Karunakaran, D.K.P., Al Seesi, S., Banday, A.R., Baumgartner, M., Olthof, A., Lemoine, C., Măndoiu, II, Kanadia, R.N., 2016. Network-based bioinformatics analysis of spatio-temporal RNA-Seq data reveals transcriptional programs underpinning normal and aberrant retinal development. BMC Genomics 17.

Kasper, D.M., Moro, A., Ristori, E., Narayanan, A., Hill-Teran, G., Fleming, E., Moreno-Mateos, M., Vejnar, C.E., Zhang, J., Lee, D., Gu, M., Gerstein, M., Giraldez, A., Nicoli, S., 2017. MicroRNAs Establish Uniform Traits during the Architecture of Vertebrate Embryos. Dev Cell 40, 552-565.e555.

Keller, R.B., Tran, T.T., Pyott, S.M., Pepin, M.G., Savarirayan, R., McGillivray, G., Nickerson, D.A., Bamshad, M.J., Byers, P.H., 2017. Monoallelic and biallelic CREB3L1 variant causes mild and severe osteogenesis imperfecta, respectively. Genet Med.

Keller, R.B., Tran, T.T., Pyott, S.M., Pepin, M.G., Savarirayan, R., McGillivray, G., Nickerson, D.A., Bamshad, M.J., Byers, P.H., 2018. Monoallelic and biallelic CREB3L1 variant causes mild and severe osteogenesis imperfecta, respectively. Genet Med 20, 411-419.

Kessels, M.Y., Huitema, L.F.A., Boeren, S., Kranenbarg, S., Schulte-Merker, S., van Leeuwen, J.L., de Vries, S.C., 2014. Proteomics Analysis of the Zebrafish Skeletal Extracellular Matrix. PLoS One 9.

Kettleborough, R.N.W., Busch-Nentwich, E.M., Harvey, S.A., Dooley, C.M., de Bruijn, E., van Eeden, F., Sealy, I., White, R.J., Herd, C., Nijman, I.J., Fényes, F., Mehroke, S., Scahill, C., Gibbons, R., Wali, N., Carruthers, S., Hall, A., Yen, J., Cuppen, E., Stemple, D.L., 2013. A systematic genome-wide analysis of zebrafish protein-coding gene function. Nature 496, 494-497.

Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., Kathiresan, S., 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 50, 1219-1224.

Kho, A.N., Pacheco, J.A., Peissig, P.L., Rasmussen, L., Newton, K.M., Weston, N., Crane, P.K., Pathak, J., Chute, C.G., Bielinski, S.J., Kullo, I.J., Li, R., Manolio, T.A., Chisholm, R.L., Denny, J.C., 2011. Electronic medical records for genetic research: results of the eMERGE consortium. Science translational medicine 3, 79re71.

Khoriaty, R., Vasievich, M.P., Ginsburg, D., 2012. The COPII pathway and hematologic disease. Blood 120, 31-38.

Khuansuwan, S., Gamse, J.T., 2014. Identification of differentially expressed genes during development of the zebrafish pineal complex using RNA sequencing. Dev Biol 395, 144-153.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology 14, R36.

Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A., Choi, S.Y., 2011a. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS One 6, e18556.

Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A., Choi, S.Y., 2011b. High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. PLoS One 6.

Kim, S.D., Pahuja, K.B., Ravazzola, M., Yoon, J., Boyadjiev, S.A., Hammamoto, S., Schekman, R., Orci, L., Kim, J., 2012. The [corrected] SEC23-SEC31 [corrected] interface plays critical role for export of procollagen from the endoplasmic reticulum. J Biol Chem 287, 10134-10144.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., Schilling, T.F., 1995. Stages of embryonic development of the zebrafish. Dev Dyn 203, 253-310.

Kimmel, C.B., Miller, C.T., Kruze, G., Ullmann, B., BreMiller, R.A., Larison, K.D., Snyder, H.C., 1998. The shaping of pharyngeal cartilages during early development of the zebrafish. Dev Biol 203, 245-263.

Kimmel, C.B., Miller, C.T., Moens, C.B., 2001. Specification and morphogenesis of the zebrafish larval head skeleton. Dev Biol 233, 239-257.

Kitagawa, H., Uyama, T., Sugahara, K., 2001. Molecular cloning and expression of a human chondroitin synthase. J Biol Chem 276, 38721-38726.

Klarin, D., Emdin, C.A., Natarajan, P., Conrad, M.F., Kathiresan, S., 2017. Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk Factor. Circ Cardiovasc Genet 10.

Klein, C., Gahl, W.A., 2018. Patients with rare diseases: from therapeutic orphans to pioneers of personalized treatments, EMBO Mol Med, pp. 1-3.

Kleinbaum , D., Kupper, L., Nizam , A., Muller, K., 1998. Applied Regression Analysis and Other Multivariable Methods Brooks/Cole, Pacific Grove, CA.

Knapik, E.W., 2000. ENU mutagenesis in zebrafish--from genes to complex diseases. Mamm Genome 11, 511-519.

Knapik, E.W., Goodman, A., Atkinson, O.S., Roberts, C.T., Shiozawa, M., Sim, C.U., Weksler-Zangen, S., Trolliet, M.R., Futrell, C., Innes, B.A., Koike, G., McLaughlin, M.G., Pierre, L., Simon, J.S., Vilallonga, E., Roy, M., Chiang, P.W., Fishman, M.C., Driever, W., Jacob, H.J., 1996. A reference cross DNA panel for zebrafish (Danio rerio) anchored with simple sequence length polymorphisms. Development 123, 451-460.

Knapik, E.W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M.C., Jacob, H.J., 1998. A microsatellite genetic linkage map for zebrafish (Danio rerio). Nat Genet 18, 338-343.

Knudson, C.B., Knudson, W., 2001. Cartilage proteoglycans. Semin Cell Dev Biol 12, 69-78.

Kohl, M., Wiese, S., Warscheid, B., 2011. Cytoscape: software for visualization and analysis of biological networks. Methods Mol Biol 696, 291-303.

