# Phenome-based approach identifies RIC1-linked Mendelian syndrome through zebrafish models, biobank associations, and clinical studies

**Gokhan Unlu**[1,2,3,8], **Xinzi Qi**[1,2], **Eric R. Gamazon**[1,2,4,10], **David B. Melville**[1,9,10], **Nisha Patel**[5,10], **Amy R. Rushing**[1,2], **Mais Hashem**[5], **Abdullah Al-Faifi**[6], **Rui Chen**[2,7], **Bingshan Li**[2,7], **Nancy J. Cox**[1,2], **Fowzan S. Alkuraya**[5], **Ela W. Knapik**[1,2,3,*]

[1]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA

[2]Vanderbilt Genetic Institute, Vanderbilt University Medical Center, Nashville, Tennessee, USA

[3]Department of Cell and Developmental Biology, Vanderbilt University, Nashville, Tennessee, USA

[4]Clare Hall, University of Cambridge, Cambridge CB3 9AL, United Kingdom

[5]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

[6]Department of Pediatrics, Security Forces Hospital, Riyadh, Saudi Arabia

[7]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, USA

## Abstract

Discovery of genotype-phenotype relationships remains a major challenge in clinical medicine. Here, we combined three sources of phenotypic data to uncover a novel mechanism for rare and common diseases resulting from collagen secretion deficits. Using zebrafish genetic screen, we identified the *ric1* gene to be essential for skeletal biology. Using a gene-based phenome-wide association study (PheWAS) in the EHR-linked BioVU biobank, we show that reduced genetically determined expression of *RIC1* is associated with musculoskeletal and dental conditions. Whole

exome sequencing (WES) identified individuals homozygous-by-descent for a rare variant in *RIC1*, and, through a guided clinical re-evaluation, they were discovered to share signs with the BioVU-associated phenome. We named this novel Mendelian syndrome CATIFA (Cleft lip, cAtaract, Tooth abnormality, Intellectual disability, Facial dysmorphism, ADHD), and revealed further disease mechanisms. This gene-based PheWAS-guided approach can accelerate the discovery of clinically relevant disease phenome and associated biological mechanisms.

## 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[1]. 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 accelerate discovery of common and rare disease mechanisms and unknown biology.

Collagens are essential components of extracellular matrices and their disrupted secretion has been linked in patients and animal models to general skeletal and craniofacial dysmorphology[2-5]. Large extracellular matrix (ECM) cargos, such as procollagen, are synthesized in the endoplasmic reticulum (ER), post-translationally modified in the Golgi and trafficked out to the ECM[6].

Zebrafish have been instrumental in identifying genes acting at each of these steps[7-9]. Variants in proteins at ER-to-Golgi secretory pathway, including the coat proteins II (COPII) SEC23A and SEC24D[10-13] and procollagen modifying enzymes[14,15] cause skeletal and craniofacial defects in zebrafish. *SEC23A* mutations lead to cranio-lenticulo-sutural-dysplasia (CLSD), a disease characterized by craniofacial and skeletal defects[16]. These studies established zebrafish as a powerful tool to study procollagen transport and model skeletal conditions. Though ER-to-Golgi transport of procollagen is relatively well-studied, how procollagen is transported from the Golgi to plasma membrane [17,18] and the medical phenome[19] resulting from dysfunction of this process remains a long-standing knowledge gap.

We show here that Ric1 and its binding partner, Rgp1, are required to activate Rab6a for procollagen transport through the Trans Golgi Network (TGN) and skeletal development in zebrafish models. We investigated human phenotypes associated with the genetically reduced expression of *RIC1* in phenome-linked DNA biobanks. Clinical re-evaluation of subjects previously found to be homozygous-by-descent for a *RIC1* variant and zebrafish knockouts validated common-disease phenome in these subjects, including abnormal tooth development and attention deficit hyperactivity disorder (ADHD). This discovery allowed us to describe a novel genetic syndrome, now termed CATIFA, and establish the mechanistic continuum between individual signs of a Mendelian disease and complex traits.

## Results:

### RIC1 is required for normal skeletogenesis

In search for novel components of the procollagen secretory pathway with essential roles in skeletal biology, we characterized the zebrafish craniofacial *round* (*rnd* [m641, m713, m715]) mutations, which were identified in a large-scale chemical mutagenesis screen[20]. We genetically mapped and positionally cloned the *round* locus to identify mutations in the *ric1* gene (KIAA1432, ENSDARG00000063362 in Zv9), (Fig. 1a; Extended Data Fig. 1a-c). By direct sequencing of cDNAs 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. 1a). Ric1 is a highly evolutionarily conserved protein sharing 71% identity from zebrafish to human. The yeast and human homologs of Ric1 protein and its binding partner Rgp1 were shown to act as a guanine nucleotide exchange factor (GEF) for Rab6 GTPase[21,22]. However, the role of Ric1-Rgp1-Rab6a in the context of vertebrate development and physiology has not been established.

The zebrafish *ric1*[−/−] mutations present with flattened head, lack of jaw protrusion, kinked pectoral fins and short head/body length index at 5 days post-fertilization (dpf) (Fig. 1b). The microcephaly-like phenotype was corroborated by smaller head size indices compared to age-matched wild-type (WT) siblings (Fig. 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 malformed and smaller than WT controls (Fig. 1d, Extended Data Fig. 1d,e). Human RIC1 (hRIC1) overexpression by mRNA was sufficient, in genetic replacement experiments, to restore jaw protrusion (Fig. 1e), head and body length in *ric1*[−/−] larvae (Fig. 1f,g), corroborating functional conservation of zebrafish and human *RIC1* genes in skeletal development.

To characterize the craniofacial architecture, we examined fluorescently marked transgenic animals by live imaging (Fig. 1h). We analyzed the organ shape of the jaw-supporting hyosymplectic cartilage (HC), an easily identifiable key-like element accessible for *in vivo* imaging. We imaged the entire HC using confocal microscopy (Fig. 1h), and created 3D reconstructions (Fig. 1i), which showed irregularities including thicker and misshapen arm and body of the HC, also a collapsed foramen, an anatomical passage point (Supplementary video 1,2). High magnification imaging of the single-cell layered HC arm (Fig. 1j,k) revealed smaller, constricted mutant chondrocytes, especially around the midline, and loss of the characteristic stack-of-coins shape (Supplementary video 3,4). Volume measurements of five HC chondrocytes at the corresponding positions of age-matched zebrafish showed reduced volume ~185 μm$^3$ in mutants, compared to average WT volume of ~350 μm$^3$ (Fig. 1l). We conclude that Ric1 function is required for normal cell shape that further conveys organ shape and, ultimately, the overall craniofacial morphology.

## BioVU analyses revealed craniodental and skeletal phenotypes associated with genetically predicted reduced *RIC1* expression

To compare the zebrafish *ric1*[−/−] phenotype to the corresponding human common-disease phenome, we examined the association of genetically predicted reduced expression of *RIC1* in the BioVU Electronic Health Records (EHR) and DNA biobank[23-25]. For this analysis, we applied *PrediXcan*[26] which estimates genetically regulated gene expression (GReX) from the publicly available GTEx[27,28] project data as a reference panel. We tested the association of GReX with phecodes[25] in ~10,000 subjects of European ancestries in BioVU to determine gene-level associations to disease traits (Fig. 2a). The *RIC1* prediction model includes single-nucleotide polymorphisms (SNPs) in the region (within 1 Mbp of start and stop sites) of the gene (Supplementary Table 1). In a traditional PheWAS, one SNP is used to test association; here, we used a variation of the method, which we called "gene-based PheWAS", because multiple SNPs are used around the gene-of-interest. We focus here on all traits (Phecodes) nominally significantly associated with reduced GReX of *RIC1*, including skeletal, respiratory (asthma), sensory organs (eye, e.g., corneal opacity), nervous (ADHD), digestive (diverticulosis), and cardiovascular (stricture of artery) (Fig. 2b,c and Supplementary Table 2).

We subsequently sought evidence for replication of the BioVU-associated phenome through parallel analyses on GReX of *RIC1* and on SNPs at the gene in the UK Biobank[29] (UKB), and other publicly available GWAS data (Figure 2d, Supplementary Table 3 and 4). Asthma is significantly associated with reduced GReX of *RIC1* in UKB after correction for the number tissues and phenotypes, and in data from other independent GWAS (Fig. 2d, Extended Data Fig. 2a-c, Supplementary Table 3 and 4). Association studies in phenome available in UKB reveal SNPs at *RIC1* are highly significantly associated with at least one phenotype in each of the following phenotypic categories: skeleton/joint, respiratory, eye, nervous, digestive, cardiovascular and blood, neoplasm and infection (Fig. 2d, Supplementary Table 4).

The most notable skeletal BioVU phecodes included 'fracture of unspecified bones', 'acquired deformities of limbs', and 'disorders of tooth development', specifically, 'tooth eruption and development' (Fig. 2c). The PheWAS findings drew our attention to zebrafish dentation, easily visible in older larvae at 7 dpf. Alizarin red staining (calcified tissue) showed the presence of pharyngeal teeth and dermal bones such as cleithrum (pectoral girdle) in wild types, whereas *ric1*[−/−] larvae lacked pharyngeal teeth, and had smaller calcification domains that failed to elongate (Extended Data Fig. 1d,e). These findings confirmed a highly evolutionarily conserved function of *RIC1* in skeletal biology and tooth development. They also revealed parallels between zebrafish and the phenotypic spectrum in human subjects with reduced GReX of *RIC1*.

## CATIFA syndrome is characterized by cleft lip, cataract, tooth abnormality, intellectual disability, facial dysmorphism, and ADHD

Prior whole exome sequencing (WES) of a pediatric cataract cohort identified a missense mutation in the *RIC1* gene[30]; patients with a variant in *RIC1* were initially characterized as having craniofacial anomalies and developmental delay, in addition to pediatric cataracts.

With limited evaluation of these patients for other organ systems and mounting evidence from the PheWAS and the animal model data supporting a more complex phenotype of RIC1 deficiency, we pursued clinical re-evaluation of the patients guided by the medical phenome associated with the reduced GReX of *RIC1* in BioVU.

We identified 8 persons in two extended, consanguineous families (Fig. 2e, Extended Data Fig. 2d and Supplementary Table 5) homozygous for the missense mutation in *RIC1* and having a range of shared signs and 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. 2f,g). 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. 2f,g and Supplementary Table 5). Other symptoms included sleep disturbance (5/8), bronchial asthma (5/8), small ears (5/8), and clumsy, hypotonic gait (8/8). We collectively refer to these features as CATIFA syndrome (Cleft lip, cAtaract, Tooth abnormality, Intellectual disability, Facial dysmorphism, ADHD). The clinical data of CATIFA patients closely matched the human phenome predicted by PheWAS in BioVU, corroborating the range of medical signs and symptoms associated with *RIC1*. It was indeed the PheWAS prediction that helped identify the core CATIFA syndrome characteristics.

Using genetic linkage mapping, we had initially linked the CATIFA phenome to a mutation on Chromosome 9 with a supporting LOD score of 3.44 (Fig. 2h). Genotyping confirmed a single nucleotide substitution (c.3794C>G), predicted to cause a missense mutation R1265P in the RIC1 protein (Fig. 2i,j and Extended Data Fig. 2e) that fully segregated with the phenotype in each of the two families in a strictly autosomal recessive fashion. We sequenced cDNA obtained from patient fibroblasts and discovered two types of transcripts. The first is a missense mutation R1265P followed by correct splicing of intron 24. The second transcript carries the R1265P mutation and fails to splice due to intron retention, leading to a stop codon at position 1266, which likely undergoes nonsense-mediated decay (NMD).

## Ric1 is an essential component of the procollagen secretory pathway

We sought to identify biological function of Ric1 in vertebrates and the mechanisms underlying CATIFA syndrome. Since Ric1 was implicated in cargo traffic[21], and mutations described here affect 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. 3a,b). Quantification showed that intracellular collagen normally occupies <5% of the cytosolic area, whereas in *ric1*$^{-/-}$ chondrocytes it takes up ~10 times more surface area (Fig. 3c). We found similar results in fibroblast-like, notochord sheath cells (part of the axial hydrostatic skeleton of vertebrate embryos), accompanied by malformation

of notochord basement membrane (BM) (Fig. 3d, Extended Data Fig. 3a,b). Procollagen retention in sheath cells and craniofacial chondrocytes suggests that secretion defect is not cell-type specific, but may be cargo-selective. Thus, we analyzed other cargos and found that they are normally trafficked (Extended Data Fig. 3c-e). We conclude that Ric1 is selectively required for procollagen trafficking.

To test whether the effect on collagen secretion is cell-autonomous, we injected low concentrations of guide RNA (gRNA) to generate *ric1*-mutant cell clones (*ric1^gRNA*) in WT background using CRISPR/Cas9 system (Extended Data Fig. 4a,b). Ric1-deficient clones recapitulated the *round* phenotype and accumulated intracellular collagen, whereas neighboring WT cells secreted collagen normally. These data suggest that Ric1 is cell-autonomously required for collagen transport.

RNA-seq with FACS-sorted zebrafish chondrocytes show that *ric1* expression at 3 dpf is higher than 4 dpf which could be explained by increased collagen transport at the onset of chondrocyte differentiation, while expression of other collagen transport-related genes remains at similar levels (Fig. 3e) . Quantitative PCR (qPCR) and RNA-seq showed low, but steady *ric1* expression throughout developmental stages[31] (Extended Data Fig. 5a,b,c). Collectively, these results suggest that *ric1*'s potent chondrocyte-specific functions drive collagen secretion even at basal expression levels (i.e., RPKM range: 0.67–1.35). *Round* mutations have reduced *ric1* expression, potentially due to nonsense-mediated decay (Extended Data Fig. 5d). Similarly, fibroblasts from the CATIFA patient have reduced levels of *RIC1* transcript (<50% of control cells) that might result from degradation of non-spliced transcripts (Extended Data Fig. 5e).

We examined subcellular organization of the collagen inclusions in chondrocytes by transmission electron microscopy (TEM) and found *ric1^−/−* cells contained vesicles of various sizes, while WT cells were devoid of such vesicular structures (Fig. 3f; Extended Data Fig. 6a-g). High magnification images revealed striated material in the core of the large vesicles, that are reminiscent of collagen fibrils normally formed in ECM, as procollagen is processed by enzymes that cleave N- and C-termini (Fig. 3f-f'). Cartilage matrix analysis by TEM showed reduced ECM cross-linking in *ric1^−/−* at 3 dpf, which appeared absent by by 4 dpf (Fig. 3g). Thus, TEM data corroborate intracellular retention of collagen in *ric1^−/−* cells and compromised ECM ultrastructure.

## Ric1-depletion leads to musculoskeletal and behavioral defects

Common and rare disease phenome associated with *RIC1* included gait disturbances and musculoskeletal deficits. Tendon and muscle attachment sites are matrix-rich structures, containing high levels of thrombospondin-4 (TSP4)[32] and collagen-I (Col1)[33]. We examined, by whole-mount immunofluorescence (WMIF) and TEM analyses, the quality of the muscles and tendons. Confocal images showed shortened and misaligned *ric1^−/−* cranial muscles (Fig. 4a, Extended Data Fig. 7a), and diffused pattern of TSP4 staining in mutants (Fig. 4a, **insets**). In WT tenocytes of the tendon posterior to the sternohyoideus (sh, marked by DAPI), collagen-I (Col1) shows discrete punctate pattern while *ric1^−/−* tenocytes accumulate large clumps in perinuclear regions (Fig. 4b), indicating secretion defects.

Tendons are critical for muscle fibers to attach, and thereby contract[34]. It was previously shown that muscle fiber attachment defects lead to fiber retraction[35]. We analyzed muscle fiber attachment quality by WMIF and observed wavy, retracted muscle fibers in intermandibularis posterior and in trunk muscles (Extended Data Fig. 7a-c). TEM images corroborated discontinuity of myofibrils, revealing lack of periodic myofilament patterning and alignment (Extended Data Fig. 7b).

