Identification of Fibronectin 1 as a candidate genetic modifier in a *Col4a1* mutant mouse model of Gould syndrome

Mao Mao[1], Tanav Popli[1§], Marion Jeanne[1♯], Kendall Hoff[1+], Saunak Sen[2,3,4], and Douglas B. Gould[1,3,5*]

[1]Departments of Ophthalmology, University of California, San Francisco, San Francisco, CA. 94143, USA; [2]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA. 94143, USA; [3]Institute of Human Genetics, University of California, San Francisco, San Francisco, CA. 94143, USA; [4]Department of Preventive Medicine, University of Tennessee Health Science Center, 66 North Pauline St, Memphis, TN 38163 USA; [5]Department of Anatomy, University of California, San Francisco, San Francisco, CA 94143, USA

*Corresponding author:
Dr. Douglas B. Gould, Departments of Ophthalmology and Anatomy, Institute for Human Genetics, UCSF School of Medicine, 10 Koret Way, Room 235, San Francisco, CA, 94143-0730, (415) 476-3592 (telephone), Douglas.Gould@ucsf.edu, ORCID identifier number 0000-0001-5127-5328
[§]Current address: University of Michigan, Department of Neurology, 1500 E Medical Center Drive, Taubman Center, Clinic 1C, Ann Arbor, MI, 48109
[♯]Current address: Genentech Inc, 1 DNA Way, South San Francisco, CA 94080
[+]Current address: Centrillion Technologies, 2500 Faber Pl. Palo Alto, CA 94303

Key words:
Gould syndrome, COL4A1, basement membrane, modifier, fibronectin, integrin signaling

SUMMARY STATEMENT

A genetic screen in mice implicates FN1 as a genetic modifier of some, but not all, aspects of Gould syndrome.

ABSTRACT

Collagen type IV alpha 1 and alpha 2 (COL4A1 and COL4A2) are major components of almost all basement membranes. *COL4A1* and *COL4A2* mutations cause a multisystem disorder called Gould syndrome which can affect any organ but typically involves the cerebral vasculature, eyes, kidneys and skeletal muscles. The manifestations of Gould syndrome are highly variable and animal studies suggest that allelic heterogeneity and genetic context contribute to the clinical variability. We previously characterized a mouse model of Gould syndrome caused by a *Col4a1* mutation in which the severities of ocular anterior segment dysgenesis (ASD), myopathy, and intracerebral hemorrhage (ICH) were dependent on genetic background. Here, we performed a genetic modifier screen to provide insight into the mechanisms contributing to Gould syndrome pathogenesis and identified a single locus (*modifier of Gould syndrome 1*; *MoGS1*) on Chromosome 1 that suppressed ASD. A separate screen showed that the same locus ameliorated myopathy. Interestingly, *MoGS1* had no effect on ICH, suggesting that this phenotype may be mechanistically distinct. We refined the *MoGS1* locus to a 4.3 Mb interval containing 18 protein coding genes, including *Fn1* which encodes the extracellular matrix component fibronectin 1. Molecular analysis showed that the *MoGS1* locus increased *Fn1* expression raising the possibility that suppression is achieved through a compensatory extracellular mechanism. Furthermore, we show evidence of increased integrin linked kinase levels and focal adhesion kinase phosphorylation in *Col4a1* mutant mice that is partially restored by the *MoGS1* locus implicating the involvement of integrin signaling. Taken together, our results suggest that tissue-specific mechanistic heterogeneity contributes to the variable expressivity of Gould syndrome and that perturbations in integrin signaling may play a role in ocular and muscular manifestations.

INTRODUCTION

Collagens are the most abundant proteins in the body, making up ~30% of the dry weight. The collagen superfamily of extracellular matrix (ECM) molecules comprises 28 members encoded by 46 genes (Ricard-Blum, 2011). Among these, type IV collagens are primordial ECM molecules and are fundamental constituents of specialized structures called basement membranes (Fidler et al., 2017). In mammals, six genes encode the type IV collagens (*COL4A1* to *COL4A6*) and their protein products assemble into three distinct heterotrimers ($\alpha1\alpha1\alpha2$, $\alpha3\alpha4\alpha5$ or $\alpha5\alpha5\alpha6$). The $\alpha1\alpha1\alpha2$ network is ubiquitous throughout development and in most adult tissues and its absence results in embryonic lethality (Poschl et al., 2004). Pathogenic mammalian *Col4a1* mutations were first identified using forward mutagenesis screens in mice with variable forms of ocular pathology (Favor et al., 2007; Gould et al., 2005; Thaung et al., 2002) and human mutations were identified in individuals with severe inherited or *de novo* porencephaly and early onset intracerebral hemorrhages (ICH) (Breedveld et al., 2006; Gould et al., 2005; Gould et al., 2006; Sibon et al., 2007; Vahedi et al., 2007). A number of studies in humans and mice have subsequently expanded the phenotypic spectrum and it is now well established that mutations in *COL4A1* and *COL4A2* cause a multi-system disorder (Jeanne and Gould, 2017; Labelle-Dumais et al., 2019; Mao et al., 2015; Meuwissen et al., 2015; Yoneda et al., 2013; Zagaglia et al., 2018) that has adopted the name Gould syndrome (https://www.gouldsyndromefoundation.org).

Gould syndrome is highly clinically heterogeneous and can have variable penetrance and severity across many organs. Cerebrovascular, ocular, renal and neuromuscular pathologies are among the most commonly described manifestations. We originally described a mouse model of Gould syndrome with a *Col4a1* splice acceptor mutation that leads to exclusion of exon 41 (Δex41) from the mature transcript and 17 amino acids from the triple-helical domain of the protein (Gould et al., 2005). However, the majority of disease-causing variants reported in humans are missense mutations in highly conserved glycine residues of the triple-helical domain (Jeanne and Gould, 2017). In order to faithfully replicate human disease, we compared the effects of different mutations in an allelic series composed of nine distinct *Col4a1* and *Col4a2* mutant mouse strains (Favor et al., 2007; Jeanne et al., 2015; Kuo et al., 2014). We established that allelic heterogeneity has important implications for penetrance and severity of various pathologies including ICH and myopathy. We found that mutations can differ in degree or kind and that allelic differences contribute to clinical variability, in part, by tissue-specific mechanistic heterogeneity (Labelle-Dumais et al., 2019). However, variable clinical manifestations including age of onset and severity among organs have been reported in people with the same recurrent or inherited mutations, and reduced penetrance and asymptomatic carriers exist in some families,

indicating the involvement of factors in addition to allelic heterogeneity (Coupry et al., 2010; de Vries et al., 2009; Rødahl et al., 2013; Shah et al., 2012).