Koike, T., Izumikawa, T., Tamura, J., Kitagawa, H., 2009. FAM20B is a kinase that phosphorylates xylose in the glycosaminoglycan-protein linkage region. Biochem J 421, 157-162.

Koreishi, M., Yu, S., Oda, M., Honjo, Y., Satoh, A., 2013. CK2 phosphorylates Sec31 and regulates ER-To-Golgi trafficking. PLoS One 8, e54382.

Kuge, O., Dascher, C., Orci, L., Rowe, T., Amherdt, M., Plutner, H., Ravazzola, M., Tanigawa, G., Rothman, J.E., Balch, W.E., 1994. Sar1 promotes vesicle budding from the endoplasmic reticulum but not Golgi compartments. J Cell Biol 125, 51-65.

Kumkhaek, C., Taylor, J.G., Zhu, J., Hoppe, C., Kato, G.J., Rodgers, G.P., 2008. Fetal haemoglobin response to hydroxycarbamide treatment and sar1a promoter polymorphisms in sickle cell anaemia. Br J Haematol 141, 254-259.

Kwan, K.M., Fujimoto, E., Grabher, C., Mangum, B.D., Hardy, M.E., Campbell, D.S., Parant, J.M., Yost, H.J., Kanki, J.P., Chien, C.B., 2007. The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev Dyn 236, 3088-3099.

Lang, M.R., Lapierre, L.A., Frotscher, M., Goldenring, J.R., Knapik, E.W., 2006. Secretory COPII coat component Sec23a is essential for craniofacial chondrocyte maturation. Nat Genet 38, 1198-1203.

Latimer, A., Jessen, J.R., 2010. Extracellular matrix assembly and organization during zebrafish gastrulation. Matrix Biol 29, 89-96.

Lauwers, E., Jacob, C., Andre, B., 2009. K63-linked ubiquitin chains as a specific signal for protein sorting into the multivesicular body pathway. J Cell Biol 185, 493-502.

Le Bleu, H.K., Kamal, F.A., Kelly, M., Ketz, J.P., Zuscik, M.J., Elbarbary, R.A., 2017. Extraction of high-quality RNA from human articular cartilage. Anal Biochem 518, 134-138.

LeClair, E.E., Mui, S.R., Huang, A., Topczewska, J.M., Topczewski, J., 2009. Craniofacial skeletal defects of adult zebrafish glypican 4 (knypek) mutants. Dev Dyn 238, 2550-2563.

Lee, M.C., Orci, L., Hamamoto, S., Futai, E., Ravazzola, M., Schekman, R., 2005. Sar1p N-terminal helix initiates membrane curvature and completes the fission of a COPII vesicle. Cell 122, 605-617.

Lefebvre, V., Huang, W., Harley, V.R., Goodfellow, P.N., de Crombrugghe, B., 1997. SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. Mol Cell Biol 17, 2336-2346.

Lengfeld, J., Wang, Q., Zohlman, A., Salvarezza, S., Morgan, S., Ren, J., Kato, K., Rodriguez-Boulan, E., Liu, B., 2012. Protein kinase C delta regulates the release of collagen type I from vascular smooth muscle cells via regulation of Cdc42. Mol Biol Cell 23, 1955-1963.

Levic, D.S., Minkel, J.R., Wang, W.D., Rybski, W.M., Melville, D.B., Knapik, E.W., 2015. Animal model of Sar1b deficiency presents lipid absorption deficits similar to Anderson disease. J Mol Med (Berl) 93, 165-176.

Lewallen, E.A., Bonin, C.A., Li, X., Smith, J., Karperien, M., Larson, A.N., Lewallen, D.G., Cool, S.M., Westendorf, J.J., Krych, A.J., Leontovich, A.A., Im, H.J., van Wijnen, A.J., 2016. The synovial microenvironment of osteoarthritic joints alters RNA-seq expression profiles of human primary articular chondrocytes. Gene 591, 456-464.

Li, B., Balasubramanian, K., Krakow, D., Cohn, D.H., 2017. Genes uniquely expressed in human growth plate chondrocytes uncover a distinct regulatory network. BMC Genomics 18, 983.

Li, B., Zhan, X., Wing, M.K., Anderson, P., Kang, H.M., Abecasis, G.R., 2013. QPLOT: A Quality Assessment Tool for Next Generation Sequencing Data. Biomed Res Int 2013.

Li, S.W., Prockop, D.J., Helminen, H., Fassler, R., Lapvetelainen, T., Kiraly, K., Peltarri, A., Arokoski, J., Lui, H., Arita, M., et al., 1995. 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.

Lindahl, K., Astrom, E., Dragomir, A., Symoens, S., Coucke, P., Larsson, S., Paschalis, E., Roschger, P., Gamsjaeger, S., Klaushofer, K., Fratzl-Zelman, N., Kindmark, A., 2018. Homozygosity for CREB3L1 premature stop codon in first case of recessive osteogenesis imperfecta associated with OASIS-deficiency to survive infancy. Bone 114, 268-277.

Lindstrom, S.H., Ryan, D.G., Shi, J., DeVries, S.H., 2014. Kainate receptor subunit diversity underlying response diversity in retinal off bipolar cells. J Physiol 592, 1457-1477.

Loftus, A.F., Hsieh, V.L., Parthasarathy, R., 2012. Modulation of membrane rigidity by the human vesicle trafficking proteins Sar1A and Sar1B. Biochem Biophys Res Commun 426, 585-589.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15, 550.

Luderman, L.N., Unlu, G., Knapik, E.W., 2017. Zebrafish Developmental Models of Skeletal Diseases. Curr Top Dev Biol 124, 81-124.

Lui, W.O., Zeng, L., Rehrmann, V., Deshpande, S., Tretiakova, M., Kaplan, E.L., Leibiger, I., Leibiger, B., Enberg, U., Hoog, A., Larsson, C., Kroll, T.G., 2008. CREB3L2-PPARgamma fusion mutation identifies a thyroid signaling pathway regulated by intramembrane proteolysis. Cancer Res 68, 7156-7164.