Zebrafish larvae rely on muscles to explore the surrounding by locomotor behavior patterns consisting of bouts of swimming, turning behavior, and motionless resting episodes[36]. Swimming episodes are divided into: burst swim with fast sharp-angle turns and slow swim with slower routine, wide-angle motions[37]. Musculoskeletal tissues are critical to execute locomotor behaviors[38]. To functionally assess motility defects, we recorded live locomotion of 16 WT and 16 $ric1^{-/-}$ zebrafish larvae. We traced tracks of each embryo's travel for 14 seconds (Fig. 4c, Extended Data Fig. 7d-f). 13/16 WT and 5/16 $ric1^{-/-}$ larvae moved from their initial position during the recording. Mean velocity of WT larvae was 4.33 mm/s while that of $ric1^{-/-}$ mutants was 0.87 mm/s (Fig. 4d). Thus, $ric1$-mutants exhibit fewer exploratory behaviors and are significantly slower (Fig. 4e), than WT counterparts. Collectively, these results show that $ric1$-deficiency leads to structural and functional musculoskeletal defects.

To gain insight into neurological CATIFA symptoms, we analyzed zebrafish brain structures. In zebrafish, the forebrain controls social orienting behavior[39], which is reflective of social engagement and attention, typically affected in ADHD. We analyzed forebrain and cerebellum by WMIF using HuC/D pan-neuronal marker. (Fig. 4f). The data show reduced forebrain size in $ric1^{-/-}$ mutants. We also examined the cerebellum, a brain region functioning in sensorimotor control. The cerebellum controls sensorimotor functions, is implicated in emotional and cognitive tasks, and was linked to intellectual disability, ADHD, and cognitive dysfunction[40]. Microscopy analysis revealed smaller cerebellar area in $ric1$-mutants and a remarkable deficit in neurons at the posterior margin of the cerebellum (Fig. 4f), which regulate motor coordination in response to vestibular information[41]. This phenotype is consistent with the clumsy walk of CATIFA subjects. To assess the overall patterning of cerebellar ECM, we examined WGA (wheat germ agglutinin marking N-glycosylated proteins) staining in $ric1^{-/-}$ larvae. We found noticeable thinning of WGA labeling in rostral cerebellum and optic tectum (Extended Data Fig. 8a, b), suggesting defects in brain matrix formation. Overall, neuronal phenotypes included significantly smaller forebrain and cerebellum, and altered axonal projection (Extended Data Fig. 8c-e), suggesting that loss of $ric1$ leads to morphological and functional deficits in the nervous system.

Of note, we have not observed detectable defects in $ric1^{-/-}$ zebrafish lens at stages analyzed (Extended Data Fig. 8f-g), although 7/8 CATIFA patients had cataract.

## Ric1-Rgp1 GEF complex regulates procollagen transport by activating Rab6a

Prior in vitro studies showed that Ric1-Rgp1 proteins interact and activate Rab6a by facilitating GTP loading (Fig. 5a). This model predicts that depletion of Rgp1 would result in a deficit similar to the $ric1$ loss-of-function phenotype, and that overexpression (OE) of

constitutively active Rab6a would bypass the Ric1-Rgp1 requirement and rescue procollagen traffic in $ric1^{-/-}$ mutant embryos.

To test the first prediction, by injecting a low concentration of $rgp1$-targeting gRNA, we generated CRISPR-induced mosaic mutants containing clones of $rgp1$-mutant cells in WT background (referred to as $rgp1^{gRNA}$ hereafter) (Extended Data Fig. 9a-c). Mosaic $rgp1^{gRNA}$ embryos resembled $ric1^{-/-}$ mutants in their skeletal phenotypes (Extended Data Fig. 9d-g). Col2 antibody and WGA (marking glucosaminoglycans in cartilage ECM) staining showed that $rgp1^{-/-}$ clones contained collagen accumulation similar to $ric1^{-/-}$ chondrocytes (Fig. 5b). 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)[42] or WT Rab6a in both WT and $ric1^{-/-}$ embryos (Fig. 5c). Specifically, we tagged the N-terminus of zebrafish Rab6a with EGFP[43] and mosaically overexpressed it under $col2a1a$ promoter using Tol2kit[44] method (Fig. 5d). IF 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 but mere OE of Rab6a was not sufficient to rescue collagen transport (Fig. 5e), suggesting Rab6a activation is required for collagen secretion.

The GEF activity of RIC1 on Rab6 was shown in vitro[21] in mammalian cells; however, procollagen traffic through post-Golgi Rab6-positive compartments has not been previously implicated in collagen secretion $in\ vivo$. The essentiality of this step in collagen secretion is revealed for the first time by the zebrafish mutations and CATIFA syndrome.

## Pathogenic *RIC1* variant leads to collagen accumulation in patient fibroblasts

To test pathogenicity of the *RIC1* R1265P variant and its role in collagen transport, we analyzed patient-derived fibroblasts. By TEM, we found distended vesicular compartments, similar to those observed in $ric1^{-/-}$ zebrafish mutants, which were notably absent from control skin fibroblasts (Fig. 6a). Consistently, we found significantly expanded Col1 staining throughout the patient fibroblasts as compared to control cells by IF. (Fig. 6b,c, Extended Data Fig. 10a,b).

Rab6a is known to act at various steps of secretory and endocytic pathways[45-47]. To identify the intracellular compartment where RAB6A GEF (RIC1-RGP1) is required for procollagen transport, we tested co-localization of Col1 with compartment-specific markers. TGN markers p230 and Golgin-97 partially co-localized with Col1 staining (Fig. 6b,d, Extended Data Fig. 10a,b). Thus, procollagen transport is blocked at the TGN, placing the RIC1-RGP1 GEF activity in the anterograde secretory pathway for procollagen secretion.

We tested whether *RIC1* function is evolutionarily conserved in regulating procollagen transport. In a genetic replacement experiment, we mosaically overexpressed human *RIC1* (hRIC1) fused to EGFP by a self-hydrolyzable viral v2A tag[48] in zebrafish. Cleavage of the v2A connector detaches hRIC1 from EGFP, thereby labeling transgenic cells with diffused

intracellular EGFP (Fig. 6e). We found that OE of the WT hRIC1 rescued intracellular collagen accumulation in *rnd*$^{-/-}$ cells in a cell-autonomous manner, while having no effect on collagen expression in WT cells (Fig. 6f, Extended Data Fig. 10c,d). We introduced the R1265P variant into hRIC1-v2A-EGFP and assayed for collagen deposits in zebrafish *ric1*$^{-/-}$ cells. Mosaic OE of the R1265P variant failed to fully restore collagen secretion, unlike the WT hRIC1 construct (Fig. 6g). The WT hRIC1 was sufficient to rescue jaw protrusion defects in *rind*$^{-/-}$ (Fig. 1e). To test whether the CATIFA variant can restore jaw morphology *in vivo*, we overexpressed either WT or R1265P (Fig. 6h) in embryos produced by *ric1*$^{+/-}$ heterozygous pairs. Over 90% of the larvae injected with WT construct show fully or partially protruding jaw morphology, compared to 75% of a control non-injected clutch (Mendelian ratio for recessive mutations), while the R1265P construct did not rescue (Fig. 6i). Alcian blue preparation of ethmoid plate shows clear chondrocyte cytoplasm and rich ECM around rescued cells, and partially restored cartilage shape, which is not seen in R1265P variant (Fig. 6j). These *in vivo* experiments confirm the pathogenicity of the variant and an essential role of the R1265 residue in ECM secretion and cartilage morphogenesis. Taken together, *RIC1* depletion likely leads to *RIC1*-associated CATIFA syndrome through disruption of procollagen secretion.

### Continuum of common and Mendelian phenome in other matrix-related conditions

We analyzed two additional Mendelian conditions to test generalizability of our approach: CLSD and Marfan syndrome, each with extensively-described patient phenome and well-characterized animal models. The BioVU medical phenome associated with reduced GReX of the CLSD-linked *SEC23A* gene successfully predicted 9/10 patient symptoms[16] (Fig. 6k, Supplementary Table 6). Seven of these phenotypes have been recapitulated in *sec23a*$^{-/-}$ zebrafish model[13] (2/3 phenotypes were not observed in zebrafish, i.e., foot deformities and hemangioma of skin, likely because of species-specific differences). For the *FBN1*-linked Marfan syndrome, characterized by cardiovascular abnormalities in patients[49,50], and mouse models[51,52], we observed 7/8 phenotypes associated to reduced GReX of *FBN1* in BioVU, and 6/8 of the phenotypes including cardiomyopathy, hernia and heart failure in published mouse models (Fig. 6l, Supplementary Table 7). These tests support generalizability of the approach presented here.

## Discussion

Collagen constitutes one third of the dry body mass in vertebrates and is essential for the structure and function of a wide-range of organs[53,54]. In disease states such as fibrosis and wound healing, collagen is efficiently secreted to ECM to enclose the injury site. Despite such abundance and clinical importance, the exact mechanism by which procollagen moves through a cell is not well-understood. Here, for the first time, we show evidence that the Ric1-Rgp1-Rab6a module in the TGN is an essential transit point for the collagen traffic *en route* to the ECM.

Disrupted collagen secretion creates a bottleneck, resulting in multi-organ pathologies (e.g., craniofacial dysmorphology, intellectual disability). Despite the strong agreement between the common-disease 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 patient numbers in BioVU or to their young age relative to some phenotypes, which are known to be age-related e.g., intestinal diverticulosis. Furthermore, ADHD and intellectual disability in CATIFA patients might be due to compromised neural ECM composition and remodeling, which are implicated in synaptic plasticity, neuronal guidance, and maintenance of normal neurological functions[40,55]. Thus, a detailed *RIC1*-linked phenome offers a rich phenotype dataset associated with post-Golgi transport deficits of collagen, providing a reference panel for clinicians.

Genotype-phenotype relationships are hardly linear in patients; thus, to describe the full phenome associated with a single gene remains a standing challenge in genetic medicine. Hence, an approach integrating linear genotype-phenotype relationships and population-level variation may provide a more complete assessment of the phenome linked to a single gene.

Model systems allow investigation of molecular mechanisms, but it is hard to assess physiological relevance 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[56,57]. The challenge of explaining disease mechanism is particularly acute in the common-disease space, but also remains prominent in Mendelian diseases.

Here, we describe a novel approach integrating use of animal models, human common-disease biobanks, and rare Mendelian disease to accelerate and refine the discovery process of disease mechanisms (Fig. 6m). BioVU phenome associated with reduced GReX of *RIC1* informed phenome investigations in both the zebrafish model and the human patients homozygous for *RIC1* variant. Molecular mechanisms characterized in zebrafish helped explain the phenome observed in CATIFA patients and BioVU subjects with reduced GReX for *RIC1*. Integration of different sources of data facilitated moving from proximal mechanisms such as reduced gene expression, to cellular mechanisms such as the trafficking of procollagen, to phenotypic mechanisms for gait deficits and swimming behavior resulting from ECM impairment.

Better ability to diagnose the complete disease phenome improves the likelihood of understanding the disease mechanisms. However, clinical evaluation is often limited to the most severe and obvious symptoms patients manifest, thereby hindering the discovery of the full phenotypic spectrum. Initial phenotyping of zebrafish models and re-iteration of gene-associated phenotypes in a biobank led us to "guided clinical re-evaluation" of patients. In our example, craniofacial dysmorphology in zebrafish *ric1*[-/-] mutants and in the BioVU dataset guided examination of skeletal signs in CATIFA patients. In turn, developmental tooth deficits discovered in the biobank prompted teeth re-examination in zebrafish, and led to inclusion as a core CATIFA sign.

Detailed knowledge of the disease mechanisms and spectrum of the associated phenome will enhance opportunities for patient care. For example, preferential use of non-invasive diagnostic methods due to predicted fragility of BM in various organ systems could be prioritized. The discovery of the genetic defects affecting ECM production, and specifically

intracellular trafficking, opens new possibilities for molecular diagnosis in pre- and postnatal care, and later in life. The gene-associated phenome in biobank patients also gives a preview of potential later-onset phenotypes (e.g., Supplementary Table 2) and better prepare clinicians and families for planning future care of patients. As genotype-linked health records grow and quantitative genetics tools[58,59] advance in the upcoming years, this approach will advance personalized medicine practices.

## Methods

### Fish Maintenance and Breeding

Zebrafish were raised under standard laboratory conditions at 28.5°C as previously described[60]. All experiments were conducted according to the guidelines established by the Institutional Animal Care and Use Committee at Vanderbilt University Medical Center (M1700020-00). *round* alleles m641, m713, m715 were isolated and described in the MGH zebrafish mutant screen[7,20] and were kept in the 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).

### Genetic Mapping and Positional Cloning of Zebrafish Mutations

The *round* locus was mapped in an F2 intercross using bulked segregant analysis and mapping. DNA samples were genotyped by PCR using SSLP (Simple Sequence Length Polymorphism) markers evenly spaced across the zebrafish genome[61]. The DNA lesions in mapped *rnd^m641*, *rnd^m713* and *rnd^m715* mutations were identified by Sanger sequencing of genomic DNA from homozygous wild-type F2 animals, heterozygous F2 animals and homozygous mutant F2 animals[10]. Transcripts were confirmed by cDNA sequencing.

Homology analysis presented in Fig. 1 was conducted via Clustal Omega (EMBL-EBI) using zebrafish Ric1 (Zv9 assembly) and human RIC1 primary sequences (Uniprot: Q4ADV7).

### Participants of the CATIFA Studies 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. 2. 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[62]. Linkage analysis of both families confirmed a single significant linkage peak (LOD 3.44) on Chr9 spanning *RIC1*.

### CRISPR/Cas9 Genome Editing of the *ric1* and *rgp1* Loci

CRISPR/Cas9 target sites within the zebrafish *ric1* gene were identified using the CHOPCHOP web tool (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'. Likewise, the following CRISPR/Cas9 target site within *rgp1* gene (GRCz10 assembly) was targeted in this study; 5'-GGTGGTGGCATCTATGGCACGGG-3'. A cloning-free method to generate sgRNA templates was performed as previously described[63]. Guide RNAs (gRNAs) were synthesized with MEGAshortscript™ T7 transcription kit (ThermoFisher Scientific).

To generate mutations with the 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 stages for phenotypic analysis, histology and imaging. 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, which are listed in extended data figures (Extended Data Fig. 4b for *ric1*, and Extended Data Fig. 9b, for *rgp1*).

### BioVU

BioVU is Vanderbilt University Medical Center's DNA biobank, with DNA on >250,000 subjects linked to electronic health records (EHR) going back an average of 10-15 years, and >25 years in some subjects [23]. Although BioVU originally recruited subjects through an opt-out procedure, subjects have been prospectively consented since 2014. For our study, we utilized whole-genome genotype data in ~10,000 participants of European ancestry. Genotyping had been performed using several genotyping platforms and SNP-level imputation was subsequently conducted using *MaCH*[64] on the Michigan Imputation Server in small batches. To determine genomic ancestry, we utilized Principal Components Analysis[65]. We included genetic variants with minor allele frequency (MAF) >1% and imputation quality score >0.80 in downstream analyses.

### Comprehensive PrediXcan Analyses and Gene-based PheWAS

To evaluate the role *RIC1* may play in the etiology of a disease trait, we utilized the PrediXcan[26] approach to build SNP-based prediction models. We estimated the genetically regulated expression (GReX) in the ten thousand BioVU subjects[66] using the GTEx resource (v6p)[27,67] as a reference transcriptome panel (containing paired tissue-specific transcriptome and genotype data), to identify associated phenotypes[68] despite the lack of directly measured gene expression data. Specifically, from the weights $\hat{\beta}_j$ derived from the gene expression imputation model for *RIC1* (consisting of 21 SNPs; see Supplementary Table 1) and the number of effect alleles $X_{ij}$ at the variant *j*, we estimated the genetically determined component of gene expression:

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

in the BioVU subjects (see Supplementary Table 2). Association between the estimated genetic component of gene expression (i.e., GReX) and a disease trait was determined using logistic regression. PrediXcan proposes a causal direction of effect between GReX and trait (because the germline genetic profile is unlikely to be altered by the trait). We use

*phecodes*[25], which aggregate ICD9 or ICD10 codes into a maximum of ~1900 phecodes, in testing the association of the genetically determined expression of *RIC1* to the medical phenome. To maximize the quality of the phenome information, we required at least 2 ICD9 or ICD10 codes on different days to instantiate a phecode for a chronic disease diagnosis. Phenome-wide significance of this gene-based test is based on Bonferroni correction of the number of phenotypes examined (~1100 phecodes have at least 20 cases in the 10,000 subjects analyzed and were tested in our analyses), or $p < 4.5 \times 10^{-5}$, although this is likely to be conservative because of the correlations among phenotypes. The actual sample size for an analysis of a specific phenotype depends on the phenotype (ranged from 4,074 to 9,021 for in RIC1 dataset).