Studies using inbred strains of mice clearly demonstrate that a mutation can have variable outcomes in different genetic contexts underscoring the important role that genetic modification plays in clinical heterogeneity. When compared to other mutations in the allelic series, *Col4a1*$^{\Delta ex41}$ tends to have the most severe phenotypes across multiple organs when maintained on a pure C57BL/6J (B6) genetic background (Gould et al., 2007; Jeanne et al., 2015; Kuo et al., 2014). However, when the *Col4a1*$^{+/\Delta ex41}$ B6 mice were crossed for a single generation to the CAST/EiJ (CAST) genetic background, ocular anterior segment dysgenesis (ASD), ICH and skeletal myopathy were all significantly reduced in the F1 progeny (CASTB6F1) (Gould et al., 2007; Jeanne et al., 2015; Labelle-Dumais et al., 2011). Interestingly, when *Col4a1*$^{+/\Delta ex41}$ B6 mice were crossed for a single generation to the 129SvEvTac (129) genetic background, ASD was ameliorated (albeit to a lesser extent than in CASTB6F1 mice) but ICH was not (Gould et al., 2007; Jeanne et al., 2015). Collectively, these data suggest that the B6 genetic background confers susceptibility to develop severe Gould syndrome related pathologies and that CAST and 129 strains have one or more locus/loci that can genetically suppress pathology in a tissue-specific manner.

Identification of genetic modifier genes can help reveal disease mechanisms and potential therapeutic targets. Phenotype-driven mapping studies using mice and other model organisms are a powerful approach to identify genetic interactions that may be difficult to discover in humans even when large pedigrees are available (Ceco and McNally, 2013; Meyer and Anderson, 2017; Vieira et al., 2015). We previously performed a genetic modifier screen for pathology caused by the *Col4a1*$^{\Delta ex41}$ mutation to provide insight into the underlying pathogenic mechanisms. Because 1) ASD is relatively severe, easily screened, and robustly suppressed by two strains and 2) the magnitude and breadth of phenotypic rescue was greater in CAST compared to 129 mice, we selected ASD and CAST as the phenotype and strain, respectively, to perform a pilot suppressor screen in a small number of mice (Gould et al., 2007). We generated CASTB6F1 mice and iteratively crossed mutant mice back to the B6 background. In each backcross, we applied selective pressure to retain one or more genetic suppresser locus by choosing the progeny with the mildest phenotype to breed for the next generation. Using a crude genome-wide scan, we identified a dominant modifier locus in CAST Chromosome (Chr) 1 that suppressed ASD. However, the approach had significant limitations including screening on a single phenotype, the inability to identify potential recessive loci and the possibility that we may have incidentally overlooked other dominant loci by imposing bottlenecks at each generation.

Here, we performed independent genetic modifier screens for ASD and skeletal myopathy on a large scale F2 cross with the potential to identify multiple dominant or recessive loci that enhanced or suppressed pathology. Both screens identified a single suppressor locus on CAST Chr 1 that we called *modifier of Gould syndrome 1* (*MoGS1*). Surprisingly, although the CAST background also suppressed ICH, the effect was not attributable to the *MoGS1* locus. These data suggest that ASD and skeletal myopathy may be mechanistically related to each other but distinct from ICH which further supports the notion of tissue-specific mechanistic heterogeneity contributing to the clinical variability of Gould syndrome (Labelle-Dumais et al., 2019). Furthermore, molecular analyses suggest that the ECM protein fibronectin 1 (FN1) is a strong candidate as the genetic suppressor at the *MoGS1* locus and provide evidence for a role of altered integrin signaling in Gould syndrome pathogenesis.

RESULTS

High-resolution genome-wide mapping identified a suppressor locus for *Col4a1*-related ASD and myopathy on mouse Chr 1. To identify modifiers of Gould syndrome, we first performed a genome-wide screen with the early onset and easily observed ASD phenotype. We phenotyped and genotyped 192 mutant F_2 progeny from (CAST X B6-*Col4a1*$^{+/\Delta ex41}$) F_1 intercrosses using a mouse medium density linkage panel with 646 informative SNPs (Fig. 1A-C). We identified a region of interest on CAST Chr 1 with a LOD score of 11.2 and Bayesian confidence interval extending from 51.3 to 73.0 Mb (Fig.1C). To determine if smaller effects of other loci might have been masked, we performed a second genome-wide scan conditioned on this locus, but no additional loci reached statistical significance (Fig. 1D). This observation suggests that the major modifying effect of the CAST background on ASD is imparted through this single dominant locus on Chr 1.

In addition to suppressing ASD, the CAST genetic background significantly suppresses skeletal myopathy in *Col4a1* mutant mice (Labelle-Dumais et al., 2011). To identify potential modifier loci for this phenotype, we performed an independent genome-wide mapping analysis using skeletal myopathy as the primary phenotype. We quantified myopathy severity as the percentage of muscle fibers containing non-peripheral nuclei (NPN) in histological sections of quadriceps (Fig. 1E). This assay is more quantitative with greater statistical power but also more labor-intensive and therefore we analyzed a randomly generated subset (n= 49) of the mutant (CAST X B6-*Col4a1*$^{+/\Delta ex41}$) F_2 progeny. The genome-wide scan revealed an interval on Chr 1 defined by the same markers as the ASD modifier locus (max LOD score = 5.61) (Fig. 1F).