Maddirevula, S., Alsahli, S., Alhabeeb, L., Patel, N., Alzahrani, F., Shamseldin, H.E., Anazi, S., Ewida, N., Alsaif, H.S., Mohamed, J.Y., Alazami, A.M., Ibrahim, N., Abdulwahab, F., Hashem, M., Abouelhoda, M., Monies, D., Al Tassan, N., Alshammari, M., Alsagheir, A., Seidahmed, M.Z., Sogati, S., Aglan, M.S., Hamad, M.H., Salih, M.A., Hamed, A.A., Alhashmi, N., Nabil, A., Alfadli, F., Abdel-Salam, G.M.H., Alkuraya, H., Peitee, W.O., Keng, W.T., Qasem, A., Mushiba, A.M., Zaki, M.S., Fassad, M.R., Alfadhel, M., Alexander, S., Sabr, Y., Temtamy, S., Ekbote, A.V., Ismail, S., Hosny, G.A., Otaify, G.A., Amr, K., Al Tala, S., Khan, A.O., Rizk, T., Alaqeel, A., Alsiddiky, A., Singh, A., Kapoor, S., Alhashem, A., Faqeih, E., Shaheen, R., Alkuraya, F.S., 2018. Expanding the phenome and variome of skeletal dysplasia. Genet Med.

Malhotra, V., Erlmann, P., Nogueira, C., 2015. Procollagen export from the endoplasmic reticulum. Biochem Soc Trans 43, 104-107.

Mancias, J.D., Goldberg, J., 2008. Structural basis of cargo membrane protein discrimination by the human COPII coat machinery. Embo j 27, 2918-2928.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., Gilad, Y., 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18, 1509-1517.

Martin, I., Jakob, M., Schafer, D., Dick, W., Spagnoli, G., Heberer, M., 2001. Quantitative analysis of gene expression in human articular cartilage from normal and osteoarthritic joints. Osteoarthritis Cartilage 9, 112-118.

Martinez, O., Schmidt, A., Salamero, J., Hoflack, B., Roa, M., Goud, B., 1994. The small GTP-binding protein rab6 functions in intra-Golgi transport. J Cell Biol 127, 1575-1588.

Matsui, Y., Yasui, N., Ozono, K., Yamagata, M., Kawabata, H., Yoshikawa, H., 2001. Loss of the SEDL gene product (Sedlin) causes X-linked spondyloepiphyseal dysplasia tarda: Identification of a molecular defect in a Japanese family. Am J Med Genet 99, 328-330.

Matsuoka, K., Orci, L., Amherdt, M., Bednarek, S.Y., Hamamoto, S., Schekman, R., Yeung, T., 1998. COPII-coated vesicle formation reconstituted with purified coat proteins and chemically defined liposomes. Cell 93, 263-275.

Matthaei, M., Hu, J., Meng, H., Lackner, E.M., Eberhart, C.G., Qian, J., Hao, H., Jun, A.S., 2013. Endothelial Cell Whole Genome Expression Analysis in a Mouse Model of Early-Onset Fuchs' Endothelial Corneal Dystrophy. Invest Ophthalmol Vis Sci 54, 1931-1940.

McCarthy, D.J., Chen, Y., Smyth, G.K., 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40, 4288-4297.

McCarty, C.A., Chisholm, R.L., Chute, C.G., Kullo, I.J., Jarvik, G.P., Larson, E.B., Li, R., Masys, D.R., Ritchie, M.D., Roden, D.M., Struewing, J.P., Wolf, W.A., e, M.T., 2011. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC medical genomics 4, 13.

Melville, D.B., Knapik, E.W., 2011. Traffic jams in fish bones: ER-to-Golgi protein transport during zebrafish development. Cell Adh Migr 5, 114-118.

Melville, D.B., Montero-Balaguer, M., Levic, D.S., Bradley, K., Smith, J.R., Hatzopoulos, A.K., Knapik, E.W., 2011. The feelgood mutation in zebrafish dysregulates COPII-dependent secretion of select extracellular matrix proteins in skeletal morphogenesis. Dis Model Mech 4, 763-776.

Mendler, M., Eich-Bender, S.G., Vaughan, L., Winterhalter, K.H., Bruckner, P., 1989. Cartilage contains mixed fibrils of collagen types II, IX, and XI. J Cell Biol 108, 191-197.

Merte, J., Jensen, D., Wright, K., Sarsfield, S., Wang, Y., Schekman, R., Ginty, D.D., 2010. Sec24b selectively sorts Vangl2 to regulate planar cell polarity during neural tube closure. Nat Cell Biol 12, 41-46; sup pp 41-48.

Mienaltowski, M.J., Birk, D.E., 2014. Structure, physiology, and biochemistry of collagens. Adv Exp Med Biol 802, 5-29.

Miller, E., Antonny, B., Hamamoto, S., Schekman, R., 2002. Cargo selection into COPII vesicles is driven by the Sec24p subunit. Embo j 21, 6105-6113.

Miller, E.A., Beilharz, T.H., Malkus, P.N., Lee, M.C., Hamamoto, S., Orci, L., Schekman, R., 2003. Multiple cargo binding sites on the COPII subunit Sec24p ensure capture of diverse membrane proteins into transport vesicles. Cell 114, 497-509.

Mironov, A.A., Mironov, A.A., Jr., Beznoussenko, G.V., Trucco, A., Lupetti, P., Smith, J.D., Geerts, W.J., Koster, A.J., Burger, K.N., Martone, M.E., Deerinck, T.J., Ellisman, M.H., Luini, A., 2003. ER-to-Golgi carriers arise through direct en bloc protrusion and multistage maturation of specialized ER exit domains. Dev Cell 5, 583-594.

Miserey-Lenkei, S., Chalancon, G., Bardin, S., Formstecher, E., Goud, B., Echard, A., 2010. Rab and actomyosin-dependent fission of transport vesicles at the Golgi complex. Nat Cell Biol 12, 645-654.