In brief, Phenome-Wide Association (PheWAS) Study we conducted tested GReX against ~1100 phenome codes[25]. In a traditional PheWAS[66], one SNP is used to test association; however, we used a variation of the method which we called "gene-based PheWAS" because we used multiple SNPs from the local (*cis*) regulatory region spanning ~ 1Mb of a genome around the gene of interest.

### UK Biobank and GWAS Analyses

To provide independent support for the phenome associated with reduced *RIC1* expression observed in BioVU, we conducted association analyses on GReX of *RIC1* in the UK Biobank and other public GWAS (summary statistics) datasets and, in parallel, performed analyses of SNP associations (Neale mega-GWAS) in the *RIC1* gene locus in the UK Biobank. For the GReX analyses, we applied models using the tissues (Extended Data Fig. 2a) in which we have high-quality prediction performance of the genetic component of *RIC1* expression. For the analysis of SNP associations, we considered the most significant SNP in the gene; this can be viewed as Tippett's (non-parametric) gene-based combination method that rejects low values of the minimum of SNP p-values[69]. We mapped the BioVU phecodes for the core phenotypes of CATIFA as well as the phenotypes reported as nominally significantly associated in BioVU (Fig. 2d, Extended Data Fig. 2b,c and Supplementary Table 3) to the UK Biobank phenotype (ICD10) codes.

### Genetic Manipulations in Zebrafish

**mRNA overexpression and phenotypic rescue:** 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 a 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 human *RIC1* coding sequence was sub-cloned into pCS2+ vector. Patient variant (R1265P) was introduced with site-directed mutagenesis. These vectors were linearized and used for in vitro transcription to synthesize capped mRNA with the mMESSAGE mMACHINE SP6 transcription kit (ThermoFisher). 1 ng mRNA (either WT hRIC1 or R1265P patient variant) per embryo was micro-injected into 1-cell stage zebrafish embryos. Embryos were raised to 5-7 dpf and scored for craniofacial morphology (with live imaging) and cartilage shape (with Alcian blue staining). Double-blinded, two independent investigators scored the rescue of craniofacial morphology at 5 dpf accounting for the level of jaw protrusion beyond the eye plane. For cartilage shape analysis,

Alcian blue stained cartilage elements from mRNA injected embryos and non-injected controls were dissected with fine-tipped forceps and wet-mounted in 90% glycerol, imaged by an AxioImager.Z1 equipped with a 20X/0.80 Plan Apochromat objective and HRc camera (Zeiss).

**Transgenic constructs:** Transgenic constructs for overexpression experiments (OE) were generated with the Gateway cloning-based Tol2kit method[44]. Full length *RIC1* was PCR-amplified with attB1(forward) and attB2 (reverse) recombineering sites and cloned into pDON221. The obtained clone was utilized as a middle entry vector in the LR cloning step of the Multistep Gateway recombineering system, in combination with the p5E-1.7kb Col2a1a promoter[70], p3E-v2A-EGFP vector (gifted by Dr. J.T. Gamse), pDestTol2pA2 and incubated with Gateway® LR clonase® II enzyme mix. The resulting destination clone was used for mosaic overexpression experiments. v2A-EGFP tag was used to label transgenic cells with cytoplasmic EGFP signal that forms upon self-cleavage of v2A-EGFP from the fusion polypeptide[48]. The R1265P *RIC1* variant identified in CATIFA patients was introduced into the destination vector via Q5® site-directed mutagenesis kit (NEB). Zebrafish *rab6a* was cloned from 4 dpf wild type (AB) cDNA into the vector pEGFP-C1. Next, the EGFP-Rab6a fusion was sub-cloned into the pDONR221 vector via BP cloning, after adding attB1 and attB2 sites with PCR amplification. The obtained clone was utilized as a middle entry vector in the LR cloning as described above. p5E-1.7kb Col2a1a promoter[70], p3E-ployA, pDestTol2pA2 were used to build *rab6a* destination vector. The Q72L mutation (constitutively active Rab6a mutant[42]) was generated on the destination vector via the Q5® site-directed mutagenesis kit (New England Biolabs). Primers used are listed in Supplementary Table 8.

**Microinjection of Tol2kit clones for mosaic overexpression.—**Zebrafish embryos at 1-cell stage were injected with a combination of 50 pg medaka transposase mRNA[71] and 10-20 pg of destination vector DNA as described previously[72], and grown to indicated stages in embryo medium. Mosaic clones were detected by EGFP signal.

**Histology, Transmission Electron Microscopy (TEM) and Tissue Staining:**

**Histology:** Histological sections were prepared with JB-4 embedding kit as previously described[73]. 4-8 μm thick sections were stained with 1% toluidine blue.

**Transmission electron microscopy & quantification of vesicles in zebrafish tissues:** Samples were processed for TEM as previously described[10]. 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. Cartilage, notochord and muscle tissues were imaged 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. Width and height of vesicles were measured using ImageJ's 'Measure' tool.

**Cartilage and bone staining:** 5 or 7 dpf embryos were stained for bone (Alizarin red) and cartilage (Alcian blue) as described previously[74]. 0.02% Alcian blue and 60 mM MgCl$_2$ in 70% ethanol, 0.005% alizarin red was used as a double staining solution. Stained embryos were imaged with Zeiss Axioimager.Z1 scope equipped with an Axiocam HRc camera. Number of analyzed embryos for each experiment were indicated in corresponding figure panels.

### Live Imaging and Immunofluorescence

**Live imaging of zebrafish embryos and morphometric analysis:** 5 dpf embryos were anesthetized in Tricaine (Sigma), mounted in 3% methylcellulose (Sigma) and imaged using a 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.

**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). Movies were generated using Imaris' movie tool to help with the 3D view.

**Immunofluorescence (IF) and fluorescence microscopy:** 14-μm thick cryo-sections of zebrafish tissues were obtained by Leica CM-1900 cryostat. IF on cryo-sections was performed as previously described[72]. The 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), anti-Col1 (abcam, ab34710), anti-myosin (Developmental Studies Hybridoma Bank, #MF20). 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); muscle staining on sections was with Alexa Fluor 488-conjugated phalloidin (ThermoFisher, #A12379). Fluorescence imaging was carried out using AxoImager 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, the intracellular area of each cell (demarcated by caax-EGFP signal), the collagen area (Col2 signal) and the 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 the maximum intensity projection views were drawn by tracing caax-EGFP signal and sub-selection images were produced. The total area of sub-selection was measured as total intracellular area. The total collagen area was measured by tracing the Col2 signal in individual cells and adding them up. Percent collagen area per cell was calculated by the following formula:

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

**Whole-mount IF:** Zebrafish embryos at the specified stages were fixed in Prefer Fixative (Anatech) for 48 hours, rinsed in PBS-0.1% tween 20, post-fixed in 4% PFA for 10 minutes, and then permeabilized in 0.5% triton X-100 in PBS. Next, embryos were bleached with 1.5% $H_2O_2$ / 1% KOH, then rinsed, blocked in 10% goat serum, 1% DMSO, 1% BSA and incubated in primary antibody (1:250 dilution) overnight at 4°C. Secondary antibody incubation (1:300 dilution) with fluorophore-conjugated antibodies was followed at 4°C overnight. After rinsing in PBS-0.1% tween 20, DAPI staining was performed (1:5000 dilution), and then embryos were post-fixed in 4%PFA for 20 minutes, afterwards, cleared and stored in 90% glycerol. Stained embryos were imaged on a Nikon Spinning Disk Confocal Microscope. Images are presented as maximum intensity z-projections generated by Nikon Elements software. Primary antibodies used are: α-myosin (Developmental Studies Hybridoma Bank #MF20), α-thrombospondin-4 (abcam #ab211143), α-acetylated tubulin (Sigma #T7451), α-HuC/D (ThermoFisher #A-21271). Secondary antibodies used are: anti-Mouse-Alexa Fluor 488 (ThermoFisher #A-11001), anti-Mouse-Alexa Fluor 555 (ThermoFisher #A28180), anti-Rabbit Alexa Fluor 488 (ThermoFisher #A-11008) and anti-Rabbit Alexa Fluor 555 (ThermoFisher #A-21428). Alexa Fluor 488-conjugated Wheat Germ Agglutinin (WGA), ThermoFisher #W11261) was performed at the secondary antibody incubation step with 1:300 dilution in blocking solution. DAPI counterstaining was performed after secondary antibody step, by incubating for 20 minutes in 1:5000 diluted DAPI in PBS-0.1% tween 20.

**Brain Anatomy Measurements:** Whole mount zebrafish embryos stained with HuC/D and WGA were imaged on Nikon Spinning Disk Confocal Microscope, with Plan Apo Lambda 10X objective, 0.45NA. Maximum intensity z-projections (150-200 μm depth) were created in Nikon Elements software and analyzed in ImageJ using Freehand selection and measurement tools. Several embryos were processed and statistically analyzed with two-tailed Mann-Whitney U-test (CI = 95%).

**Lens imaging:** Lens imaging was performed as previously described [75]. Briefly, whole intact eyes were dissected from live, anesthetized fish larvae and examined under DIC illumination on Zeiss AxioImager.Z1 equipped with 20X/0.80 Plan Apochromat objective and Flash4.0 camera (Hamamatsu). For transmitted light imaging, lens was further dissected

from eye and image was acquired by AxioCam HRc (Zeiss) camera on AxioImager.Z1equipped with 20X/0.80 Plan Apochromat objective.

## Human cell culture, TEM and IF analysis

**Culture conditions:** Dermal fibroblasts from CATIFA patient (15DG2428) was isolated from skin biopsy and cultured in MEM (GIBCO) supplied with bovine serum at 37°C with 5% $CO_2$. BJ skin fibroblasts (ATCC® CRL-2522™) were used as the control for all experiments described further.

**TEM:** Cells grown in 10 $cm^2$ dishes to confluency were fixed in 2.5% glutaraldehyde and in 0.1 M sodium cacodylate by incubating at room temperature for 1 hour first, then overnight at 4°C, and were then further processed for TEM imaging in Vanderbilt Cell Imaging Shared Resource core.

**IF:** Prior to collagen-I staining, cells were treated with ascorbate for 1 hour at 37°C with 5% $CO_2$ as described previously[76]. Cells were fixed in 4%PFA at room temperature for 20 minutes, permeabilized in 1% triton X-100 in PBS, then blocked in 1% BSA. 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 the nuclear counterstain. Slides were imaged with the 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 cells:** Fluorescence intensity of collagen-I signal per cell and total cellular area were measured in ImageJ using the '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. Colocalization analysis of Collagen-I and the p230 (TGN marker) was performed using the Coloc2 tool in ImageJ. Pearson's r-values were recorded from multiple cells and compared using the Student's t-test.

## Quantitative Transcript Analyses

*Quantitative real-time PCR (qRT-PCR)* was performed as described previously[10]. Using TRIzol reagent (ThermoFisher Scientific), total RNA was extracted from 12-15 zebrafish embryos for each sample at the specified developmental stage. The TRIzol method was performed to extract total RNA from cultured fibroblasts as well, following the manufacturer's manual. 500 ng of total RNA was used as a template to make cDNA through reverse transcription using M-MLV reverse transcriptase (Promega) and poly-T primer. Each qRT-PCR reaction was performed with 20 ng of total cDNA, SYBR Green Real-Time PCR Master Mix and 2 μM of each primer. Primers used are listed in Supplementary Table 8. qRT-PCR reactions were run on CFX96 (Bio-Rad) system. Data were analyzed with the - Ct method, Student's t-test (CI = 95%). Each sample was independently obtained from

three independent zebrafish clutches. A similar approach was taken for qPCR for cultured human fibroblasts

**Fluorescence-associated cell sorting (FACS):** GFP+ and GFP− cells were sorted from Tg(Col2a1a:caax-eGFP)[70] transgenic zebrafish embryos at 3 or 4 dpf stages with FACS Aria II Flow cytometer (BD) as described by Unlu et al (unpublished). Total RNA was isolated from corresponding cell population and used for RNA sequencing. Approximately 100 embryos were used to generate each sample.

**cDNA library preparation and next generation RNA sequencing:** cDNA libraries were processed for RNA-seq by the Vanderbilt Technologies for Advanced Genomics (VANTAGE) core facility. 30 million 50-bp single end reads were collected per sample. Two independent samples for each experimental group were included as replicates.

**Sequence alignment and differential expression analysis:** Reads were aligned to the zebrafish transcriptome (GRCz10 release 80) and genome (GRCz10) with Tophat2[77] default parameters. Differentially expressed genes were called with DESeq2[78] in Bioconductor[79].

### Locomotor Behavior Analysis

2 WT and 2 *ric1*[−/−] mutant zebrafish embryos at 4 dpf stage were transferred into a 27-cm glass bottom dish filled with embryo medium, up to the level of glass-dish barrier. Locomotion behavior of the embryos was recorded for 14-seconds, 0.07 sec / frame with an HRc camera mounted on Stemi-2000C stereoscope (Zeiss) at 0.65X magnification. Movie recording was performed by ZEN2 software (blue edition) with 1×1 binning, 1 ms exposure, B/W setting, at 1388 × 1040-pixel resolution. Movies were opened in ImageJ and analyzed with a MTrackJ plug-in. The position of each embryo was manually tracked frame by frame. Total displacement of embryos was obtained with MTrackJ's Measure tool. The total displacement / total time (14 s) formula was used to calculate mean velocity. Distance to previous position (D2P) is also recorded by MTrackJ for each single displacement event. Maximum D2P value (MaxD2P) of each embryo was used to calculate 'maximum velocity reached during the travel' by the following formula: MaxD2P/ time between frames (0.07 s). A total of 8 videos were generated, accounting for 16 WT and 16 mutant embryos. Initial (x, y) coordinates (generated by MTrackJ, Measure Tool, Points sheet) of each embryo was taken as (0, 0) origin point and all following coordinates were adjusted with reference to the origin, creating relative coordinates values. Relative (x, y) coordinates were plotted in Prism 7 (GraphPad) to display individual tracks. Swim behavior patterns of embryos were scored by analyzing individual frames and categorized as described previously[37].

### Statistics

For head size indices, collagen accumulation by IF and qPCR data, significance was assessed by non-parametric, two-tailed, Mann-Whitney *U*-test for comparison of two groups. Statistical analysis was performed with Prism 7.0 (GraphPad). Confidence interval (CI) is 95%. Mean and standard deviation bars are indicated on graphs. The data represented as box-whisker plots show the range between 25th and 75th percentile as boxes, minimum

and maximum values as whiskers with median at the center. Collagen content and colocalization in human fibroblasts were statistically analyzed by two-tailed Student's t-test using Prism 7.0 (GraphPad). All zebrafish assays were performed on three or more embryos unless otherwise specified; number of embryos analyzed for each assay was noted in figure panels as 'n'.

### Data Availability

Summary statistics from SNP-level and gene-level (PrediXcan) associations are available in Supplementary Tables. The protected data for the GTEx project (for example, genotype and RNA-sequence data) are available via access request to dbGaP accession number phs000424.v6.p1. All requests for raw (e.g., genotype and phenotype) data and materials are reviewed by Vanderbilt University Medical Center to determine whether the request is subject to any intellectual property or confidentiality obligations. For example, patient-related data not included in the paper may be subject to patient confidentiality. Any such data and materials that can be shared will be released via a Material Transfer Agreement. All gene, transcript and protein sequences can be found at NCBI and Ensembl databases (accession numbers are provided in Methods section).