To validate the biological effect of the *MoGS1* locus, we independently generated an incipient congenic strain by iterative backcrossing the CAST-derived locus onto the susceptible B6 genetic background for 5 generations (N5). At N5, the genetic background of the mice is ~97% pure B6 on average, except that they carry the CAST Chr 1 interval that includes the *MoGS1* locus. Incipient congenic mice were intercrossed to generate *Col4a1*$^{+/\Delta ex41}$ mice with zero (B/B), one (C/B) or two (C/C) copies of the CAST-derived chromosomal interval (Fig. 2). Out of 24 eyes from *Col4a1*$^{+/\Delta ex41}$ mice that were homozygous for the B6 allele (*MoGS1*$^{B/B}$) at the congenic interval, 29% (7) were moderate, and 71% (17) were severe (Fig. 2A). In contrast, out of 28 eyes from the heterozygous group (*MoGS1*$^{C/B}$) there were 21.5% (6), 46.5% (13) and 32% (9) of eyes that were scored with mild, moderate and severe ASD, respectively. In mice homozygous for the CAST allele (*MoGS1*$^{C/C}$), there were 50% (10), 35% (7) and 15% (3) of eyes that were mild, moderate and severe, respectively. In a parallel experiment, we tested the effect of this congenic locus on myopathy and observed a dosage effect, whereby *Col4a1*$^{+/\Delta ex41}$ mice that were heterozygous at the *MoGS1* locus (*MoGS1*$^{C/B}$) showed a trend toward reduced myopathy while homozygous *MoGS1*$^{C/C}$ incipient congenic mice had significantly milder myopathy compared to mice that were homozygous for the B6 allele (*MoGS1*$^{B/B}$). Together, these results support the existence of one or more semi-dominant modifier genes at the *MoGS1* locus on Chr 1.

MoGS1 does not reduce porencephaly penetrance or ICH severity in _Col4a1_ mutant mice

Individuals with Gould syndrome have highly penetrant and clinically variable cerebrovascular diseases that include porencephaly and ICH (Bilguvar et al., 2009; de Vries et al., 2009; Giorgio et al., 2015; Vilain et al., 2002). ICH was previously reported to be significantly reduced in *Col4a1* mutant mice on a CASTB6F1 background compared to those maintained on a B6 background suggesting that the CAST genetic background also suppresses this phenotype (Jeanne et al., 2015). To test whether the *MoGS1* locus might also genetically modify cerebrovascular diseases, we assessed porencephaly penetrance and ICH severity of the incipient congenic mice. We found that porencephaly penetrance and ICH severity were similar among *Col4a1*$^{+/\Delta ex41}$ mice irrespective of their genotypes at the *MoGS1* locus (Fig. 2C, D).

Fine mapping with subcongenic lines refined _MoGS1_ to a 4.3 Mb interval

To refine the interval of interest at the *MoGS1* locus, we generated subcongenic lines of the CAST derived chromosomal fragments on a B6 background (N5) and tested which line suppresses ASD and skeletal myopathy in *Col4a1*$^{+/\Delta ex41}$ mice (Fig. 3A). Line 1 which included an interval of ~10 Mb from the CAST genome did not have a modifying effect on either phenotype (Fig. S1). In contrast, a subcongenic line (line 2) containing the distal portion of the *MoGS1* locus (from 68.7 – 73.0 Mb) showed a significant

protective effect for both ASD (Fig. 3B) and myopathy (Fig. 3C). While approximately 87% of eyes from *Col4a1*$^{+/\Delta ex41}$;*MoGS1*$^{B/B}$ mice had severe ASD (1 mild, 4 moderate and 33 severe), only ~35% of the eyes from *Col4a1*$^{+/\Delta ex41}$;*MoGS1*$^{C/C}$ mice were severe (8 mild, 14 moderate, and 12 severe). Likewise, *Col4a1*$^{+/\Delta ex41}$;*MoGS1*$^{C/C}$ mice had significantly fewer NPN compared to *Col4a1*$^{+/\Delta ex41}$;*MoGS1*$^{B/B}$ mice. Heterozygosity for this distal interval had intermediate effects for both phenotypes.

Candidate gene analysis

The refined *MoGS1* locus is 4.3 Mb, and contains 18 protein-coding genes, and 38 predicted non-protein coding genes including 13 lncRNAs, 6 snRNA genes, pseudogenes, and others (Table S1). Using publicly available databases (Keane et al., 2011), we examined the refined *MoGS1* interval for sequence differences between the B6 and CAST genomes. Among all variations, there were 7,430 single nucleotide polymorphisms (SNPs), 667 small insertions and deletions (indels) and 17 large structural variations including large insertions or deletions. Thirty-five SNPs were predicted to be missense variants affecting 10 protein-coding genes (Table 1). Among those genes, *Ankar* had three missense variants predicted to be deleterious by SIFT (Vaser et al., 2016) and *Spag16* had one predicted stop-gain variant; however, neither gene is an obvious functional candidate. Based on the Gene Expression Database (GXD) from Mouse Genome Informatics (informatics.jax.org/expression), 13 of the 18 protein coding genes have been shown to be expressed both in eyes and skeletal muscles. Two genes, *Vwc2l* and *Bard1*, are only expressed in eye and three genes, *Ankar*, *Spag16*, and *Abca12*, have no reported expression in either tissue. Since both ocular and muscular defects in *Col4a1*$^{+/\Delta ex41}$ mice were suppressed by the *MoGS1* locus, we hypothesized that the underlying pathogenic mechanism(s) may be shared between these two tissues. Therefore, despite the presence of possible pathogenic variants, we excluded *Ankar, Spag16*, *Abca12, Vwc2l and Bard1* from further analysis.

Next, we tested whether any of the 13 genes with reported expression in ocular and muscular tissues are differentially expressed in B6 mice with or without the CAST interval at the *MoGS1* locus. Four genes, *Mreg*, *Fn1*, *Pecr1*, and *Igbp2*, showed significantly increased expression in *MoGS1*$^{C/C}$ eyes at postnatal day 0 (P0) compared to *MoGS1*$^{B/B}$ eyes by qPCR (Table 1 and Fig. 4A). FN1 is an extracellular matrix protein with multiple important roles in development and tissue homeostasis by interacting with cell surface receptors, extracellular matrix proteins and growth factors (Zollinger and Smith, 2017). Notably, FN1 has binding sites for type IV collagen (Laurie et al., 1986; Laurie et al., 1982) and localizes adjacent to basement membranes in many tissues (Laurie et al., 1982). Moreover, like collagens, FN1 interacts with cells via integrin receptors (Johansson et al., 1997; Vandenberg et al., 1991), making it a strong functional candidate as a genetic modifier of COL4A1-related pathology.