Mitchell, R., Huitema, L., Skinner, R., Brunt, L., Severn, C., Schulte-Merker, S., Hammond, C., 2013. New tools for studying osteoarthritis genetics in zebrafish. Osteoarthritis Cartilage 21, 269-278.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., Valen, E., 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res 42, W401-407.

Montanaro, L., Parisini, P., Greggi, T., Di Silvestre, M., Campoccia, D., Rizzi, S., Arciola, C.R., 2006. Evidence of a linkage between matrilin-1 gene (MATN1) and idiopathic scoliosis. Scoliosis 1, 21.

Montero-Balaguer, M., Lang, M.R., Sachdev, S.W., Knappmeyer, C., Stewart, R.A., De La Guardia, A., Hatzopoulos, A.K., Knapik, E.W., 2006. The mother superior mutation ablates foxd3 activity in neural crest progenitor cells and depletes neural crest derivatives in zebrafish. Dev Dyn 235, 3199-3212.

Montero-Balaguer, M., Swirsding, K., Orsenigo, F., Cotelli, F., Mione, M., Dejana, E., 2009. Stable vascular connections and remodeling require full expression of VE-cadherin in zebrafish embryos. PloS one 4, e5772.

Moosa, S., Chung, B.H., Tung, J.Y., Altmuller, J., Thiele, H., Nurnberg, P., Netzer, C., Nishimura, G., Wollnik, B., 2015. Mutations in SEC24D cause autosomal recessive osteogenesis imperfecta. Clin Genet.

Mumm, S., Christie, P.T., Finnegan, P., Jones, J., Dixon, P.H., Pannett, A.A., Harding, B., Gottesman, G.S., Thakker, R.V., Whyte, M.P., 2000. A five-base pair deletion in the sedlin gene causes spondyloepiphyseal dysplasia tarda in a six-generation Arkansas kindred. J Clin Endocrinol Metab 85, 3343-3347.

Mumm, S., Zhang, X., Vacca, M., D'Esposito, M., Whyte, M.P., 2001. The sedlin gene for spondyloepiphyseal dysplasia tarda escapes X-inactivation and contains a non-canonical splice site. Gene 273, 285-293.

Myllyharju, J., Kivirikko, K.I., 2004. Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet 20, 33-43.

Nakano, A., Muramatsu, M., 1989. A novel GTP-binding protein, Sar1p, is involved in transport from the endoplasmic reticulum to the Golgi apparatus. J Cell Biol 109, 2677-2691.

Nauroy, P., Hughes, S., Naba, A., Ruggiero, F., 2018. The in-silico zebrafish matrisome: A new tool to study extracellular matrix gene and protein functions. Matrix Biol 65, 5-13.

Neuhauss, S.C., Solnica-Krezel, L., Schier, A.F., Zwartkruis, F., Stemple, D.L., Malicki, J., Abdelilah, S., Stainier, D.Y., Driever, W., 1996. Mutations affecting craniofacial development in zebrafish. Development 123, 357-367.

Niu, X., Gao, C., Jan Lo, L., Luo, Y., Meng, C., Hong, J., Hong, W., Peng, J., 2012. Sec13 safeguards the integrity of the endoplasmic reticulum and organogenesis of the digestive system in zebrafish. Dev Biol 367, 197-207.

Norum, M., Tang, E., Chavoshi, T., Schwarz, H., Linke, D., Uv, A., Moussian, B., 2010. Trafficking through COPII Stabilises Cell Polarity and Drives Secretion during Drosophila Epidermal Differentiation. PLoS One 5.

Novick, P., Field, C., Schekman, R., 1980. Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway. Cell 21, 205-215.

Ohisa, S., Inohaya, K., Takano, Y., Kudo, A., 2010. sec24d encoding a component of COPII is essential for vertebra formation, revealed by the analysis of the medaka mutant, vbi. Dev Biol 342, 85-95.

Ohno, S., Murakami, K., Tanimoto, K., Sugiyama, H., Makihira, S., Shibata, T., Yoneno, K., Kato, Y., Tanne, K., 2003. Immunohistochemical study of matrilin-1 in arthritic articular cartilage of the mandibular condyle. J Oral Pathol Med 32, 237-242.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., Mann, M., 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127, 635-648.

Paccaud, J.P., Reith, W., Carpentier, J.L., Ravazzola, M., Amherdt, M., Schekman, R., Orci, L., 1996. Cloning and functional characterization of mammalian homologues of the COPII component Sec23. Mol Biol Cell 7, 1535-1546.

Panagopoulos, I., Moller, E., Dahlen, A., Isaksson, M., Mandahl, N., Vlamis-Gardikas, A., Mertens, F., 2007. Characterization of the native CREB3L2 transcription factor and the FUS/CREB3L2 chimera. Genes Chromosomes Cancer 46, 181-191.

Patel, N., Anand, D., Monies, D., Maddirevula, S., Khan, A.O., Algoufi, T., Alowain, M., Faqeih, E., Alshammari, M., Qudair, A., Alsharif, H., Aljubran, F., Alsaif, H.S., Ibrahim, N., Abdulwahab, F.M., Hashem, M., Alsedairy, H., Aldahmesh, M.A., Lachke, S.A., Alkuraya, F.S., 2017. Novel phenotypes and loci identified through clinical genomics approaches to pediatric cataract. Hum Genet 136, 205-225.

Pendergrass, S.A., Brown-Gentry, K., Dudek, S., Frase, A., Torstenson, E.S., Goodloe, R., Ambite, J.L., Avery, C.L., Buyske, S., Buzkova, P., Deelman, E., Fesinmeyer, M.D., Haiman, C.A., Heiss, G., Hindorff, L.A., Hsu, C.N., Jackson, R.D., Kooperberg, C., Le Marchand, L., Lin, Y., Matise, T.C., Monroe, K.R., Moreland, L., Park, S.L., Reiner, A., Wallace, R., Wilkens, L.R., Crawford, D.C., Ritchie, M.D., 2013. Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. PLoS genetics 9, e1003087.