## Extended Data
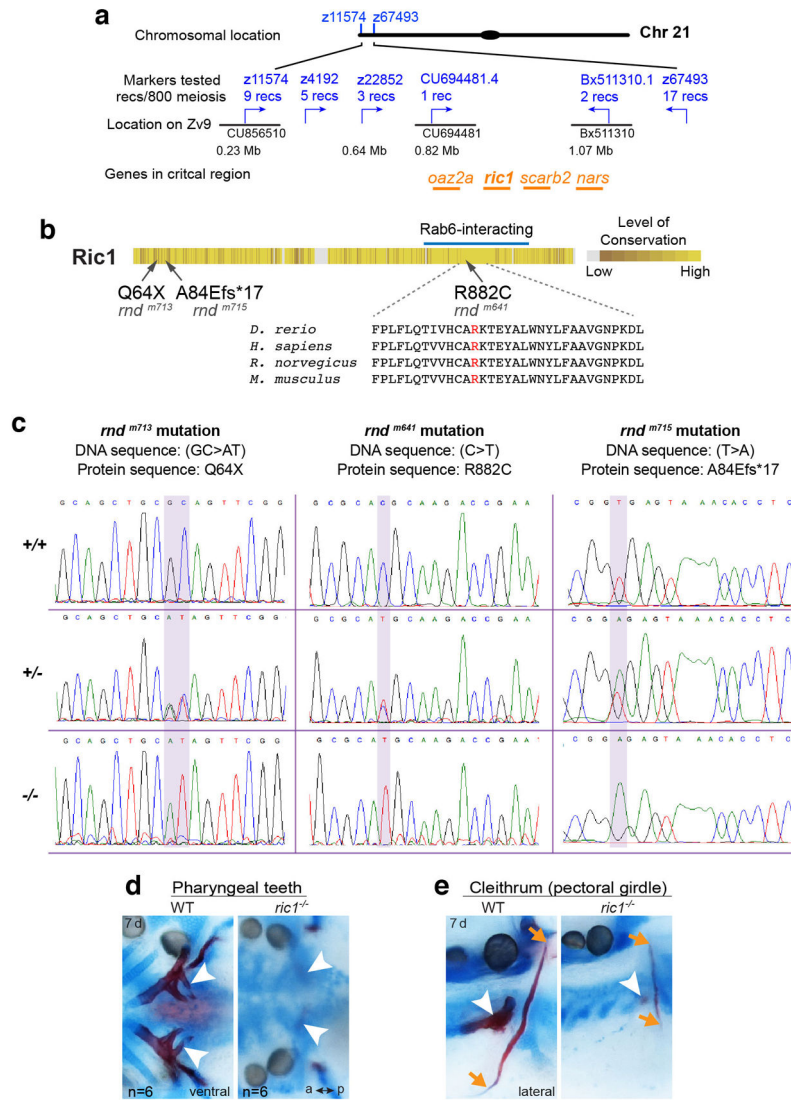
**Extended Data Fig. 1. Genetic analysis of *ric1* mutations and dental phenotypes**
**a,** Genetic linkage analysis and positional cloning mapped the *rnd* mutation to zebrafish chromosome 21. Microsatellite markers and number of recombination events (recs, blue) in critical region that contains 4 protein-coding 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. Multiple sequence alignment shows highly conserved R882 residue (red) of the Rab6-interacting region 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 sites. 10 independent embryos of each genotype were analyzed and all revealed identical results as presented here. **d,** Ventral view of the teeth (arrows) stained by Alcian blue (cartilage) and Alizarin red (ossification), n=6 animals per group, **e,** lateral view of the pharyngeal teeth (white arrows) on 7th

pharyngeal arch, n=6 animals per group. Orange arrows point to lack of cleithrum bone elongation in *ric1*$^{-/-}$ mutant embryo, 7 dpf; a: anterior, p: posterior.
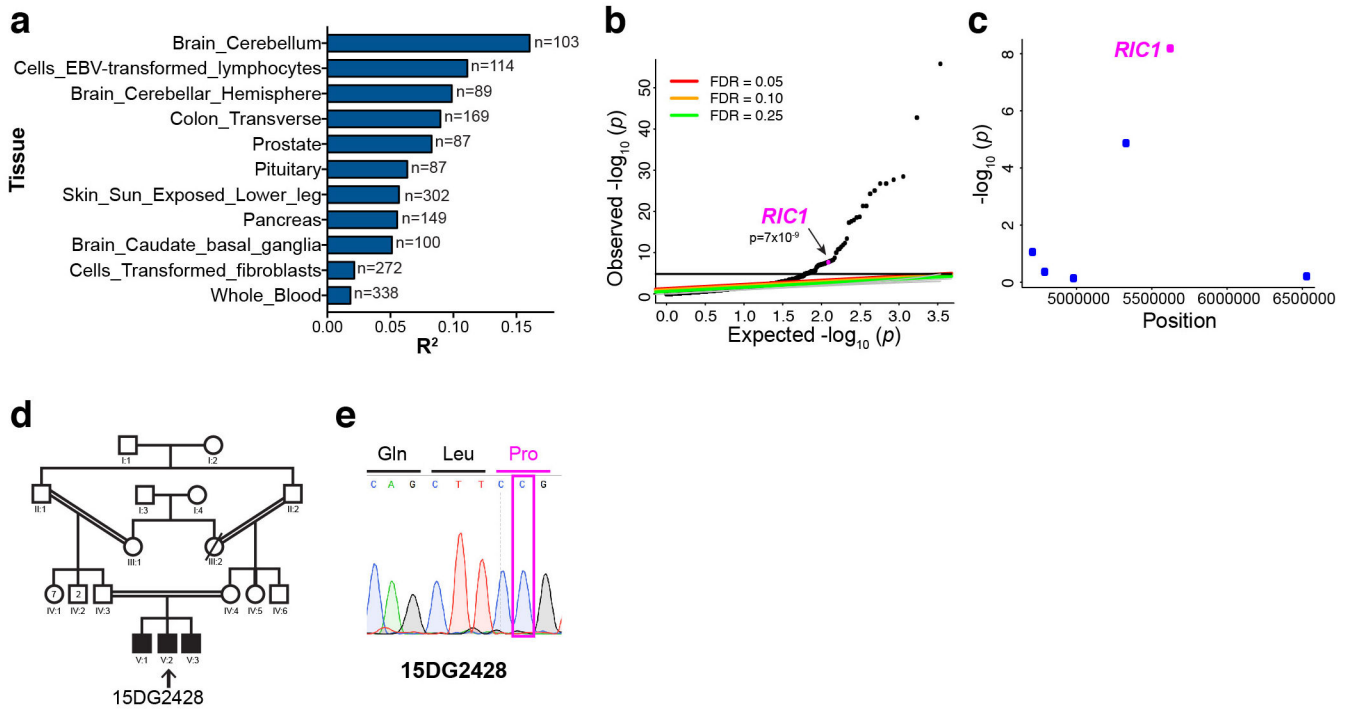
**Extended Data Fig. 2. Genetic analysis of RIC1 in biobanks and additional CATIFA patients**
**a,** Correlation of imputation ($R^2$) values for *RIC1* expression across tissues. Sample sizes are indicated as 'n' next to each corresponding tissue on the graph. **b,** Q-Q plot showing the distribution of all association p-values with asthma in UK biobank. *RIC1* association to asthma is statistically significant with $p=7\times10^{-9}$. Significance threshold is shown with black line set to $p=1.5\times10^{-5}$. **c,** Association of genes 4Mb around RIC1 locus with asthma. Note that association of RIC1 reduced expression with asthma is the most significant within this region. **d,** Pedigree of the additional family contributing to this study (modified from Patel et al., 2017[30]). Standard pedigree symbols used; affected individuals are shaded, arrow points to patient 15DG2428 whose skin biopsy fibroblasts were sequenced (**e**), and analyzed by cellular and molecular methods. The mutation site is boxed, and results in Arg to Pro substitution in the protein sequence.
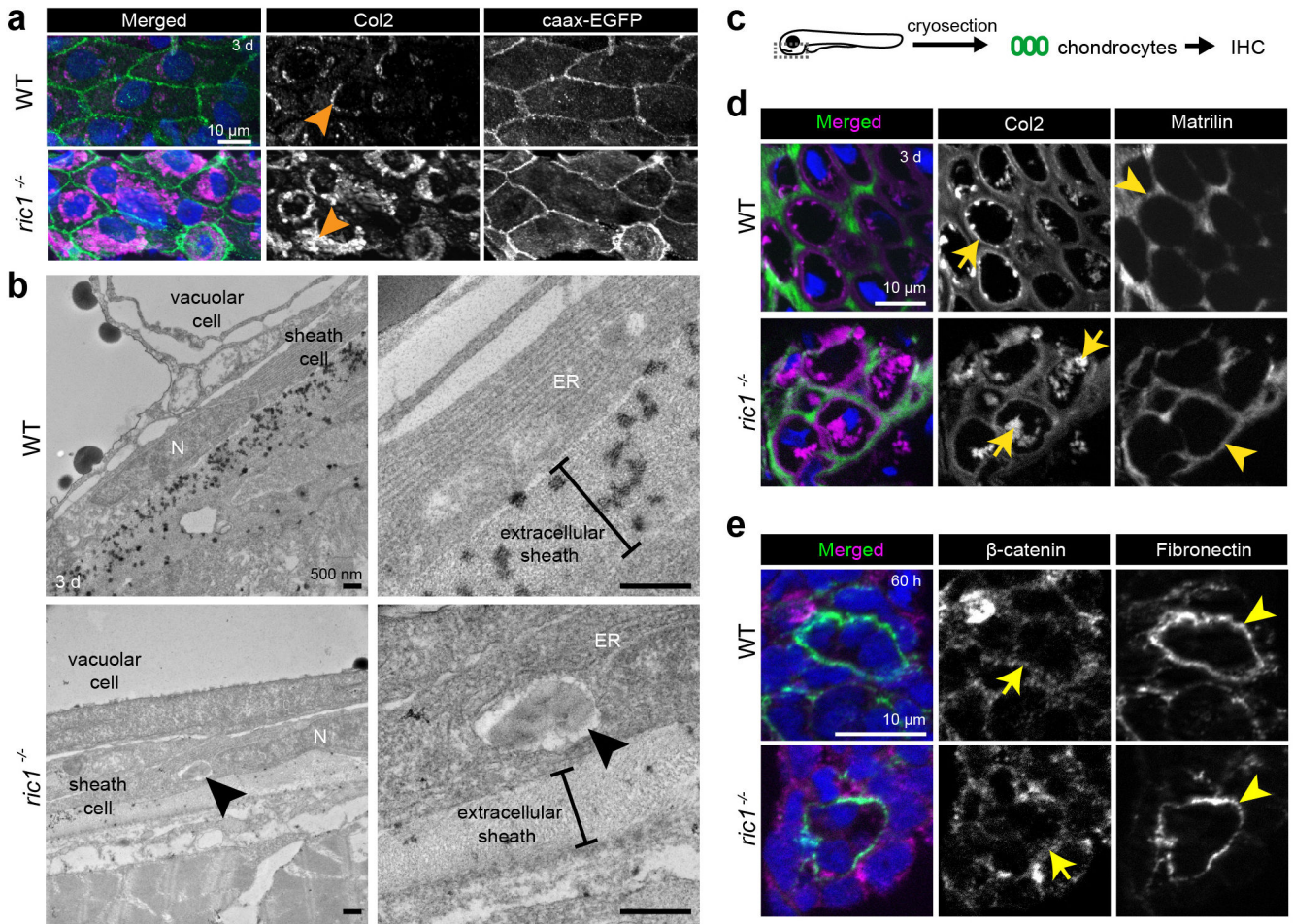
**Extended Data Fig. 3. Analysis of matrix components in sheath cells and notochord basement membrane (BM)**

**a,** Representative images of Tg(Col2a1a:caax-EGFP) transgenic zebrafish stained with Col2 antibody in WT and *ric1*$^{-/-}$ notochord sections (Collagen-2, magenta; EGFP, green). Note accumulation of intracellular Col2 in mutant sheath cells (arrowheads). 3 animals per group were analyzed and similar results were confirmed. **b,** TEM images of notochord tissue at 3 dpf. Magnified views of sheath cells and extracellular sheath / basement membrane (BM) are displayed on the right panels. Arrowheads point to intracellular inclusions. ER: endoplasmic reticulum, N: nucleus. Size bars = 500 nm. 3 cells per genotype were imaged and similar observations were confirmed. Collagen II fails to be transported out of Ric1-deficient chondrocytes, while other ECM cargos are secreted normally. **c,** Experimental design for immunohistochemistry (IHC) analysis with antibodies against Col2, matrilin, β-catenin and fibronectin epitopes on cryosections. **d,** Col2 (arrows) and Matrilin (arrowheads) co-immunostaining in 3 dpf chondrocytes. 3 animals per group and 5 cells per animal were analyzed and similar results were confirmed. **e,** β-catenin (arrows) and Fibronectin (arrowheads) co-immunostaining at 60 hours post fertilization (hpf) stage. DAPI (blue) marks nuclei. 3 animals per group and 5 cells per animal were analyzed and similar results were confirmed.
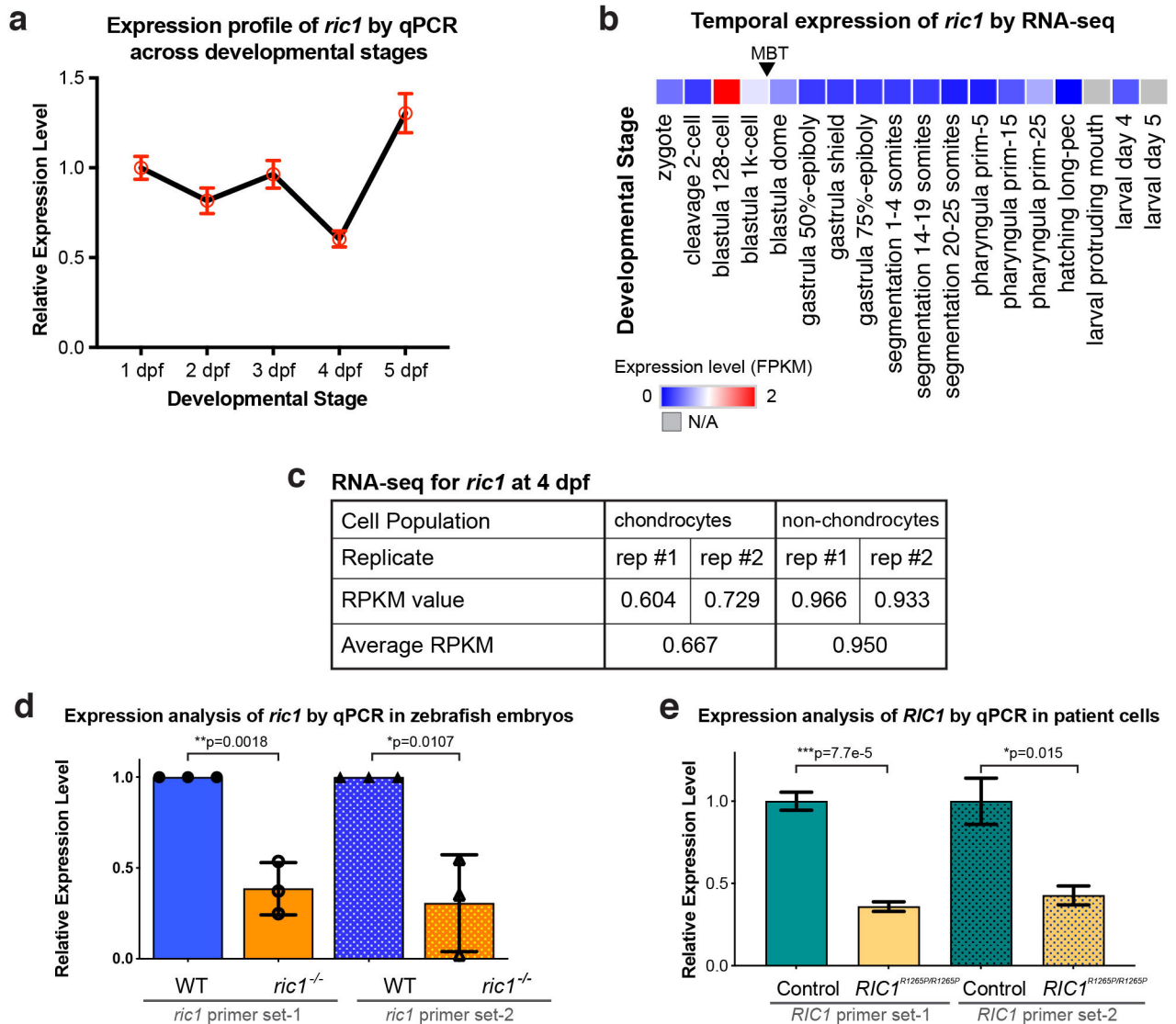
**a**