Functional testing of FN1 and integrin signaling in MoGS1 mice

In silico analysis comparing the CAST and B6 alleles of *Fn1* revealed 3 missense SNPs, 1 splice region SNP, 7 SNPs in the UTR, 238 other noncoding SNPs, and 2 structural variations (Table S2). Murine FN1 has 12 isoforms (White et al., 2008), making it difficult to predict the functional consequence(s) of a particular variant. The primary consequence of *Col4a1* mutations is impaired secretion of mutant $\alpha1\alpha1\alpha2$ heterotrimers into the basement membranes (Kuo et al., 2014) and it is possible that increased FN1 confers partial compensation for extracellular $\alpha1\alpha1\alpha2$ deficiency. Consistent with this hypothesis, we observed increased *Fn1* expression in developing eyes from $MoGS1^{C/C}$ mice compared to $MoGS1^{B/B}$ mice (Fig. 4A). Furthermore, we found that FN1 levels were higher in P10 quadriceps from $Col4a1^{+/\Delta ex41}$ compared to $Col4a1^{+/+}$ mice and in $MoGS1^{C/C}$ compared to $MoGS1^{B/B}$ mice (Fig. 4B, C). FN1 plays important roles in cell adhesion, migration and signaling during tissue morphogenesis, which is primarily mediated through its interaction with integrins (Miyamoto et al., 1998; Sakai et al., 2003). Therefore, we tested the effect of *MoGS1* on integrin signaling by evaluating the protein levels and phosphorylation status of two downstream effectors, integrin linked kinase (ILK) and focal adhesion kinase (FAK) (Hu and Luo, 2013; Humphries et al., 2019) (Fig. 4B, C). Western blot analysis of P10 quadriceps revealed that ILK levels were higher in $Col4a1^{+/\Delta ex41}$ mice irrespective of whether they carried the $MoGS1^{B/B}$ or $MoGS1^{C/C}$ interval. Similarly, FAK phosphorylation (pFAK) and pFAK/FAK ratio were significantly higher in $Col4a1^{+/\Delta ex41}$;$MoGS1^{B/B}$ compared to $Col4a1^{+/+}$;$MoGS1^{B/B}$ mice, and were reduced by the $MoGS1^{C/C}$ interval. No difference was detected for the broadly used beta subunit of the integrin receptor, ITGB1, between genotypes. Taken together, these data suggest that elevated integrin signaling may contribute to Gould syndrome and that ASD and myopathy may be partially rescued by compensatory *Fn1* expression.

DISCUSSION

Here, we performed a genetic modifier screen in *Col4a1* mutant mice to gain insight into the pathogenic mechanisms that contribute to Gould syndrome. To this end, we used a large-scale F2 cross of CASTxB6 and conducted independent screens for ASD and skeletal myopathy – two phenotypes that are commonly observed in individuals with Gould syndrome. Independent analyses for both phenotypes revealed a single, shared, semi-dominant locus on CAST Chr 1 that corroborates a previously identified locus (Gould et al., 2007). To validate this genetic data, we iteratively backcrossed the locus onto the B6 background and showed that a refined interval, referred to as *MoGS1*, suppressed ASD and myopathy in *Col4a1* mutant mice. However, the *MoGS1* locus had no effect on porencephaly penetrance and ICH severity – two phenotypes associated with Gould syndrome and previously

characterized in *Col4a1* mutant mice. The failure of *MoGS1* to suppress ICH severity was unexpected since the CASTB6F1 genetic context significantly suppresses this phenotype. These data suggest that the effect of the *MoGS1* locus is tissue-specific and that other modifiers of ICH exist in the CAST background.

Using subcongenic strains we refined the minimum critical interval for the suppressor locus to a 4.3 Mb region containing 18 protein-coding genes. Of the 18 positional candidate genes, 13 are expressed both in eyes and muscles, of which 6 contained missense variants between the CAST and B6 genomes; however, none were predicted to be deleterious. Mutations in one of these six genes, *Smarcal1*, causes a rare multisystem disorder (Schimke Immunoosseous dysplasia; OMIM# 242900) characterized by spondyloepiphyseal dysplasia, nephrotic syndrome, T cell immunodeficiency, and a portion of patients develop corneal opacity, myopia, astigmatism and optic atrophy (Boerkoel et al., 2000). SMARCAL1 (SWI/SNF related matrix associated, actin dependent regulator of chromatin, subfamily a-like 1) is chromatin remodeling protein involved in transcriptional regulation and DNA replication, repair and recombination (Bansbach et al., 2010). Compared to the B6 reference genome, CAST has two missense variations in *Smarcal1* that were predicted to be tolerated, and no change of expression was detected in P0 eyes. Of the four positional candidate genes that were differentially expressed, *Fn1* is a strong functional candidate. FN1 is a ubiquitously expressed extracellular glycoprotein which plays important roles in multiple processes by interacting with cell surface receptors, growth factors and other extracellular matrix proteins. In cell culture, FN1 promotes collagen IV deposition, assembly and incorporation into the extracellular matrix (Filla et al., 2017; Ngandu Mpoyi et al., 2020). Mice deficient for *Fn1* have aberrant lens placode formation and develop microphthalmia and cataracts (Hayes et al., 2012; Huang et al., 2011). Moreover, lack of FN1 in the skeletal muscle stem cell niche impairs muscle regeneration in aged mice (Lukjanenko et al., 2016). Collectively, these findings support our genetic and molecular evidence suggesting that that FN1 is a strong candidate as the gene responsible for the effects conferred by the *MoGS1* locus.