Pin, J.P., Duvoisin, R., 1995. The metabotropic glutamate receptors: structure and functions. Neuropharmacology 34, 1-26.

Pinzani, M., Rosselli, M., Zuckermann, M., 2011. Liver cirrhosis. Best Pract Res Clin Gastroenterol 25, 281-290.

Piotrowski, T., Schilling, T.F., Brand, M., Jiang, Y.J., Heisenberg, C.P., Beuchle, D., Grandel, H., van Eeden, F.J., Furutani-Seiki, M., Granato, M., Haffter, P., Hammerschmidt, M., Kane, D.A., Kelsh, R.N., Mullins, M.C., Odenthal, J., Warga, R.M., Nusslein-Volhard, C., 1996. Jaw and branchial arch mutants in zebrafish II: anterior arches and cartilage differentiation. Development 123, 345-356.

Polishchuk, E.V., Di Pentima, A., Luini, A., Polishchuk, R.S., 2003. Mechanism of Constitutive Export from the Golgi: Bulk Flow via the Formation, Protrusion, and En Bloc Cleavage of large trans-Golgi Network Tubular Domains$\boxed{V}$. Mol Biol Cell 14, 4470-4485.

Polishchuk, R.S., Capestrano, M., Polishchuk, E.V., 2009. Shaping tubular carriers for intracellular membrane transport. FEBS Lett 583, 3847-3856.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics 38, 904-909.

Provost, E., Rhee, J., Leach, S.D., 2007. Viral 2A peptides allow expression of multiple proteins from a single ORF in transgenic zebrafish embryos. Genesis 45, 625-629.

Pusapati, G.V., Luchetti, G., Pfeffer, S.R., 2012. Ric1-Rgp1 complex is a guanine nucleotide exchange factor for the late Golgi Rab6A GTPase and an effector of the medial Golgi Rab33B GTPase. J Biol Chem 287, 42129-42137.

Reddi, A.H., Gay, R., Gay, S., Miller, E.J., 1977. Transitions in collagen types during matrix-induced cartilage, bone, and bone marrow formation. Proc Natl Acad Sci U S A 74, 5589-5592.

Ritchie, M.D., Denny, J.C., Crawford, D.C., Ramirez, A.H., Weiner, J.B., Pulley, J.M., Basford, M.A., Brown-Gentry, K., Balser, J.R., Masys, D.R., Haines, J.L., Roden, D.M., 2010. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am J Hum Genet 86, 560-572.

Rivals, I., Personnaz, L., Taing, L., Potier, M.C., 2007. Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics 23, 401-407.

Roden, D.M., 2017. Phenome-wide association studies: a new method for functional genomics in humans. J Physiol 595, 4109-4115.

Roden, D.M., Pulley, J.M., Basford, M.A., Bernard, G.R., Clayton, E.W., Balser, J.R., Masys, D.R., 2008. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther 84, 362-369.

Russo, R., Esposito, M.R., Asci, R., Gambale, A., Perrotta, S., Ramenghi, U., Forni, G.L., Uygun, V., Delaunay, J., Iolascon, A., 2010. Mutational spectrum in congenital dyserythropoietic anemia type II: identification of 19 novel variants in SEC23B gene. Am J Hematol 85, 915-920.

S, A., 2010. FastQC: a quality control tool for high throughput sequence data.

Sachdev, S.W., Dietz, U.H., Oshima, Y., Lang, M.R., Knapik, E.W., Hiraki, Y., Shukunami, C., 2001. Sequence analysis of zebrafish chondromodulin-1 and expression profile in the notochord and chondrogenic regions during cartilage morphogenesis. Mech Dev 105, 157-162.

Sadowski, M., Suryadinata, R., Tan, A.R., Roesley, S.N., Sarcevic, B., 2012. Protein monoubiquitination and polyubiquitination generate structural diversity to control distinct biological processes. IUBMB Life 64, 136-142.

Saito, A., Hino, S., Murakami, T., Kanemoto, S., Kondo, S., Saitoh, M., Nishimura, R., Yoneda, T., Furuichi, T., Ikegawa, S., Ikawa, M., Okabe, M., Imaizumi, K., 2009a. Regulation of endoplasmic reticulum stress response by a BBF2H7-mediated Sec23a pathway is essential for chondrogenesis. Nat Cell Biol 11, 1197-1204.

Saito, A., Kanemoto, S., Zhang, Y., Asada, R., Hino, K., Imaizumi, K., 2014. Chondrocyte proliferation regulated by secreted luminal domain of ER stress transducer BBF2H7/CREB3L2. Mol Cell 53, 127-139.

Saito, K., Chen, M., Bard, F., Chen, S., Zhou, H., Woodley, D., Polischuk, R., Schekman, R., Malhotra, V., 2009b. TANGO1 facilitates cargo loading at endoplasmic reticulum exit sites. Cell 136, 891-902.

Saito, K., Katada, T., 2015. Mechanisms for exporting large-sized cargoes from the endoplasmic reticulum. Cell Mol Life Sci.

Saito, K., Yamashiro, K., Ichikawa, Y., Erlmann, P., Kontani, K., Malhotra, V., Katada, T., 2011. cTAGE5 mediates collagen secretion through interaction with TANGO1 at endoplasmic reticulum exit sites. Mol Biol Cell 22, 2301-2308.

Salama, N.R., Chuang, J.S., Schekman, R.W., 1997. Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum. Mol Biol Cell 8, 205-217.

Sarmah, S., Barrallo-Gimeno, A., Melville, D.B., Topczewski, J., Solnica-Krezel, L., Knapik, E.W., 2010. Sec24D-dependent transport of extracellular matrix proteins is required for zebrafish skeletal morphogenesis. PLoS One 5, e10367.

Schalkwijk, J., Zweers, M.C., Steijlen, P.M., Dean, W.B., Taylor, G., van Vlijmen, I.M., van Haren, B., Miller, W.L., Bristow, J., 2001. A recessive form of the Ehlers-Danlos syndrome caused by tenascin-X deficiency. N Engl J Med 345, 1167-1175.