WT | *ric1^gRNA* mosaic mutant

Col2 WGA DAPI

Col2

*ric1^-/-*  WT  WT

WGA
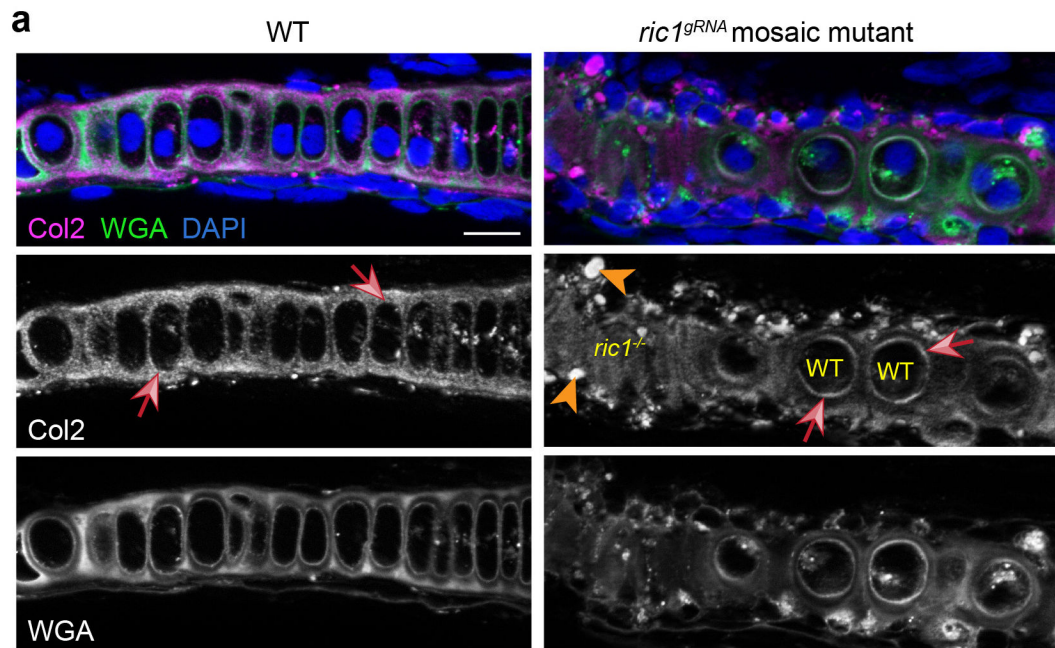
**b**

## CRISPR/Cas9 edited sites detected in *ric1*

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

**Extended Data Fig. 4. 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. Arrowheads in *ric1^gRNA* mosaic mutant (*ric1^-/-* chondrocyte clones in the WT embryo injected with gRNA targeting *ric1*) point to collagen accumulations. Arrows indicate secreted, extracellular Col2. DAPI (blue) marks nuclei. This experiment was repeated with similar results using 4 animals per each genotype. **b,** Representative mutations detected in *ric1^gRNA* mosaic mutants by direct sequencing. PAM: Protospacer adjacent motif.

**a**

**Expression profile of *ric1* by qPCR across developmental stages**



**b**

**Temporal expression of *ric1* by RNA-seq**



**c**

**RNA-seq for *ric1* at 4 dpf**

| Cell Population | chondrocytes | | non-chondrocytes | |
|---|---|---|---|---|
| Replicate | rep #1 | rep #2 | rep #1 | rep #2 |
| RPKM value | 0.604 | 0.729 | 0.966 | 0.933 |
| Average RPKM | 0.667 | | 0.950 | |

**d**

**Expression analysis of *ric1* by qPCR in zebrafish embryos**



**e**

**Expression analysis of *RIC1* by qPCR in patient cells**



**Extended Data Fig. 5.** *Ric1* **expression in zebrafish and *RIC1*$^{R1265P/R1265P}$ patient fibroblasts**
**a,** Temporal expression profile of *ric1* across zebrafish developmental stages by PCR. 3 independent experiments were run and presented here. Each experiment contained pooled RNA from 10 embryos of the indicated genotype and stage. Center line indicates the mean; bars show standard error of the mean (SEM). **b,** Temporal expression profile of *ric1* by RNA-seq from whole embryos (data from Expression Atlas, White et al., 2017[31]). **c,** Gene expression analysis of *ric1* in sorted, live chondrocytes and non-chondrocytic cells by RNA-seq. R/FPKM: reads/ fragments per kilobase of transcript in million reads. **d,** QPCR analysis of *ric1* in whole zebrafish embryos at 5 dpf stage, replicated with two primer sets. 3 independent experiments were run and presented here. Each experiment contained pooled RNA from 10 embryos of the indicated genotype and stage. Center line indicates the mean; bars show SD. **e,** Expression analysis of *RIC1* in CATIFA patient's fibroblasts (*RIC1*$^{R1265P/R1265P}$) by qPCR. Two independent primer sets were used for each assay. QPCR data in **d** and **e** were normalized to β-actin or GAPDH. Mean expression levels

relative to control (WT) are reported in center lines, SD bars are indicated. Two-tailed Mann-Whitney U-test, CI = 95%.
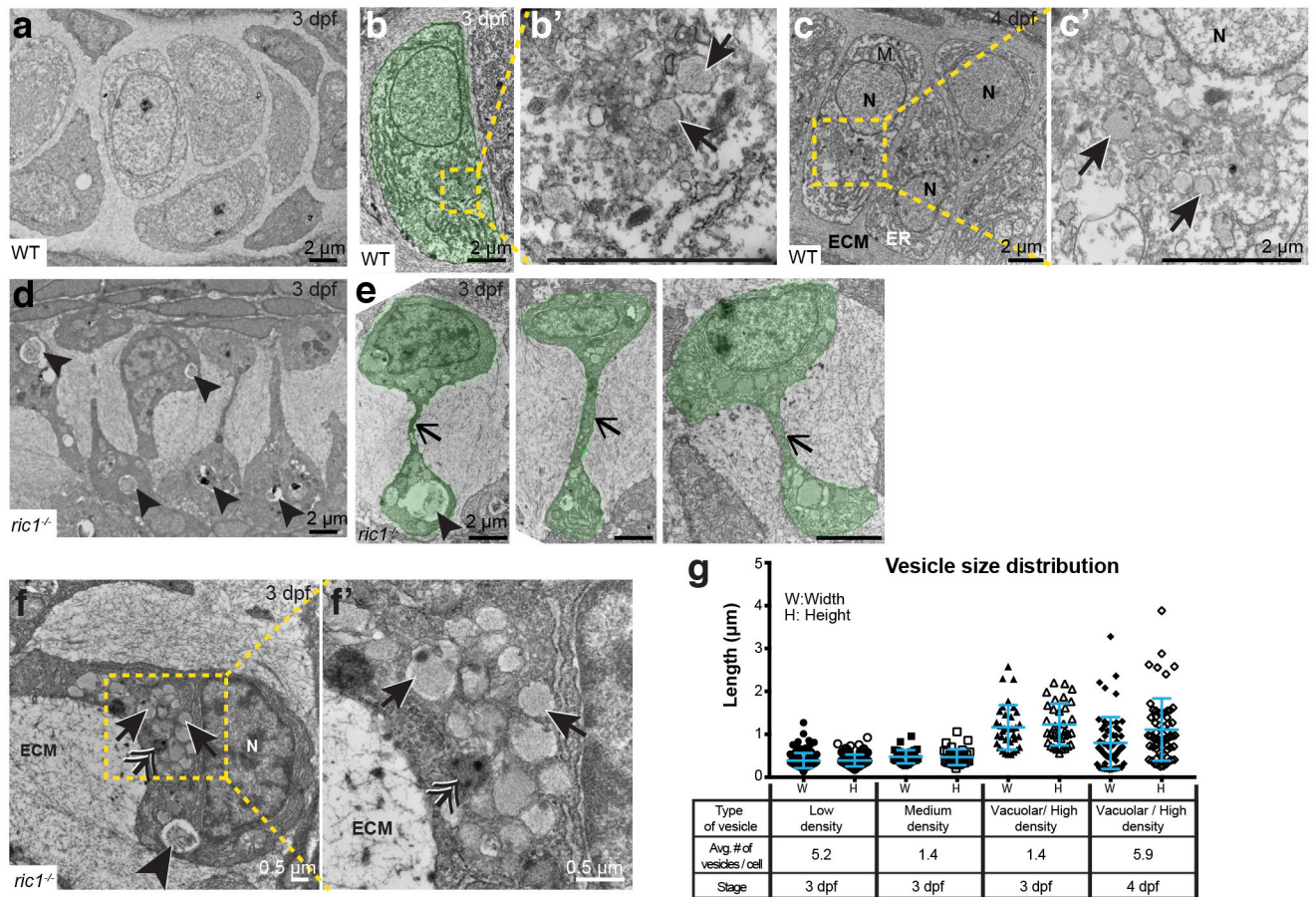
**Extended Data Fig. 6. Constricted cell shapes of *ric1*⁻/⁻ chondrocytes at 3 dpf and normal ultrastructure of WT cells by TEM at 3 and 4 dpf**

**a,** TEM image of WT craniofacial cartilage showing normal chondrocyte shapes. **b,** A representative WT chondrocyte at 3 dpf **b',** Higher magnification reveals vesicular structures associated with normal secretion (arrows). **c,** A representative TEM image of WT chondrocytes at 4 dpf. **c',** Magnified view of boxed area is shown on the right, arrows point to normal vesicular compartments. Intracellular vesicles accumulate in *ric1*⁻/⁻ chondrocytes. **d,** *ric1*⁻/⁻ craniofacial cartilage chondrocytes contain large inclusions (arrowheads). **e,** Representative examples of chondrocytes constricted at the midline (arrows), green overlays on **b** & **e** are drawn after image acquisition to help with cell shape demarcation. **f,** TEM micrograph of a representative 3 dpf *ric1*⁻/⁻ 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). **f',** Magnified view of the boxed region from **f. g,** Size distribution and average number of vesicles in 3 and 4 dpf *ric1*⁻/⁻ chondrocytes. (20 cells at 3dpf; 10 cells at 4dpf stage). Vesicle diameters were measured both in parallel (W: width) and perpendicular (H: height) dimension to image plane; and plotted as length (μm) values. ECM: extracellular matrix, ER: endoplasmic reticulum, M: mitochondrion, N: nucleus. Mean and SD values are indicated with bars. Three *ric1*⁻/⁻ embryos at 3 dpf and one at 4dpf were imaged with similar results. One

representative WT of each stage was imaged with similar results across 10 cells imaged per embryo at each stage analyzed.
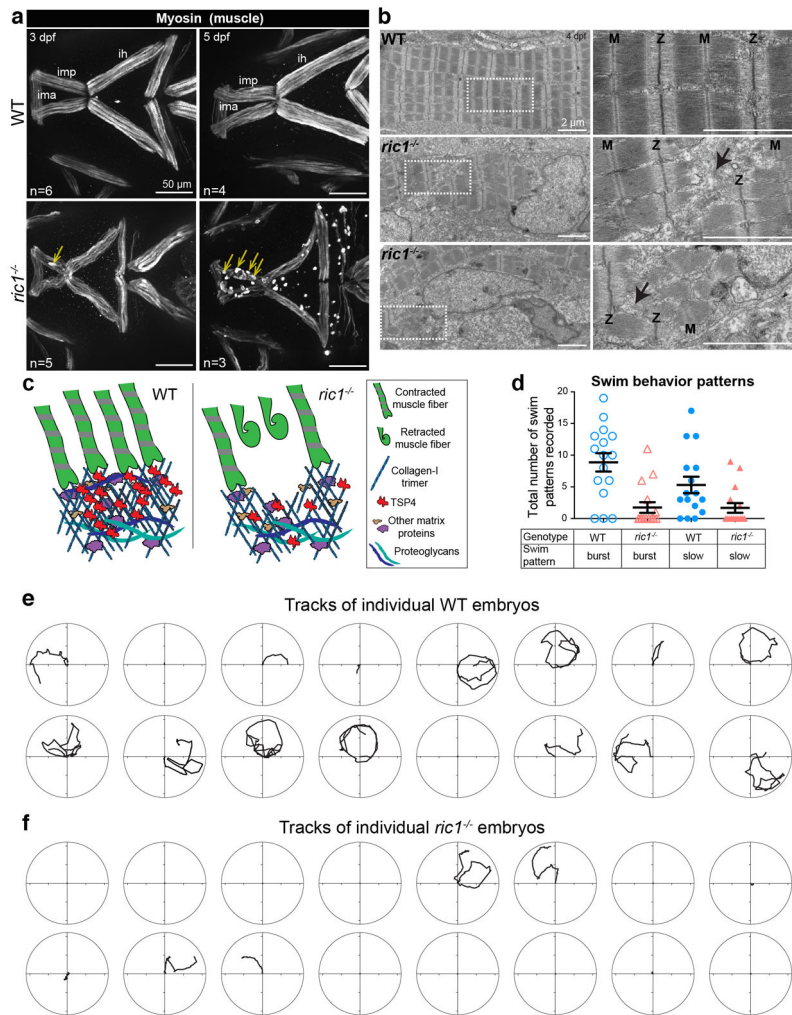
**Extended Data Fig. 7. Muscle attachment defects in *ric1*⁻/⁻ zebrafish**