In general, the presence of mutant COL4A1 or COL4A2 leads to intracellular accumulation of mutant heterotrimers at the expense of their secretion. Intracellular heterotrimer accumulation represents a potential cell autonomous (proximal) insult (Jeanne and Gould, 2017; Mao et al., 2015). Impaired heterotrimer secretion can also lead to a cell non-autonomous (distal) insult caused by extracellular heterotrimer deficiency which can simultaneously perturb any number of presently unidentified cellular pathways. Intracellular accumulation and extracellular deficiency can be addressed simultaneously by targeting the proximal defect of protein misfolding. For example, pharmacologically promoting

heterotrimer secretion using a chemical chaperone, 4-phenylbutyrate (4PBA), alleviates diverse pathologies in *Col4a1* mutant mice (Hayashi et al., 2018; Jeanne et al., 2015; Jones et al., 2019; Labelle-Dumais et al., 2019). However, even in the absence of 4PBA, secretion of mutant heterotrimers is not completely abolished and the presence of mutant heterotrimers in the basement membrane represents a third and distinct class of insult. We demonstrated this proof of concept by showing that mice with a *Col4a1*^{G394V} mutation have disproportionately severe myopathy, which is exacerbated when they are treated with 4PBA (Labelle-Dumais et al., 2019). The affected residue, glycine 394, is adjacent to a putative integrin-binding domain (Parkin et al., 2011) implicating impaired integrin binding in skeletal myopathy. Thus, for mutations that impair subdomains of the heterotrimer responsible for executing specific extracellular functions, promoting secretion of mutant heterotrimers would still be predicted to ameliorate many extracellular functions. However, the increased levels of mutant heterotrimers in basement membranes may also exacerbate specific disease pathway(s) related to the function(s) of the impacted subdomain (Labelle-Dumais et al., 2019). Identifying the multitude of extracellular roles of COL4A1/A2 and how they are executed will be key in understanding fundamental aspects of matrix biology and the tissue-specific pathogenic mechanisms in Gould syndrome. Mapping functional subdomains on the $\alpha1\alpha1\alpha2$ heterotrimers will be important for genetically stratifying patients and may influence the prognosis and potential therapeutic approaches.

The unbiased nature of genetic screens leaves open the possibility for finding modifiers of either proximal or distal insults. By independently analyzing two phenotypes using an inbred strain (CAST) with broad phenotypic suppression, we sought to find modifiers that might represent therapeutic targets for the breadth of Gould Syndrome phenotypes. *MoGS1*, the only significant locus that we identified, conferred tissue-specific effects by suppressing ASD and myopathy but not ICH. This observation is consistent with a modifier that has an extracellular function and suggests that the two phenotypes have at least partially overlapping distal pathogenic mechanisms. In a previous study using the *Col4a1*$^{\Delta ex41}$ mutation we found that, compared to B6, the CASTB6F1 background appeared to reduce intracellular accumulation but did not increase extracellular COL4A1 levels (Jeanne et al., 2015). This observation implicated intracellular accumulation in pathogenesis but did not rule out the potential of compensation by other extracellular factors. The fact that *MoGS1* does not suppress ICH severity indicates that intracellular accumulation of mutant proteins may indeed be the primary pathogenic insult for that phenotype.

When we compared disease severity across a murine allelic series of 9 *Col4a1* and *Col4a2* mutations, we found that the *Col4a1*^{G394V} mutation had relatively little intracellular accumulation and that

Col4a1$^{+/G394V}$ mice had mild ICH but severe ASD and myopathy (Kuo et al., 2014). The proximity of the *Col4a1*^{G394V} mutation to a predicted integrin binding domain suggests that impaired integrin binding contributes to ocular dysgenesis and myopathy. Therefore, one possibility is that FN1 ameliorates ASD and skeletal myopathy caused by *Col4a1* mutations through compensatory integrin-mediated interactions. Consistent with this hypothesis, we detected evidence for increased integrin signaling in *Col4a1*$^{+/\Delta ex41}$ mice that was reduced by the *MoGS1* locus from CAST in the context of increased *Fn1* expression. Moreover, ILK and FAK are both previously implicated in lens development (Cammas et al., 2012; Harburger and Calderwood, 2009; Hayes et al., 2012; Samuelsson et al., 2007; Teo et al., 2014) and myopathy (Boppart and Mahmassani, 2019; Gheyara et al., 2007). Functional validation of this pathway with experimental manipulation and preclinical interventions would represent a significant advance in the understanding of the pathogenic mechanisms that contribute to Gould syndrome.

MATERIALS AND METHODS

Mice

All experiments were compliant with the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research and approved by the Institutional Animal Care and Use Committee at the University of California, San Francisco. *Col4a1*$^{+/\Delta ex41}$ mice were originally identified in a mutagenesis screen conducted at The Jackson Laboratory (Bar Harbor, ME) (Gould et al., 2007; Gould et al., 2005). Since then *Col4a1*$^{+/\Delta ex41}$ mice has been backcrossed to B6 for at least 20 generations. CAST mice were obtained from The Jackson Laboratory. All animals were maintained in full-barrier facilities free of specific pathogens on a 12-hour light/dark cycle with food and water ad libitum. Both male and female mice were used in this study.

Slit lamp biomicroscopy

Ocular anterior segment examinations were performed on mice at 1.0- 1.5 months of age using a slit lamp biomicroscope (Topcon SL-D7; Topcon Medical Systems, Oakland, NJ) attached to a digital SLR camera (Nikon D200; Nikon, Melville, NY).

Muscle analysis

At 2 months of age, mice were subjected to treadmill exercise and their quadriceps were harvested two days later. Exercise was 30 minutes with a 15° downhill grade on a treadmill equipped with a shock plate (Columbus Instruments, Columbus, OH). Animals were started at 6 m/min and increased by 3 m/min every 2 min until maximum of 15 m/min speed was reached. Quadriceps were dissected and frozen in liquid nitrogen cooled isopentane. Cryosections (10 μm) were collected at regular intervals

and stained with hematoxylin and eosin (H&E) for histopathology and the number of NPN were evaluated on a total of 12 sections for each muscle and the percentage of muscle fibers with NPN was quantified. The observers were masked to genotypes and counted between 2000–5000 muscle fibers per animal.

Brain histological analysis

ICH was assessed at 2 months of age by Perl's Prussian blue staining as previously described (Jeanne et al., 2015). Briefly, mice underwent transcardial perfusion with saline followed by 4% paraformaldehyde (PFA) and then posted fixed in 4% PFA overnight. Brains were dissected, cryoprotected in 30% sucrose and embedded in optimal cutting temperature compound (Sakura Finetek, Torrance, CA). Coronal cryo-sections (35 µm) regularly spaced along the rostro-caudal axis were stained with Prussian blue/Fast red and imaged. On each section, the percentage of brain area with Prussian blue staining was calculated using ImageJ software (National Institutes of Health). ICH severity was expressed as the average percentage of hemosiderin surface area on 28 sections for each brain. The presence or absence of porencephaly on sections used in ICH analysis was also recorded.