Schibler, L., Gibbs, L., Benoist-Lasselin, C., Decraene, C., Martinovic, J., Loget, P., Delezoide, A.L., Gonzales, M., Munnich, A., Jais, J.P., Legeai-Mallet, L., 2009. New insight on FGFR3-related chondrodysplasias molecular physiopathology revealed by human chondrocyte gene expression profiling. PLoS One 4, e7633.

Schilling, T.F., Piotrowski, T., Grandel, H., Brand, M., Heisenberg, C.P., Jiang, Y.J., Beuchle, D., Hammerschmidt, M., Kane, D.A., Mullins, M.C., van Eeden, F.J., Kelsh, R.N., Furutani-Seiki, M., Granato, M., Haffter, P., Odenthal, J., Warga, R.M., Trowe, T., Nusslein-Volhard, C., 1996. Jaw and branchial arch mutants in zebrafish I: branchial arches. Development 123, 329-344.

Schmidt, K., Cavodeassi, F., Feng, Y., Stephens, D.J., 2013. Early stages of retinal development depend on Sec13 function. Biol Open 2, 256-266.

Schreml, J., Durmaz, B., Cogulu, O., Keupp, K., Beleggia, F., Pohl, E., Milz, E., Coker, M., Ucar, S.K., Nurnberg, G., Nurnberg, P., Kuhn, J., Ozkinay, F., 2014. The missing "link": an autosomal recessive short stature syndrome caused by a hypofunctional XYLT1 mutation. Hum Genet 133, 29-39.

Schwarz, K., Iolascon, A., Verissimo, F., Trede, N.S., Horsley, W., Chen, W., Paw, B.H., Hopfner, K.P., Holzmann, K., Russo, R., Esposito, M.R., Spano, D., De Falco, L., Heinrich, K., Joggerst, B., Rojewski, M.T., Perrotta, S., Denecke, J., Pannicke, U., Delaunay, J., Pepperkok, R., Heimpel, H., 2009. Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II. Nat Genet 41, 936-940.

Senkov, O., Andjus, P., Radenovic, L., Soriano, E., Dityatev, A., 2014. Neural ECM molecules in synaptic plasticity, learning, and memory. Prog Brain Res 214, 53-80.

Sharpe, L.J., Luu, W., Brown, A.J., 2011. Akt phosphorylates Sec24: new clues into the regulation of ER-to-Golgi trafficking. Traffic 12, 19-27.

Shaywitz, D.A., Espenshade, P.J., Gimeno, R.E., Kaiser, C.A., 1997. COPII subunit interactions in the assembly of the vesicle coat. J Biol Chem 272, 25413-25416.

Shoulders, C.C., Stephens, D.J., Jones, B., 2004. The intracellular transport of chylomicrons requires the small GTPase, Sar1b. Curr Opin Lipidol 15, 191-197.

Shoulders, M.D., Raines, R.T., 2009. Collagen structure and stability. Annu Rev Biochem 78, 929-958.

Shugrue, C.A., Kolen, E.R., Peters, H., Czernik, A., Kaiser, C., Matovcik, L., Hubbard, A.L., Gorelick, F., 1999. Identification of the putative mammalian orthologue of Sec31P, a component of the COPII coat. J Cell Sci 112 (Pt 24), 4547-4556.

Siddiqi, S., Siddiqi, S.A., Mansbach, C.M., 2nd, 2010. Sec24C is required for docking the prechylomicron transport vesicle with the Golgi. J Lipid Res 51, 1093-1100.

Siniossoglou, S., Peak-Chew, S.Y., Pelham, H.R., 2000. Ric1p and Rgp1p form a complex that catalyses nucleotide exchange on Ypt6p. Embo j 19, 4885-4894.

Sisson, B.E., Dale, R.M., Mui, S.R., Topczewska, J.M., Topczewski, J., 2015. A role of glypican4 and wnt5b in chondrocyte stacking underlying craniofacial cartilage morphogenesis. Mech Dev 138 Pt 3, 279-290.

Soleman, S., Filippov, M.A., Dityatev, A., Fawcett, J.W., 2013. Targeting the neural extracellular matrix in neurological disorders. Neuroscience 253, 194-213.

Soul, J., Dunn, S.L., Anand, S., Serracino-Inglott, F., Schwartz, J.M., Boot-Handford, R.P., Hardingham, T.E., 2018. Stratification of knee osteoarthritis: two major patient subgroups identified by genome-wide expression analysis of articular cartilage. Ann Rheum Dis 77, 423.

St-Jacques, B., Hammerschmidt, M., McMahon, A.P., 1999. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. Genes Dev 13, 2072-2086.

Stagg, S.M., Gurkan, C., Fowler, D.M., LaPointe, P., Foss, T.R., Potter, C.S., Carragher, B., Balch, W.E., 2006. Structure of the Sec13/31 COPII coat cage. Nature 439, 234-238.

Stankewich, M.C., Stabach, P.R., Morrow, J.S., 2006. Human Sec31B: a family of new mammalian orthologues of yeast Sec31p that associate with the COPII coat. J Cell Sci 119, 958-969.

Stegle, O., Parts, L., Piipari, M., Winn, J., Durbin, R., 2012. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nature protocols 7, 500-507.

Stephens, D.J., 2012. CELL BIOLOGY: Collagen secretion explained. Nature 482, 474-475.

Stephens, D.J., Pepperkok, R., 2002. Imaging of procollagen transport reveals COPI-dependent cargo sorting during ER-to-Golgi transport in mammalian cells. J Cell Sci 115, 1149-1160.

Sucic, S., El-Kasaby, A., Kudlacek, O., Sarker, S., Sitte, H.H., Marin, P., Freissmuth, M., 2011. The serotonin transporter is an exclusive client of the coat protein complex II (COPII) component SEC24C. J Biol Chem 286, 16482-16490.