**a,** Whole-mount zebrafish embryos immunostained for myosin. Ventral head views are displayed. Arrows point to retracted fibers, in the form of bright punctate structures. ih: interhyoideus, imp: intermandibularis posterior, ima: intermandibularis anterior. **b,** Transmission electron micrographs of cranial muscles. Note non-continuous muscle fibers in *ric1*⁻/⁻ (arrow). M: M-line, Z: Z-line. Right panels show magnified views of boxed regions in the left panels (2 independent mutant examples, one representative WT). **c,** Schematic of tendon matrix and consequent muscle fiber attachment defects. **d,** Quantification of swim behavior patterns, categorized as either burst (fast, sharp-angle turn) or slow (slower routine, wide-angle motion) swim pattern (Budick & O'Malley, 2000). Mean and SEM are shown as bars. Each data point represents a single embryo, n=16 independent embryos per genotype. Locomotion tracks of individual embryos, **e,** WT embryos, **f,** *ric1*⁻/⁻ mutant embryos.
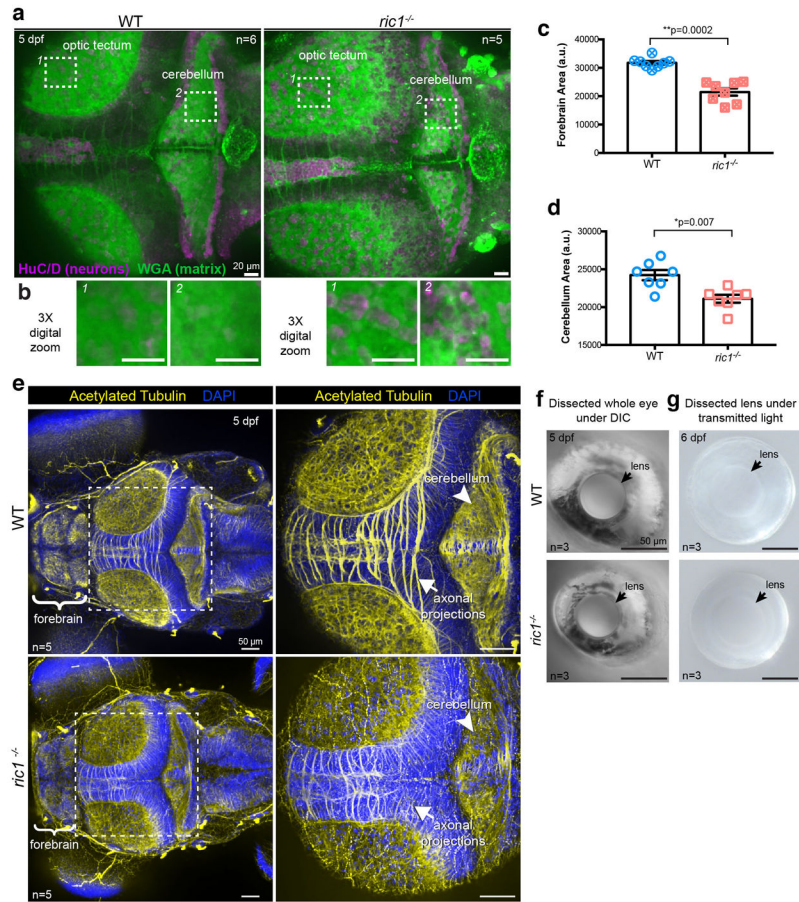
**Extended Data Fig. 8. Analysis of brain and eye phenotypes in $ric1^{-/-}$ zebrafish**

**a**, Whole mounted zebrafish embryos immunostained for HuC/D (pan-neuronal marker) and WGA (marking N-glycosylated proteins in neural ECM) at 5 dpf (Higher objective view of the same embryos in Fig. 4f). **b**, Digital zoom-in views showing sparser matrix in optic tectum and cerebellum regions of $ric1^{-/-}$. **c,** Forebrain (n=8 independent animals per genotype) and **d**, Cerebellum area measurements (n=7 independent animals per genotype) performed by ImageJ, Measure Tool using whole mount immunostained images. Arbitrary units (a.u.) were plotted. Bars indicate the mean, lines are SD. Two-tailed Mann-Whitney U-test, CI = 95%. **e**, Acetylated tubulin whole-mount immunostaining that labels axonal projections. Right panel shows magnified views. Notable poor bundle structure of axonal projections in $ric1^{-/-}$ (arrows) and sparse axon tracks in cerebellum (arrowheads). **f**, Images of dissected whole zebrafish eyes under DIC illumination. **g**, Transmitted light illumination images of dissected zebrafish lenses. Note comparable smooth lens surface of WT and $ric1^{-/-}$ in both panels. n: number of embryos analyzed.
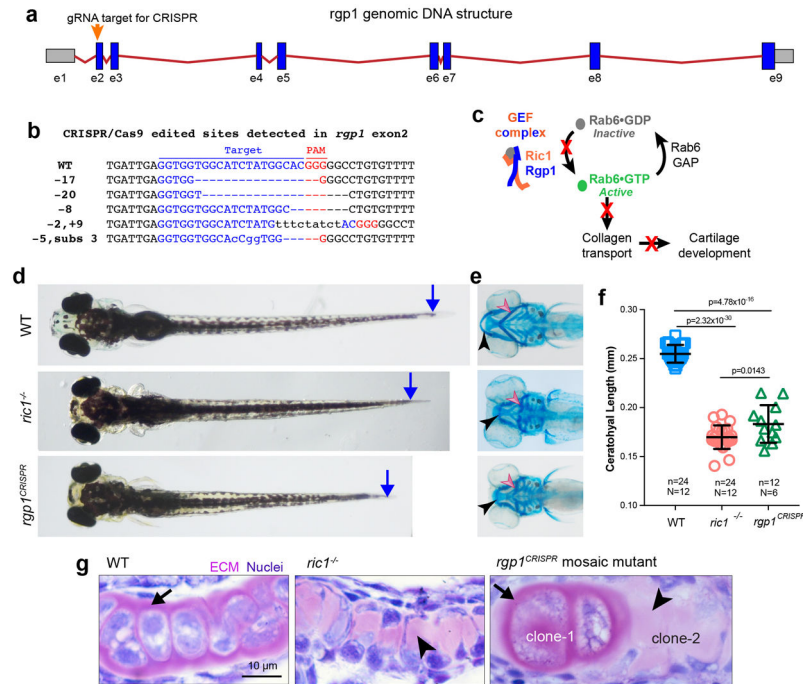
**Extended Data Fig. 9. Rgp1-depletion recapitulates *ric1*$^{-/-}$ mutant phenotype**
**a,** Schematic for genomic structure of *rgp1* gene. Exons are shown as boxes, protein coding in blue and untranslated in gray, introns as red lines. Orange arrow marks guide RNA target site within exon 2. **b,** 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, two-tailed, CI= 95%). Central line shows mean; bars indicate SD. **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. This experiment was repeated with similar results from 3 independent animals from each indicated genotype.
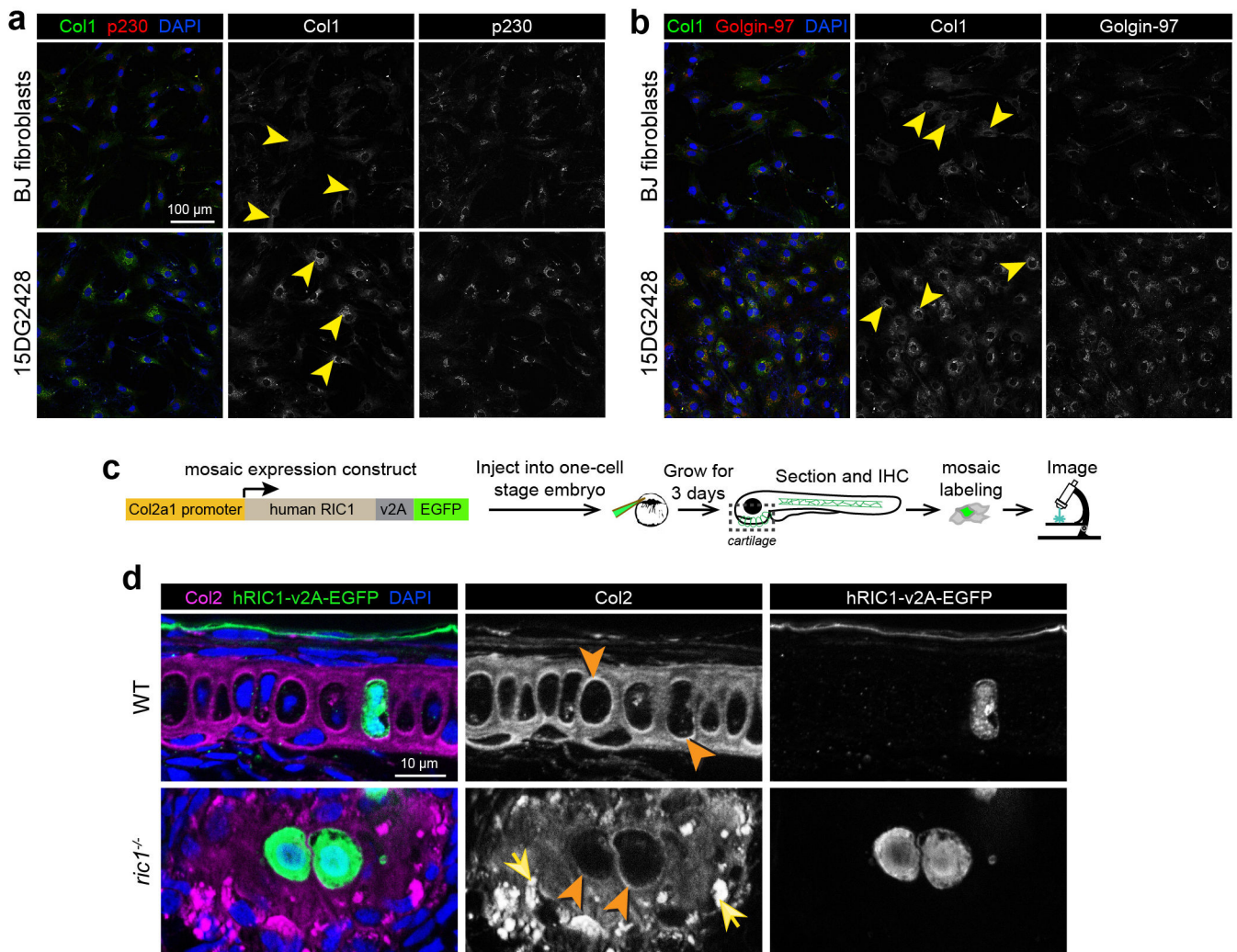
**Extended Data Fig. 10. Collagen-I accumulates in TGN-associated compartments in $RIC1^{R1265P/R1265P}$ patient's fibroblasts**

**a,** Representative images (of 3 experiments of independently grown cell cultures) for co-IF of BJ fibroblasts (control) and 15DG2428 ($RIC1^{R1265P/R1265P}$ 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 3 experiments of independently grown cell cultures) for co-IF with Col1 and alternative TGN marker Golgin-97 taken under the same conditions. Arrows point to TGN-associated Col1 signal. Overexpression of human RIC1 restores collagen secretion in *ric1*-deficient zebrafish chondrocytes. **c,** Experimental design to mosaically overexpress human *RIC1* (hRIC1) gene in zebrafish. EGFP is linked to hRIC1 via self-cleavable viral 2A (v2A) peptide. **d,** Co-immunostaining of Col2 and EGFP in cartilage. Green cells express the construct and neighboring cells act as endogenous WT or mutant controls. Arrows point to collagen accumulation; arrowheads show secreted, extracellular collagen. Representative of 3 cells per group, with similar results.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

## References

1. Klein C & Gahl WA Patients with rare diseases: from therapeutic orphans to pioneers of personalized treatments. EMBO Mol Med 10, 1–3 (2018). [PubMed: 29180354]

2. Maddirevula S et al. Expanding the phenome and variome of skeletal dysplasia. Genet Med, 12, 1609–1616 (2018).

3. Arnold WV & Fertala A Skeletal diseases caused by mutations that affect collagen structure and function. Int J Biochem Cell Biol 45, 1556–1567, (2013). [PubMed: 23707199]

4. Luderman LN, Unlu G & Knapik EW Zebrafish Developmental Models of Skeletal Diseases. Curr Top Dev Biol 124, 81–124, (2017). [PubMed: 28335865]

5. Unlu G, Levic DS, Melville DB & Knapik EW Trafficking mechanisms of extracellular matrix macromolecules: insights from vertebrate development and human diseases. Int J Biochem Cell Biol 47, 57–67, (2014). [PubMed: 24333299]

6. Canty EG & Kadler KE Procollagen trafficking, processing and fibrillogenesis. J Cell Sci 118, 1341–1353, (2005). [PubMed: 15788652]

7. Driever W et al. A genetic screen for mutations affecting embryogenesis in zebrafish. Development 123, 37–46 (1996). [PubMed: 9007227]

8. Knapik EW ENU mutagenesis in zebrafish--from genes to complex diseases. Mamm Genome 11, 511–519 (2000). [PubMed: 10886014]

9. Latimer A & Jessen JR Extracellular matrix assembly and organization during zebrafish gastrulation. Matrix Biol 29, 89–96, (2010). [PubMed: 19840849]

10. Melville DB et al. The feelgood mutation in zebrafish dysregulates COPII-dependent secretion of select extracellular matrix proteins in skeletal morphogenesis. Dis Model Mech 4, 763–776, (2011). [PubMed: 21729877]

11. Levic DS et al. Animal model of Sar1b deficiency presents lipid absorption deficits similar to Anderson disease. J Mol Med (Berl) 93, 165–176, (2015). [PubMed: 25559265]

12. Sarmah S et al. Sec24D-dependent transport of extracellular matrix proteins is required for zebrafish skeletal morphogenesis. PLoS One 5, e10367, (2010). [PubMed: 20442775]

13. Lang MR, Lapierre LA, Frotscher M, Goldenring JR & Knapik EW Secretory COPII coat component Sec23a is essential for craniofacial chondrocyte maturation. Nat Genet 38, 1198–1203, (2006). [PubMed: 16980978]

14. Eames BF et al. UDP xylose synthase 1 is required for morphogenesis and histogenesis of the craniofacial skeleton. Dev Biol 341, 400–415, (2010). [PubMed: 20226781]

15. Eames BF et al. Mutations in fam20b and xylt1 Reveal That Cartilage Matrix Controls Timing of Endochondral Ossification by Inhibiting Chondrocyte Maturation. PLoS Genet 7, (2011).

16. Boyadjiev SA et al. Cranio-lenticulo-sutural dysplasia is caused by a SEC23A mutation leading to abnormal endoplasmic-reticulum-to-Golgi trafficking. Nat Genet 38, 1192–1197, (2006). [PubMed: 16980979]

17. Wakana Y et al. A new class of carriers that transport selective cargo from the trans Golgi network to the cell surface. EMBO J 31, 3976–3990, (2012). [PubMed: 22909819]

18. Malhotra V, Erlmann P & Nogueira C Procollagen export from the endoplasmic reticulum. Biochem Soc Trans 43, 104–107, (2015). [PubMed: 25619253]

19. Roden DM Phenome-wide association studies: a new method for functional genomics in humans. J Physiol 595, 4109–4115, (2017). [PubMed: 28229460]

20. Neuhauss SC et al. Mutations affecting craniofacial development in zebrafish. Development 123, 357–367 (1996). [PubMed: 9007255]

21. Pusapati GV, Luchetti G & Pfeffer SR 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, (2012). [PubMed: 23091056]

22. Siniossoglou S, Peak-Chew SY & Pelham HR Ric1p and Rgp1p form a complex that catalyses nucleotide exchange on Ypt6p. EMBO J 19, 4885–4894, doi:10.1093/emboj/19.18.4885 (2000). [PubMed: 10990452]

23. Roden DM et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther 84, 362–369, (2008). [PubMed: 18500243]

24. Ritchie MD et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am J Hum Genet 86, 560–572, (2010). [PubMed: 20362271]

25. Denny JC et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 26, 1205–1210, (2010). [PubMed: 20335276]

26. Gamazon ER et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 47, 1091–1098, (2015). [PubMed: 26258848]

27. GTEx Consortium. Genetic effects on gene expression across human tissues. Nature 550, 204–213, (2017). [PubMed: 29022597]

28. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580–585, (2013). [PubMed: 23715323]

29. Bycroft C et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209, (2018). [PubMed: 30305743]

30. Patel N et al. Novel phenotypes and loci identified through clinical genomics approaches to pediatric cataract. Hum Genet 136, 205–225, (2017). [PubMed: 27878435]

31. White RJ et al. A high-resolution mRNA expression time course of embryonic development in zebrafish. Elife 6, (2017).

32. Subramanian A & Schilling TF Thrombospondin-4 controls matrix assembly during development and repair of myotendinous junctions. Elife 3, (2014).

33. Heinemeier KM et al. Expression of collagen and related growth factors in rat tendon and skeletal muscle in response to specific contraction types. J Physiol 582, 1303–1316, (2007). [PubMed: 17540706]

34. Pingel J et al. 3-D ultrastructure and collagen composition of healthy and overloaded human tendon: evidence of tenocyte and matrix buckling. J Anat 224, 548–555, (2014). [PubMed: 24571576]

35. Schejter ED & Baylies MK Born to run: creating the muscle fiber. Curr Opin Cell Biol 22, 566–574, (2010). [PubMed: 20817426]

36. Fuiman LA & Webb PW Ontogeny of routine swimming activity and performance in zebra danios (Teleostei: Cyprinidae). Anim. Behav. 36, 250–261 (1988).