Genome scan

Genomic DNA was obtained from 192 (CAST X B6) F_2 mice carrying the *Col4a1*$^{+/\Delta ex41}$ mutation and genotyped at the UCSF Genomic Core Facility with a commercial single nucleotide polymorphism (SNP) panel (Illumina, San Diego, CA) which contains 646 informative SNPs for (CAST X B6) F_2 progeny. A genome-wide suppressor screen was performed in R (R Core Team, 2019) using the package R/qtl (Broman et al., 2003) treating the trait as binary. Genome-wide significance was established using 1000 permutation testing (Churchill and Doerge, 1994). A confidence interval for each QTL location was then calculated using Bayesian confidence sets (Sen and Churchill, 2001). The ASD score for each animal was averaged from both eyes and treated as a dichotomous variable where score of 1 was mild, and score of 1.5 -3 was severe. A subset (49) of these 192 mice was randomly selected in a separate mapping analysis for the muscle modifier as a quantitative trait.

Quantitative polymerase chain reaction (qPCR)

Eyes from P0 mice were dissected in PBS, immediately transferred in RNA*later* (ThermoFisher Scientific, USA) and stored at -80°C until use. Total RNA was extracted using TRIzol Reagent (ThermoFisher Scientific) according to the manufacturer's instructions and reverse transcribed to cDNA

using iScript cDNA synthesis kit (Bio-Rad, Hercules, CA). Quantitative PCR was performed on a Bio-Rad CFX96 real-time system using SsoFast Evagreen mix (Bio-Rad) and primers are listed in Table S3. Briefly, 10 ng of cDNA and 1.25 µM primers were used per reaction in a final volume of 10 µl. Each cycle consisted of denaturation at 95°C for 5s, followed by annealing and extension at 60°C for 5s for a total of 45 cycles. All experiments were run with technical duplicates and 4-6 biological replicates were used per group. The relative expression of each gene was normalized to *Hprt1* or *Gapdh* and analyzed using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

Western blot analyses

Quadriceps from P10 pups were dissected and total proteins extracted using radioimmunoprecipitation assay (RIPA) buffer (VWR, Radnor, PA) supplemented with Halt Protease and Phosphatase Inhibitor Cocktail (ThermoFisher Scientific), EDTA and 2 mM phenylmethylsulfonyl fluoride. Total proteins (10 µg) were separated on 4–15% gradient SDS-PAGE gels (Bio-Rad) and transferred to polyvinylidene fluoride (PVDF) membranes (BioRad). Membranes were blocked for 1 hr at room temperature in 5% non-fat milk in TBS containing 0.1% Tween-20 (TBST), incubated overnight at 4°C in primary antibodies diluted in 2% non-fat milk in TBST. Primary antibodies and dilutions used are as follow: rabbit anti-FN1 (Abcam #Ab2413 1:10000, Abcam), rabbit anti-pFAK (1:1000, Invitrogen #44624G), rabbit anti-FAK (1:1000, Santa Cruz #SC-557), rabbit anti-ILK (1:1000, Cell Signaling Technology #3862), mouse anti-ITGB1 (1:1000, BD Biosciences #610467), and mouse anti-GAPDH (1:500000, Millipore #MAB374). Following washes in TBST, membranes were incubated for 1 hr at room temperature with species-specific horseradish peroxidase conjugated secondary antibodies (1:10000, Jackson ImmunoResearch, West Grove, PA) diluted in 2% non-fat milk in TBST. Immunoreactivity was visualized using chemiluminescence (ECL or Luminata Forte substrate, ThermoFisher Scientific) and imaged using a ChemiDoc MP Imaging System (Bio-Rad) or exposed to X-ray films. Densitometric analyses were performed on low exposure images using the Quantity One software (Bio-Rad).

Statistics

Statistical analyses and graphs were prepared using GraphPad Prism v8.0 Software (GraphPad, La Jolla, CA). Multiple comparisons were carried using one-way ANOVA followed by Tukey post-test or Kruskal-Wallis test followed by Dunn's post-test. *p* values less than 0.05 were considered statistically significant.

ACKNOWLEDGEMENTS

We thank members of the Gould Laboratory for their critical reviews of the manuscript. We also thank Cassandre Labelle-Dumais for helpful discussion and guidance with the muscle studies.

COMPETING INTERESTS

M.J. is an employee of Genentech, Inc., a member of the Roche group. K.H. is an employee of Centrillion Technologies.

FUNDING

This work was supported by NIH grants R01EY019887 (DBG), R01NS096173 (DBG), Research to Prevent Blindness (DBG), That Man May See (MM and DBG), and Knights Templar Eye Foundation (MM) and Bright Focus (DBG). This research was supported, in part, by the UCSF Vision Core shared resource of the NIH/NEI P30 EY002162, and by an unrestricted grant from Research to Prevent Blindness, New York, NY.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: D.B.G and M.M. Performed experiments: M.M, T.P, M.J, and K.H. Data analysis: M.M, T.P, M.J, K.H, S.S and D.B.G. Contributed reagents, materials, and/or analysis tools: S.S. Writing of the manuscript: M.M and D.B.G.