Swaroop, A., Yang-Feng, T.L., Liu, W., Gieser, L., Barrow, L.L., Chen, K.C., Agarwal, N., Meisler, M.H., Smith, D.I., 1994. Molecular characterization of a novel human gene, SEC13R, related to the yeast secretory pathway gene SEC13, and mapping to a conserved linkage group on human chromosome 3p24-p25 and mouse chromosome 6. Hum Mol Genet 3, 1281-1286.

Symoens, S., Malfait, F., D’hondt, S., Callewaert, B., Dheedene, A., Steyaert, W., Bächinger, H.P., De Paepe, A., Kayserili, H., Coucke, P.J., 2013. Deficiency for the ER-stress transducer OASIS causes severe recessive osteogenesis imperfecta in humans. Orphanet J Rare Dis 8, 154.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., von Mering, C., 2015. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43, D447-452.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., Jensen, L.J., von Mering, C., 2017. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res 45, D362-368.

Szul, T., Sztul, E., 2011. COPII and COPI traffic at the ER-Golgi interface. Physiology (Bethesda) 26, 348-364.

Tanegashima, K., Zhao, H., Rebbert, M.L., Dawid, I.B., 2009. Coordinated activation of the secretory pathway during notochord formation in the Xenopus embryo. Development 136, 3543-3548.

Tang, B.L., Kausalya, J., Low, D.Y., Lock, M.L., Hong, W., 1999. A family of mammalian proteins homologous to yeast Sec24p. Biochem Biophys Res Commun 258, 679-684.

Tang, B.L., Zhang, T., Low, D.Y., Wong, E.T., Horstmann, H., Hong, W., 2000. Mammalian homologues of yeast sec31p. An ubiquitously expressed form is localized to endoplasmic reticulum (ER) exit sites and is essential for ER-Golgi transport. J Biol Chem 275, 13597-13604.

Tao, J., Zhu, M., Wang, H., Afelik, S., Vasievich, M.P., Chen, X.W., Zhu, G., Jensen, J., Ginsburg, D., Zhang, B., 2012. SEC23B is required for the maintenance of murine professional secretory tissues. Proc Natl Acad Sci U S A 109, E2001-2009.

Tomoishi, S., Fukushima, S., Shinohara, K., Katada, T., Saito, K., 2017. CREB3L2-mediated expression of Sec23A/Sec24D is involved in hepatic stellate cell activation through ER-Golgi transport. Sci Rep 7, 7992.

Townley, A.K., Feng, Y., Schmidt, K., Carter, D.A., Porter, R., Verkade, P., Stephens, D.J., 2008. Efficient coupling of Sec23-Sec24 to Sec13-Sec31 drives COPII-dependent collagen secretion and is essential for normal craniofacial development. J Cell Sci 121, 3025-3034.

Townley, A.K., Schmidt, K., Hodgson, L., Stephens, D.J., 2012. Epithelial organization and cyst lumen expansion require efficient Sec13-Sec31-driven secretion. J Cell Sci 125, 673-684.

Unger, S., Lausch, E., Rossi, A., Megarbane, A., Sillence, D., Alcausin, M., Aytes, A., Mendoza-Londono, R., Nampoothiri, S., Afroze, B., Hall, B., Lo, I.F., Lam, S.T., Hoefele, J., Rost, I., Wakeling, E., Mangold, E., Godbole, K., Vatanavicharn, N., Franco, L.M., Chandler, K., Hollander, S., Velten, T., Reicherter, K., Spranger, J., Robertson, S., Bonafe, L., Zabel, B., Superti-Furga, A., 2010. Phenotypic features of carbohydrate sulfotransferase 3 (CHST3) deficiency in 24 patients: congenital dislocations and vertebral changes as principal diagnostic features. Am J Med Genet A 152a, 2543-2549.

Unlu, G., Levic, D.S., Melville, D.B., Knapik, E.W., 2014. Trafficking mechanisms of extracellular matrix macromolecules: insights from vertebrate development and human diseases. Int J Biochem Cell Biol 47, 57-67.

Vacaru, A.M., Unlu, G., Spitzner, M., Mione, M., Knapik, E.W., Sadler, K.C., 2014. In vivo cell biology in zebrafish - providing insights into vertebrate development and disease. J Cell Sci 127, 485-495.

Valente, C., Polishchuk, R., De Matteis, M.A., 2010. Rab6 and myosin II at the cutting edge of membrane fission. Nat Cell Biol 12, 635-638.

Varshney, G.K., Pei, W., LaFave, M.C., Idol, J., Xu, L., Gallardo, V., Carrington, B., Bishop, K., Jones, M., Li, M., Harper, U., Huang, S.C., Prakash, A., Chen, W., Sood, R., Ledin, J., Burgess, S.M., 2015. High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. Genome Res.

Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., Schier, A.F., 2010. Chromatin signature of embryonic pluripotency is established during genome activation. Nature 464, 922-926.

Venditti, R., Scanu, T., Santoro, M., Di Tullio, G., Spaar, A., Gaibisso, R., Beznoussenko, G.V., Mironov, A.A., Mironov, A., Jr., Zelante, L., Piemontese, M.R., Notarangelo, A., Malhotra, V., Vertel, B.M., Wilson, C., De Matteis, M.A., 2012. Sedlin controls the ER export of procollagen by regulating the Sar1 cycle. Science 337, 1668-1672.

Vornehm, S.I., Dudhia, J., Von der Mark, K., Aigner, T., 1996. Expression of collagen types IX and XI and other major cartilage matrix components by human fetal chondrocytes in vivo. Matrix Biol 15, 91-98.

Wadhwa, R., Kaul, S.C., Komatsu, Y., Ikawa, Y., Sarai, A., Sugimoto, Y., 1993. Identification and differential expression of yeast SEC23-related gene (Msec23) in mouse tissues. FEBS Lett 315, 193-196.

Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., Björkegren, J.L.M., Im, H.K., Pasaniuc, B., Rivas, M.A., Kundaje, A., 2018. Transcriptome-wide association studies: opportunities and challenges. biorxiv.