37. Budick SA & O'Malley DM Locomotor repertoire of the larval zebrafish: swimming, turning and prey capture. J Exp Biol 203, 2565–2579 (2000). [PubMed: 10934000]

38. Granato M et al. Genes controlling and mediating locomotion behavior of the zebrafish embryo and larva. Development 123, 399–413 (1996). [PubMed: 9007258]

39. Stednitz SJ et al. Forebrain Control of Behaviorally Driven Social Orienting in Zebrafish. Curr Biol 28, 2445–2451.e2443, (2018). [PubMed: 30057306]

40. Senkov O, Andjus P, Radenovic L, Soriano E & Dityatev A Neural ECM molecules in synaptic plasticity, learning, and memory. Prog Brain Res 214, 53–80, (2014). [PubMed: 25410353]

41. Aizenberg M & Schuman EM Cerebellar-dependent learning in larval zebrafish. J Neurosci 31, 8708–8712, (2011). [PubMed: 21677154]

42. Martinez O et al. The small GTP-binding protein rab6 functions in intra-Golgi transport. J Cell Biol 127, 1575–1588 (1994). [PubMed: 7798313]

43. White J et al. Rab6 Coordinates a Novel Golgi to ER Retrograde Transport Pathway in Live Cells. J Cell Biol 147, 743–760 (1999). [PubMed: 10562278]

44. Kwan KM et al. The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev Dyn 236, 3088–3099, (2007). [PubMed: 17937395]

45. Valente C, Polishchuk R & De Matteis MA Rab6 and myosin II at the cutting edge of membrane fission. Nat Cell Biol 12, 635–638, (2010). [PubMed: 20596045]

46. Del Nery E et al. Rab6A and Rab6A' GTPases play non-overlapping roles in membrane trafficking. Traffic 7, 394–407, (2006). [PubMed: 16536738]

47. Miserey-Lenkei S et al. Rab and actomyosin-dependent fission of transport vesicles at the Golgi complex. Nat Cell Biol 12, 645–654, (2010). [PubMed: 20562865]

48. Kim JH et al. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS One 6, e18556, (2011). [PubMed: 21602908]

49. Faivre L et al. Effect of Mutation Type and Location on Clinical Outcome in 1,013 Probands with Marfan Syndrome or Related Phenotypes and FBN1 Mutations: An International Study. Am J Hum Genet 81, 454–466 (2007). [PubMed: 17701892]

50. Detaint D et al. Cardiovascular manifestations in men and women carrying a FBN1 mutation. Eur Heart J 31, 2223–2229, (2010). [PubMed: 20709720]

51. Judge DP et al. Evidence for a critical contribution of haploinsufficiency in the complex pathogenesis of Marfan syndrome. J Clin Invest 114, 172–181, (2004). [PubMed: 15254584]

52. Habashi JP et al. Losartan, an AT1 antagonist, prevents aortic aneurysm in a mouse model of Marfan syndrome. Science 312, 117–121, (2006). [PubMed: 16601194]

53. Mienaltowski MJ & Birk DE Structure, physiology, and biochemistry of collagens. Adv Exp Med Biol 802, 5–29, (2014). [PubMed: 24443018]

54. Ishikawa Y & Bachinger HP A molecular ensemble in the rER for procollagen maturation. Biochim Biophys Acta 1833, 2479–2491, (2013). [PubMed: 23602968]

55. Soleman S, Filippov MA, Dityatev A & Fawcett JW Targeting the neural extracellular matrix in neurological disorders. Neuroscience 253, 194–213, (2013). [PubMed: 24012743]

56. Hatzopoulos AK Disease Models & Mechanisms in the Age of Big Data. Dis Model Mech 12, (2019).

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

58. Bastarache L et al. Phenotype risk scores identify patients with unrecognized Mendelian disease patterns. Science 359, 1233–1239, (2018). [PubMed: 29590070]

59. Khera AV et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 50, 1219–1224, (2018). [PubMed: 30104762]

## Methods-only References

60. Muller II, Knapik EW & Hatzopoulos AK Expression of the protein related to Dan and Cerberus gene--prdc--During eye, pharyngeal arch, somite, and swim bladder development in zebrafish. Dev Dyn 235, 2881–2888, (2006). [PubMed: 16921498]

61. Knapik EW et al. A microsatellite genetic linkage map for zebrafish (Danio rerio). Nat Genet 18, 338–343, (1998). [PubMed: 9537415]

62. Hoffmann K & Lindner TH easyLINKAGE-Plus--automated linkage analyses using large-scale SNP data. Bioinformatics 21, 3565–3567, (2005). [PubMed: 16014370]

63. Varshney GK et al. High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. Genome Res, (2015).

64. Li Y, Willer CJ, Ding J, Scheet P & Abecasis GR MaCH: Using Sequence and Genotype Data to Estimate Haplotypes and Unobserved Genotypes. Genet Epidemiol 34, 816–834, (2010). [PubMed: 21058334]

65. Price AL et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904–909, (2006). [PubMed: 16862161]

66. Denny JC et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol 31, 1102–1110, (2013). [PubMed: 24270849]

67. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660, (2015). [PubMed: 25954001]

68. Unlu G et al. GRIK5 Genetically Regulated Expression Associated with Eye and Vascular Phenomes: Discovery through Iteration among Biobanks, Electronic Health Records, and Zebrafish. Am J Hum Genet 104, 503–519, (2019). [PubMed: 30827500]

69. Chen L, Hsu L, Gamazon E, Cox N & Nicolae D An Exponential Combination Procedure for Set-Based Association Tests in Sequencing Studies. Am J Hum Genet 91, 977–986, (2012). [PubMed: 23159251]

70. Dale RM & Topczewski J Identification of an evolutionarily conserved regulatory element of the zebrafish col2a1a gene. Dev Biol 357, 518–531, (2011). [PubMed: 21723274]

71. Kawakami K & Shima A Identification of the Tol2 transposase of the medaka fish Oryzias latipes that catalyzes excision of a nonautonomous Tol2 element in zebrafish Danio rerio. Gene 240, 239–244, (1999). [PubMed: 10564832]

72. Cox NJ et al. Dynamic Glycosylation Governs the Vertebrate COPII Protein Trafficking Pathway. Biochemistry 57, 91–107, (2018). [PubMed: 29161034]

73. Niu X et al. The nuclear pore complex function of Sec13 is required for cell survival during retinal development. J Biol Chem, (2014).

74. Walker MB & Kimmel CB A two-color acid-free cartilage and bone stain for zebrafish larvae. Biotech Histochem 82, 23–28, (2007). [PubMed: 17510811]

75. Zou P et al. A conserved role of alphaA-crystallin in the development of the zebrafish embryonic lens. Exp Eye Res 138, 104–113, (2015). [PubMed: 26149094]

76. Gorur A et al. COPII-coated membranes function as transport carriers of intracellular procollagen I. J Cell Biol 216, 1745–1759, (2017). [PubMed: 28428367]

77. Kim D et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36, (2013). [PubMed: 23618408]

78. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550, (2014). [PubMed: 25516281]

79. Huber W et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12, 115–121, (2015). [PubMed: 25633503]
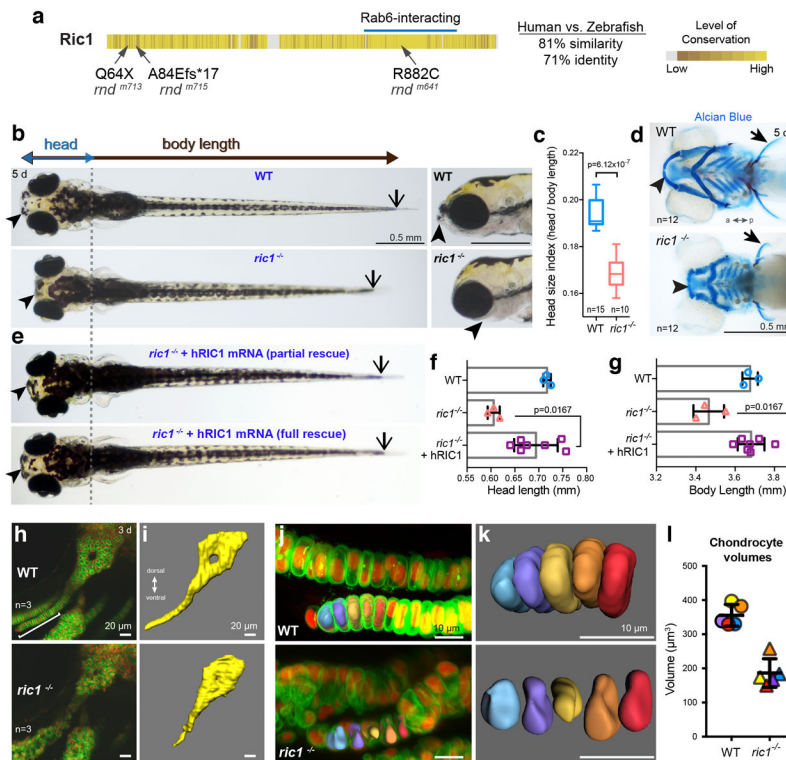
**Figure 1. Ric1 is required for craniofacial skeleton development and shape.**

**a,** Zebrafish Ric1 protein is highly conserved with 81% similarity to human RIC1 (Clustal Omega, EMBL-EBI). Positional cloning identified mutations in *round* alleles (*rnd*[m641], *rnd*[m713], *rnd*[m715]; see arrows), Rab6 interacting region (blue bar, also Extended Data Fig. 1). **b,** Live images of *ric1*[−/−] mutant larvae show micrognathia (arrowhead), small head and shortened trunk (dashed line to arrow). **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). Boxes show 25th to 75th percentiles and medians. Whiskers are min to max values. **d,** Alcian blue and Alizarin red staining of the *ric1*[−/−] mutant shows shorter, malformed Meckel's cartilage (arrowheads) and shorter, kinked pectoral fins (arrows). All three alleles exhibit indistinguishable phenotypes and fail to complement one another in genetic complementation assays. **e,** Live images of *ric1*[−/−] mutant embryos (5 dpf, days post-fertilization) overexpressing human *RIC1* mRNA, show rescue of jaw protrusion (arrowheads) and elongation of the body length (arrows). Quantification of the rescue experiments, head in **f,** and body length in **g**. Statistical comparison by two-tailed Mann-Whitney U-test, CI = 95%, n=3 independent animals for *ric1*[−/−] and WT; n=7 animals for hRIC1 rescue group. **h,** Hyosymplectic cartilage (HC) dysmorphology shown in double transgenic Tg(Col2a1a:caax-EGFP; Col2a1a:H2A-mCherry) zebrafish that marks plasma membrane (green) and nucleus (red), white bar marks the symplectic arm (SA); maximum intensity projections of live z-stack images by confocal microscopy in WT and *ric1*[−/−] embryos (n=3 independent animals per group). **i,** 3D rendered structures of HC shapes highlight the malformed and short cartilage shape in *ric1*[−/−]. **j,** Magnified views of SA chondrocytes. 3D rendered volumes are overlaid onto maximum intensity projection views.

**k,** 3D rendering underscores the dysmorphic shape and reduced volume of *ric1*$^{-/-}$ chondrocytes. **l,** Volume measurements of 5 cells per genotype by *Imaris* 8, color-coded to match the cells in **k**. Lines indicate mean and standard deviations in **f, g** and **l**.
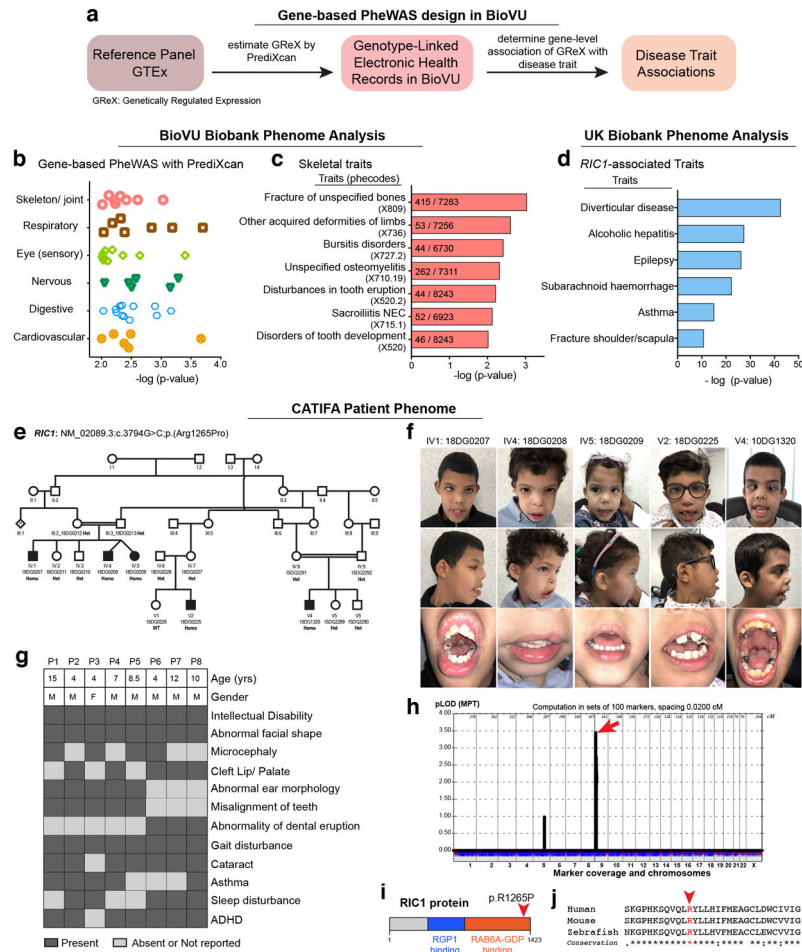
**Figure 2. Human common-disease phenome significantly associated with genetically determined reduced *RIC1* expression in BioVU biobank and CATIFA patient phenome.**

**a,** Design of gene-based PheWAS study conducted in BioVU biobank. **b,** 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 and sample numbers are in the Supplementary Table 2. Traits are categorized into systems (y-axis), and significance is displayed on x-axis. Significance is tested by logistic regression analysis (two-sided). Multiple testing adjustment is done using Bonferroni correction. **c,** Individual skeletal phecodes from panel **b** are listed on y-axis. 'Number of cases / total cases analyzed' is indicated as insets within bars. x-axis shows significance levels (logistic regression, two-sided, Bonferroni-corrected). **d,** SNPs in the *RIC1* gene are highly significantly associated, in the UK Biobank, with *RIC1* phenome observed in the BioVU biobank. Representative traits and the p-value for the most significant SNP in *RIC1* plotted; complete list of SNP-level and gene-level associations can be found in Supplementary Table 4. **e,** Pedigree of a large multiplex, consanguineous family with affected individuals shaded. Standard pedigree symbols used, genotypes are listed for tested individuals. Het: Heterozygous, Homo: Homozygous. **f,** Photographs of patients with CATIFA syndrome. Written consent for the use of photographs was obtained from the parents of affected individuals. **g,** Human Phenotype Ontology heat map of patients' common clinical features.

Table of complete clinical description for each subject is presented in Supplementary Table 5. **h,** Linkage analysis and LOD score value of CATIFA syndrome-linked RIC1 c3794G>C variant on chromosome 9 (arrow). **i,** Schematic of RIC1 protein marking the R1265 variant site within the Rab6 binding region. **j,** Translated RIC1 protein sequence is highly conserved among vertebrates; human, mouse and zebrafish, including R1265 residue (arrowhead).