Table 1. List of protein-coding genes in the refined *MoGS* locus

Gene name	Entrez gene ID	# of missense SNPs	# of deleterious SNPs	# of stop-gain SNPs	# of SNPs in splice region	# of SNPs in 3'UTR	Eye expression	Muscle expression	Relative mRNA expression	*p - value*
Erbb4	13869	0	0	0	0	0	Yes (> E18)	Yes (< E18)	0.88	0.559
Ikzf2	22779	2	0	0	0	40	Yes (> P4)	Yes (> P4)	0.79	0.054
Spag16	66722	8	0	1	5	39	-	-	NA	NA
Vwc2l	320460	0	0	0	0	21	Yes (> P4)	-	NA	NA
Bard1	12021	3	0	0	0	29	Yes (> P4)	-	NA	NA
Abca12	74591	5	0	0	4	4	-	-	NA	NA
Atic	108147	3	0	0	2	17	Yes (> E15)	Yes (> P4)	0.82	0.267
Fn1	14268	3	0	0	10	5	Yes (> E8)	Yes (> E13)	1.45	0.007**
Mreg	381269	0	0	0	0	16	Yes (> P4)	Yes (> P4)	1.43	0.026*
Pecr	111175	0	0	0	1	8	Yes (> P4)	Yes (> P4)	0.80	0.026*
Tmem169	271711	1	0	0	1	7	Yes (> P4)	Yes (> P4)	0.97	0.724
Xrcc5	22596	3	0	0	5	5	Yes (> P4)	Yes (> P4)	0.88	0.496
March4	381270	0	0	0	0	13	Yes (> P4)	Yes (> P4)	1.20	0.452
Smarcal1	54380	2	0	0	0	0	Yes (> P4)	Yes (> P4)	0.95	0.459
Ankar	319695	5	3	0	0	0	-	-	NA	NA
Rpl37a	19981	0	0	0	2	1	Yes (> P4)	Yes (> P4)	1.07	0.192
Igfbp2	16008	0	0	0	0	1	Yes (> E13)	Yes (> P4)	0.83	0.023*
Igfbp5	16011	0	0	0	0	36	Yes (> E13)	Yes (> E15)	1.34	0.076

E, embryonic day; P, postnatal; -, expression not detectable; NA, data not available; * $p < 0.05$; ** $p < 0.01$ by Student's t-test

REFERENCES

Bansbach, C. E., Boerkoel, C. F. and Cortez, D. (2010). SMARCAL1 and replication stress: an explanation for SIOD? *Nucleus* **1**, 245-8.

Bilguvar, K., DiLuna, M. L., Bizzarro, M. J., Bayri, Y., Schneider, K. C., Lifton, R. P., Gunel, M. and Ment, L. R. (2009). COL4A1 mutation in preterm intraventricular hemorrhage. *J Pediatr* **155**, 743-5.

Boerkoel, C. F., O'Neill, S., André, J. L., Benke, P. J., Bogdanović, R., Bulla, M., Burguet, A., Cockfield, S., Cordeiro, I., Ehrich, J. H. et al. (2000). Manifestations and treatment of Schimke immuno-osseous dysplasia: 14 new cases and a review of the literature. *Eur J Pediatr* **159**, 1-7.

Boppart, M. D. and Mahmassani, Z. S. (2019). Integrin signaling: linking mechanical stimulation to skeletal muscle hypertrophy. *Am J Physiol Cell Physiol* **317**, C629-C641.

Breedveld, G., de Coo, I. F., Lequin, M. H., Arts, W. F., Heutink, P., Gould, D. B., John, S. W., Oostra, B. and Mancini, G. M. (2006). Novel mutations in three families confirm a major role of COL4A1 in hereditary porencephaly. *J Med Genet* **43**, 490-5.

Broman, K. W., Wu, H., Sen, S. and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889-90.

Cammas, L., Wolfe, J., Choi, S. Y., Dedhar, S. and Beggs, H. E. (2012). Integrin-linked kinase deletion in the developing lens leads to capsule rupture, impaired fiber migration and non-apoptotic epithelial cell death. *Invest Ophthalmol Vis Sci* **53**, 3067-81.

Ceco, E. and McNally, E. M. (2013). Modifying muscular dystrophy through transforming growth factor-β. *FEBS J* **280**, 4198-209.

Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-71.

Coupry, I., Sibon, I., Mortemousque, B., Rouanet, F., Mine, M. and Goizet, C. (2010). Ophthalmological features associated with COL4A1 mutations. *Arch Ophthalmol* **128**, 483-9.

de Vries, L. S., Koopman, C., Groenendaal, F., Van Schooneveld, M., Verheijen, F. W., Verbeek, E., Witkamp, T. D., van der Worp, H. B. and Mancini, G. (2009). COL4A1 mutation in two preterm siblings with antenatal onset of parenchymal hemorrhage. *Ann Neurol* **65**, 12-8.

Favor, J., Gloeckner, C. J., Janik, D., Klempt, M., Neuhauser-Klaus, A., Pretsch, W., Schmahl, W. and Quintanilla-Fend, L. (2007). Type IV procollagen missense mutations associated with defects of the eye, vascular stability, the brain, kidney function and embryonic or postnatal viability in the mouse, Mus musculus: an extension of the Col4a1 allelic series and the identification of the first two Col4a2 mutant alleles. *Genetics* **175**, 725-36.

Fidler, A. L., Darris, C. E., Chetyrkin, S. V., Pedchenko, V. K., Boudko, S. P., Brown, K. L., Gray Jerome, W., Hudson, J. K., Rokas, A. and Hudson, B. G. (2017). Collagen IV and basement membrane at the evolutionary dawn of metazoan tissues. *Elife* **6**.

Filla, M. S., Dimeo, K. D., Tong, T. and Peters, D. M. (2017). Disruption of fibronectin matrix affects type IV collagen, fibrillin and laminin deposition into extracellular matrix of human trabecular meshwork (HTM) cells. *Exp Eye Res* **165**, 7-19.

Gheyara, A. L., Vallejo-Illarramendi, A., Zang, K., Mei, L., St-Arnaud, R., Dedhar, S. and Reichardt, L. F. (2007). Deletion of integrin-linked kinase from skeletal muscles of mice resembles muscular dystrophy due to alpha 7 beta 1-integrin deficiency. *Am J Pathol* **171**, 1966-77.

Giorgio, E., Vaula, G., Bosco, G., Giacone, S., Mancini, C., Calcia, A., Cavalieri, S., Di Gregorio, E., Rigault De Longrais, R., Leombruni, S. et al. (2015). Two families with novel missense mutations in COL4A1: When diagnosis can be missed. *J Neurol Sci* **352**, 99-104.

Gould, D. B., Marchant, J. K., Savinova, O. V., Smith, R. S. and John, S. W. (2007). Col4a1 mutation causes endoplasmic reticulum stress and genetically modifiable ocular dysgenesis. *Hum Mol Genet* **16**, 798-807.

Gould, D. B., Phalan, F. C., Breedveld, G. J., van Mil, S. E., Smith, R. S., Schimenti, J. C., Aguglia, U., van der Knaap, M. S., Heutink, P. and John, S. W. (2005). Mutations in Col4a1 cause perinatal cerebral hemorrhage and porencephaly. *Science* **308**, 1167-71.