Wakana, Y., van Galen, J., Meissner, F., Scarpa, M., Polishchuk, R.S., Mann, M., Malhotra, V., 2012. A new class of carriers that transport selective cargo from the trans Golgi network to the cell surface. Embo j 31, 3976-3990.

Walker, M.B., Kimmel, C.B., 2007. A two-color acid-free cartilage and bone stain for zebrafish larvae. Biotech Histochem 82, 23-28.

Wang, J., Duncan, D., Shi, Z., Zhang, B., 2013. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res 41, W77-83.

Wansleeben, C., Feitsma, H., Montcouquiol, M., Kroon, C., Cuppen, E., Meijlink, F., 2010. Planar cell polarity defects and defective Vangl2 trafficking in mutants for the COPII gene Sec24b. Development 137, 1067-1073.

Watson, P., Townley, A.K., Koka, P., Palmer, K.J., Stephens, D.J., 2006. Sec16 defines endoplasmic reticulum exit sites and is required for secretory cargo export in mammalian cells. Traffic 7, 1678-1687.

Wei, W.Q., Denny, J.C., 2015. Extracting research-quality phenotypes from electronic health records to support precision medicine. Genome Med 7.

Wendeler, M.W., Paccaud, J.P., Hauri, H.P., 2007. Role of Sec24 isoforms in selective export of membrane proteins from the endoplasmic reticulum. EMBO Rep 8, 258-264.

White, J., Johannes, L., Mallard, F., Girod, A., Grill, S., Reinsch, S., Keller, P., Tzschaschel, B., Echard, A., Goud, B., Stelzer, E.H., 1999. Rab6 Coordinates a Novel Golgi to ER Retrograde Transport Pathway in Live Cells. J Cell Biol 147, 743-760.

White, R.J., Collins, J.E., Sealy, I.M., Wali, N., Dooley, C.M., Digby, Z., Stemple, D.L., Murphy, D.N., Billis, K., Hourlier, T., Fullgrabe, A., Davis, M.P., Enright, A.J., Busch-Nentwich, E.M., 2017. A high-resolution mRNA expression time course of embryonic development in zebrafish. Elife 6.

Williams, B.B., Cantrell, V.A., Mundell, N.A., Bennett, A.C., Quick, R.E., Jessen, J.R., 2012a. VANGL2 regulates membrane trafficking of MMP14 to control cell polarity and migration. J Cell Sci 125, 2141-2147.

Williams, B.B., Mundell, N., Dunlap, J., Jessen, J., 2012b. The planar cell polarity protein VANGL2 coordinates remodeling of the extracellular matrix, Commun Integr Biol, pp. 325-328.

Willing, M.C., Deschenes, S.P., Scott, D.A., Byers, P.H., Slayton, R.L., Pitts, S.H., Arikat, H., Roberts, E.J., 1994. Osteogenesis imperfecta type I: molecular heterogeneity for COL1A1 null alleles of type I collagen. Am J Hum Genet 55, 638-647.

Wilson, D.G., Phamluong, K., Li, L., Sun, M., Cao, T.C., Liu, P.S., Modrusan, Z., Sandoval, W.N., Rangell, L., Carano, R.A., Peterson, A.S., Solloway, M.J., 2011. Global defects in collagen secretion in a Mia3/TANGO1 knockout mouse. J Cell Biol 193, 935-951.

Wolford, B.N., Willer, C.J., Surakka, I., 2018. Electronic health records: the next wave of complex disease genetics. Hum Mol Genet 27, R14-r21.

Wu, L., Bluguermann, C., Kyupelyan, L., Latour, B., Gonzalez, S., Shah, S., Galic, Z., Ge, S., Zhu, Y., Petrigliano, F.A., Nsair, A., Miriuka, S.G., Li, X., Lyons, K.M., Crooks, G.M., McAllister, D.R., Van Handel, B., Adams, J.S., Evseenko, D., 2013. Human developmental chondrogenesis as a basis for engineering chondrocytes from pluripotent stem cells. Stem Cell Reports 1, 575-589.

Yan, Y.L., Miller, C.T., Nissen, R.M., Singer, A., Liu, D., Kirn, A., Draper, B., Willoughby, J., Morcos, P.A., Amsterdam, A., Chung, B.C., Westerfield, M., Haffter, P., Hopkins, N., Kimmel, C., Postlethwait, J.H., 2002. A zebrafish sox9 gene required for cartilage morphogenesis. Development 129, 5065-5079.

Yanagi, M., Kawasaki, R., Wang, J.J., Wong, T.Y., Crowston, J., Kiuchi, Y., 2011. Vascular risk factors in glaucoma: a review. Clin Exp Ophthalmol 39, 252-258.

Yang, X.Y., Zhou, X.Y., Wang, Q.Q., Li, H., Chen, Y., Lei, Y.P., Ma, X.H., Kong, P., Shi, Y., Jin, L., Zhang, T., Wang, H.Y., 2013. Mutations in the COPII Vesicle

Component Gene SEC24B are Associated with Human Neural Tube Defects. Hum Mutat.

Ye, J., Rawson, R.B., Komuro, R., Chen, X., Dave, U.P., Prywes, R., Brown, M.S., Goldstein, J.L., 2000. ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. Mol Cell 6, 1355-1364.

Yin, L., Jao, L.E., Chen, W., 2015. Generation of Targeted Mutations in Zebrafish Using the CRISPR/Cas System. Methods Mol Biol 1332, 205-217.

Yoshihisa, T., Barlowe, C., Schekman, R., 1993. Requirement for a GTPase-activating protein in vesicle budding from the endoplasmic reticulum. Science 259, 1466-1468.

Young, S.Z., Taylor, M.M., Bordey, A., 2011. Neurotransmitters couple brain activity to subventricular zone neurogenesis. Eur J Neurosci 33, 1123-1132.

Yu, X., Lyu, D., Dong, X., He, J., Yao, K., 2014. Hypertension and risk of cataract: a meta-analysis. PLoS One 9, e114012.

Zhang, B., Kirov, S., Snoddy, J., 2005. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 33, W741-748.