**a,** Experimental design for immunohistochemistry (IHC) analysis of chondrocytes in the head. **b,** Representative images of Tg(Col2a1a:caax-EGFP) transgenic zebrafish stained with Col2 antibody (14 μm-thick cryosections) in WT and *ric1*[−/−] cartilage, (Collagen 2, magenta; EGFP, green). In WT Col2 signal is in extracellular space, outside caax-EGFP boundaries (dashed line), whereas in *ric1*[−/−] collagen accumulates in chondrocytes (inside caax-EGFP boundaries, arrows). **c, d,** Quantification of the percentage of cytoplasmic area occupied by Col2 signal in chondrocytes (**c**), and notochord sheath cells (**d**). For chondrocytes: N=3 embryos per group, n=30 cells per group. 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, two-tailed, CI= 95%. Mean and SD values are indicated with bars. **e,** Gene expression levels by RNA-seq from live, FACS-sorted chondrocytes, columns

**Figure 3. Ric1 modulates procollagen transport.**

indicate mean values, bars show SD. RPKM: reads per kilobase of transcript per million mapped reads. 2 independent samples analyzed per genotype (WT and $ric1^{-/-}$); 100 animals were included in each sample. **f,** TEM images of 4 dpf $ric1^{-/-}$ craniofacial chondrocytes show large vacuolar structures (arrowhead), **f'** higher magnification of boxed area, and **f''** further zoom in of the blue box showing striated ultrastructure of intracellular collagen fibrils. Control WT images in Extended Data Fig. 6. **g,** 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. TEM imaging was repeated with three independent $ric1^{-/-}$ animals at 3 dpf, one at 4 dpf, one WT sibling at 3 dpf and one at 4 dpf. 5 cells were imaged per animal. ECM: extracellular matrix, N: nucleus.
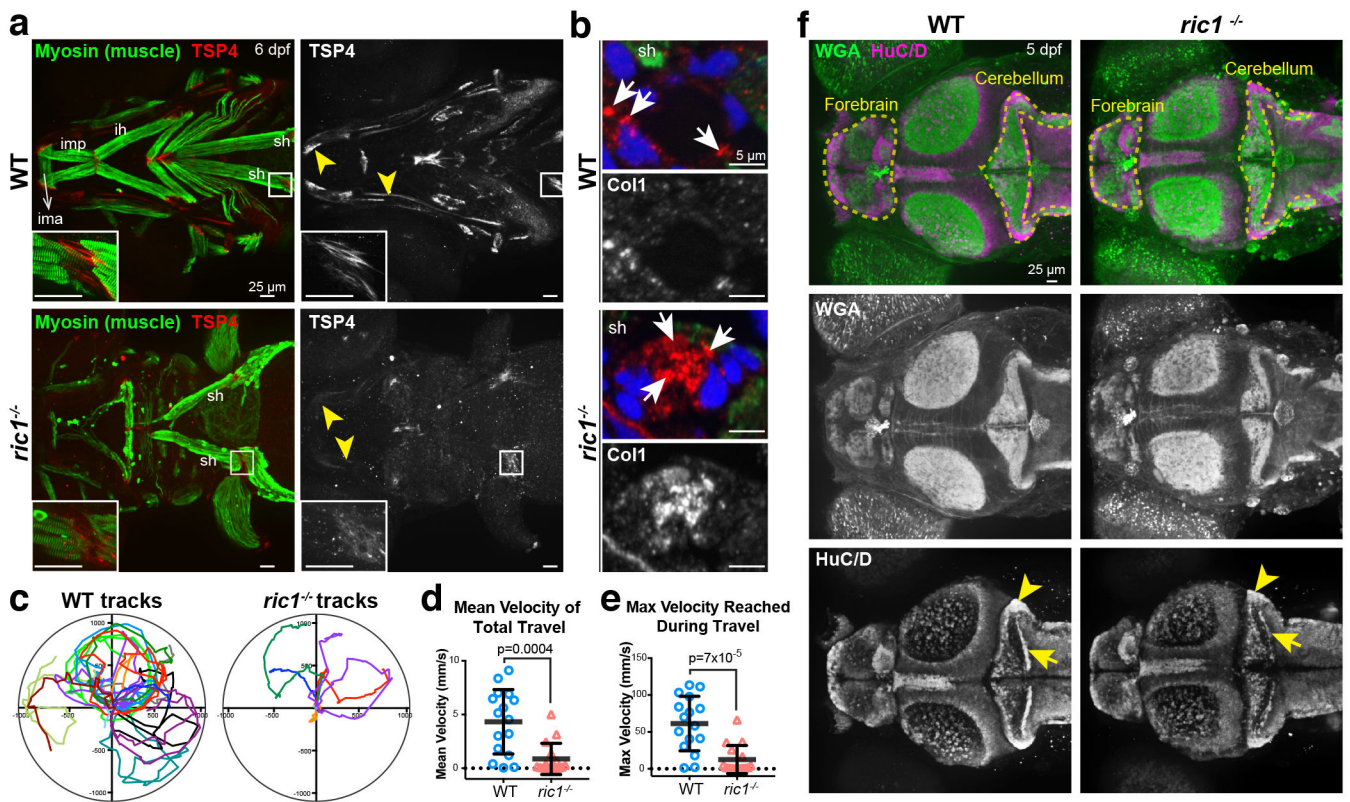
**Figure 4. Ric1 depletion leads to musculoskeletal, brain and locomotion defects.**
**a,** Whole-mount immunostaining of WT and *ric1*−/− mutant larvae for myosin (MF20, muscle) and thrombospondin-4 (TSP4) for tendon area. Ventral view of head muscles, inset: magnified view of boxed tendon region; arrowheads point to TSP4+ domains in the Meckel's (left) and palatoquadrate (right) elements. n=4 animals for each group. Abbreviations for muscles, ih: interhyoideus, imp: intermandibularis posterior, ima: intermandibularis anterior, sh: sternohyoideus. **b,** Cryo-section and immunostaining of tendon area along sh muscle for collagen-I (Col1, red), phalloidin (muscle, green) and DAPI nuclear counterstain (blue). Tenocyte nuclei are marked by DAPI, arrows to the Col1 punctae. n=3 animals for each group **c,** Locomotion tracking of zebrafish embryos at 4 dpf stage. 16 WT and 16 *ric1*−/− embryos were tracked and plotted. Each color line indicates a separate animal (individual tracks in Extended Data Fig. 7). **d, e,** Graphs showing mean velocity of total travel (**d**) and maximum velocity reached during recorded travel (**e**) for each embryo. Mean and SD bars are displayed. Two-tailed Mann-Whitney U-test, CI = 95%, 16 WT and 16 *ric1*−/− larvae were analyzed. **f,** Whole mounted zebrafish embryos immunostained for HuC/D (pan-neuronal marker) and WGA (wheat germ agglutinin) at 5 dpf, maximum intensity projection. Forebrain and cerebellum structures are demarcated by dashed lines. Quantification is in Extended Data Fig. 8. 8 independent larvae per each group were analyzed. Arrowheads point to eminentia granularis; arrows to lobus caudalis cerebelli.
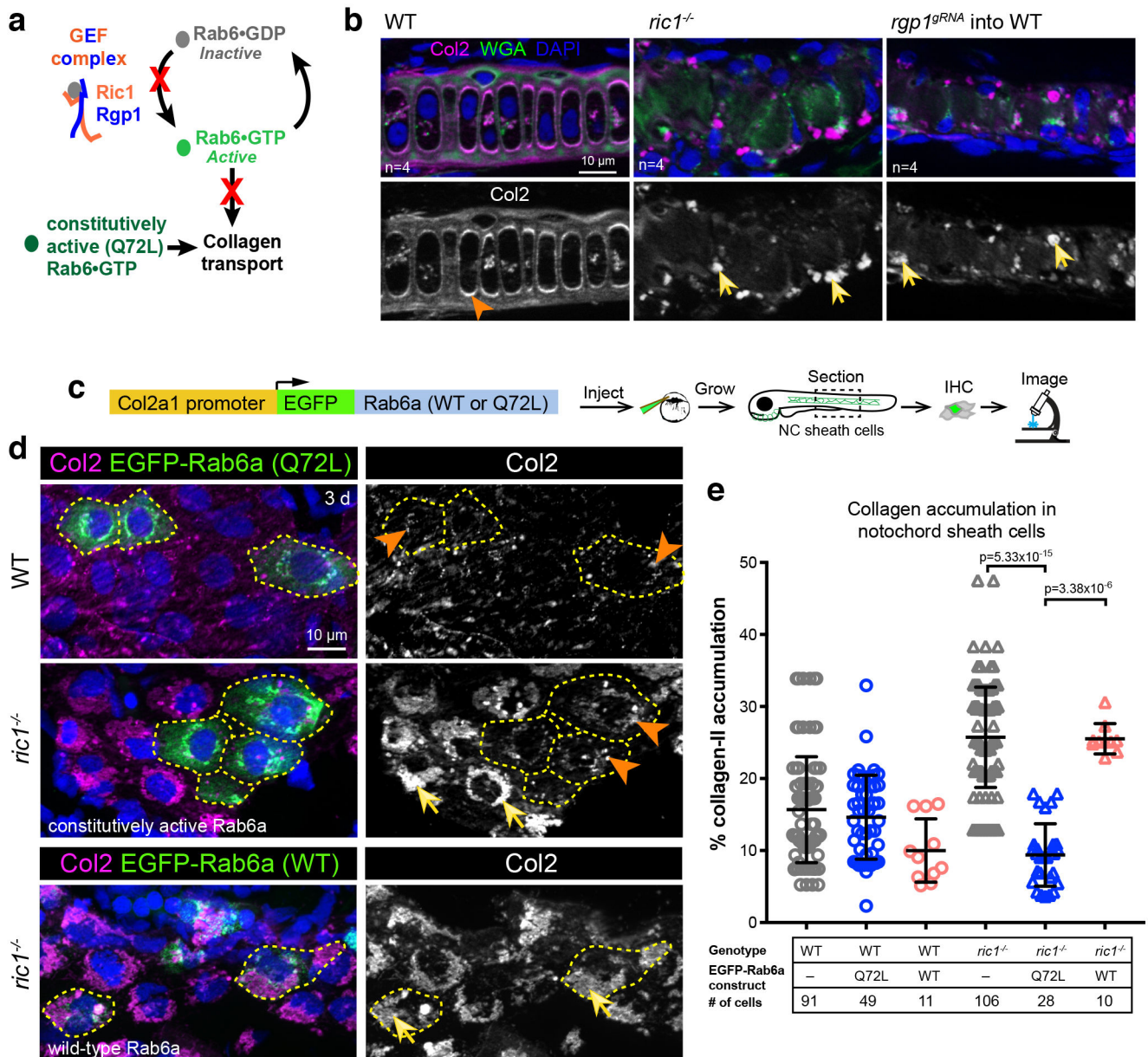
**Figure 5. Ric1-Rgp1 GEF complex regulates procollagen transport via Rab6a activation.**
**a,** Model of Rab6a activation by the Ric1-Rgp1 GEF complex to regulate collagen transport, and constitutively active Rab6a bypassing GEF requirement for collagen transport. **b,** IF labeling with Col2 antibody and wheat germ agglutinin (WGA) in 3 dpf zebrafish cartilage shows disrupted cell shape and tissue organization in *ric1*-mutants and *rgp1*[gRNA] (rgp1-guide RNA for CRISPR-Cas9 genome editing, Extended Data Fig. 9) injected WT embryos. Yellow arrows point to intracellular collagen accumulation, orange arrowhead points to extracellular Col2. This experiment was repeated with similar results, 3 independent animals per each group. **c,** Experimental design for mosaic overexpression (OE) 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). Yellow arrows point to intracellular accumulations; orange 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. OE of constitutively active Rab6a (Q72L) rescues intracellular collagen accumulation to WT levels (CI=95%, two-tailed Mann-Whitney U-test); whereas WT Rab6a does not. Mean and SD values are indicated with bars.
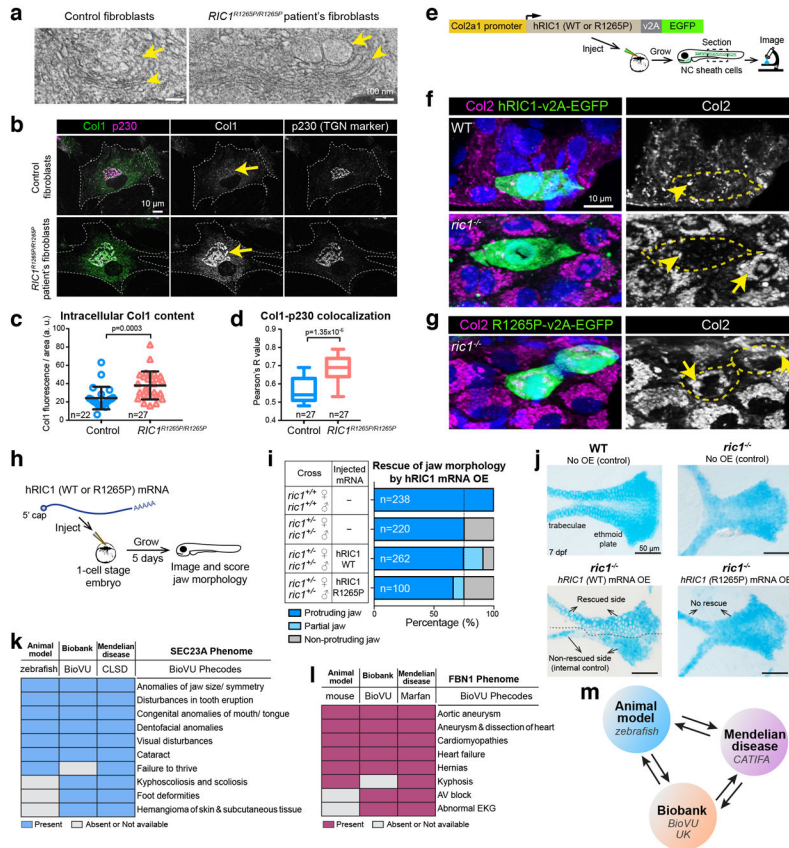
**Figure 6. Pathogenic *RIC1* variant leads to collagen accumulation in CATIFA fibroblasts.**
**a,** TEM images of control BJ fibroblasts and *RIC1^R1265P/R1265P^* patient's dermal fibroblasts show Golgi complex stacks (arrowheads). Arrow points to enlarged Golgi lumen and post-Golgi structures in patient fibroblasts. 5 cells of each group were imaged by TEM. **b,** Co-localization of Collagen 1 (Col1) and the TGN marker p230 (arrow) by IF in *RIC1^R1265P/R1265P^* cells. **c, d,** Quantification of the intracellular Col1 content (**c**), and Col1 co-localization with p230 (**d**) based on images in **b**. (Student's t-test, two-tailed, CI= 95%). Mean value (center line) and SD bars are indicated in 'c'; center line indicates the median and whiskers show min to max values in 'd'. **e,** Experimental design for mosaic OE of 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 (arrowheads). **g,** Mosaic OE of CATIFA-linked R1265P variant of hRIC1 only partially clears collagen deposits (arrows) while neighboring cells continue to retain Col2 and serve as *ric1^−/−^* controls. Similar observations were made in 3 cells for each group in f, g. **h,** Experimental design for hRIC1 mRNA OE and analysis of craniofacial morphology rescue. **i,** Phenotypic scoring results of protruding jaw morphology in hRIC1 mRNA overexpressing larvae and controls. n: number of animals analyzed. **j,** Dissected Alcian blue preparations of ethmoid plate and trabeculae (dorsal head skeleton) of hRIC1 mRNA injected larvae and controls show mosaic rescue of *ric1^−/−^* cells (arrows) with characteristic clear cytoplasm as in the WT skeletal element. The R1265P

variant rescue more closely resembles cell morphology in the *round* mutants; 3 independent animals of WT and *ric1*−/−. **k,** Phenome comparison table for *SEC23A* gene, associated with cranio-lenticulo-sutural dysplasia (CLSD) in animal models, gene-based PheWAS analysis in BioVU and Mendelian disease patients. **l,** Phenome comparison table for *FBN1* (fibrillin-1) gene, associated with Marfan syndrome. AV: atrioventricular, EKG: electrocardiogram. **m,** Integrated approach for discovery of gene function and disease mechanism using the animal model, common-disease phenome in a biobank and monogenic, rare Mendelian disease phenome of *RIC1*/CATIFA syndrome.