Gould, D. B., Phalan, F. C., van Mil, S. E., Sundberg, J. P., Vahedi, K., Massin, P., Bousser, M. G., Heutink, P., Miner, J. H., Tournier-Lasserve, E. et al. (2006). Role of COL4A1 in small-vessel disease and hemorrhagic stroke. *N Engl J Med* **354**, 1489-96.

Harburger, D. S. and Calderwood, D. A. (2009). Integrin signalling at a glance. *J Cell Sci* **122**, 159-63.

Hayashi, G., Labelle-Dumais, C. and Gould, D. B. (2018). Use of sodium 4-phenylbutyrate to define therapeutic parameters for reducing intracerebral hemorrhage and myopathy in Col4a1 mutant mice. *Dis Model Mech* **11**.

Hayes, J. M., Hartsock, A., Clark, B. S., Napier, H. R., Link, B. A. and Gross, J. M. (2012). Integrin alpha5/fibronectin1 and focal adhesion kinase are required for lens fiber morphogenesis in zebrafish. *Mol Biol Cell* **23**, 4725-38.

Hu, P. and Luo, B. H. (2013). Integrin bi-directional signaling across the plasma membrane. *J Cell Physiol* **228**, 306-12.

Huang, J., Rajagopal, R., Liu, Y., Dattilo, L. K., Shaham, O., Ashery-Padan, R. and Beebe, D. C. (2011). The mechanism of lens placode formation: a case of matrix-mediated morphogenesis. *Dev Biol* **355**, 32-42.

Humphries, J. D., Chastney, M. R., Askari, J. A. and Humphries, M. J. (2019). Signal transduction via integrin adhesion complexes. *Curr Opin Cell Biol* **56**, 14-21.

Jeanne, M. and Gould, D. B. (2017). Genotype-phenotype correlations in pathology caused by collagen type IV alpha 1 and 2 mutations. *Matrix Biol* **57-58**, 29-44.

Jeanne, M., Jorgensen, J. and Gould, D. B. (2015). Molecular and Genetic Analyses of Collagen Type IV Mutant Mouse Models of Spontaneous Intracerebral Hemorrhage Identify Mechanisms for Stroke Prevention. *Circulation* **131**, 1555-65.

Johansson, S., Svineng, G., Wennerberg, K., Armulik, A. and Lohikangas, L. (1997). Fibronectin-integrin interactions. *Front Biosci* **2**, d126-46.

Jones, F. E., Murray, L. S., McNeilly, S., Dean, A., Aman, A., Lu, Y., Nikolova, N., Malomgre, R., Horsburgh, K., Holmes, W. M. et al. (2019). 4-Sodium phenyl butyric acid has both efficacy and counter-indicative effects in the treatment of Col4a1 disease. *Hum Mol Genet* **28**, 628-638.

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M. et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-94.

Kuo, D. S., Labelle-Dumais, C., Mao, M., Jeanne, M., Kauffman, W. B., Allen, J., Favor, J. and Gould, D. B. (2014). Allelic heterogeneity contributes to variability in ocular dysgenesis, myopathy and brain malformations caused by Col4a1 and Col4a2 mutations. *Hum Mol Genet* **23**, 1709-22.

Labelle-Dumais, C., Dilworth, D. J., Harrington, E. P., de Leau, M., Lyons, D., Kabaeva, Z., Manzini, M. C., Dobyns, W. B., Walsh, C. A., Michele, D. E. et al. (2011). COL4A1 mutations cause ocular dysgenesis, neuronal localization defects, and myopathy in mice and Walker-Warburg syndrome in humans. *PLoS Genet* **7**, e1002062.

Labelle-Dumais, C., Schuitema, V., Hayashi, G., Hoff, K., Gong, W., Dao, D. Q., Ullian, E. M., Oishi, P., Margeta, M. and Gould, D. B. (2019). COL4A1 Mutations Cause Neuromuscular Disease with Tissue-Specific Mechanistic Heterogeneity. *Am J Hum Genet* **104**, 847-860.

Laurie, G. W., Bing, J. T., Kleinman, H. K., Hassell, J. R., Aumailley, M., Martin, G. R. and Feldmann, R. J. (1986). Localization of binding sites for laminin, heparan sulfate proteoglycan and fibronectin on basement membrane (type IV) collagen. *J Mol Biol* **189**, 205-16.

Laurie, G. W., Leblond, C. P. and Martin, G. R. (1982). Localization of type IV collagen, laminin, heparan sulfate proteoglycan, and fibronectin to the basal lamina of basement membranes. *J Cell Biol* **95**, 340-4.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8.

Lukjanenko, L., Jung, M. J., Hegde, N., Perruisseau-Carrier, C., Migliavacca, E., Rozo, M., Karaz, S., Jacot, G., Schmidt, M., Li, L. et al. (2016). Loss of fibronectin from the aged stem cell niche affects the regenerative capacity of skeletal muscle in mice. *Nat Med* **22**, 897-905.

Mao, M., Alavi, M. V., Labelle-Dumais, C. and Gould, D. B. (2015). Type IV Collagens and Basement Membrane Diseases: Cell Biology and Pathogenic Mechanisms. *Curr Top Membr* **76**, 61-116.

Meuwissen, M. E., Halley, D. J., Smit, L. S., Lequin, M. H., Cobben, J. M., de Coo, R., van Harssel, J., Sallevelt, S., Woldringh, G., van der Knaap, M. S. et al. (2015). The expanding phenotype of COL4A1 and COL4A2 mutations: clinical data on 13 newly identified families and a review of the literature. *Genet Med* **17**, 843-53.

Meyer, K. J. and Anderson, M. G. (2017). Genetic modifiers as relevant biological variables of eye disorders. *Hum Mol Genet* **26**, R58-r67.

Miyamoto, S., Katz, B. Z., Lafrenie, R. M. and Yamada, K. M. (1998). Fibronectin and integrins in cell adhesion, signaling, and morphogenesis. *Ann N Y Acad Sci* **857**, 119-29.

Ngandu Mpoyi, E., Cantini, M., Sin, Y. Y., Fleming, L., Zhou, D. W., Costell, M., Lu, Y., Kadler, K., Garcia, A. J., Van Agtmael, T. et al. (2020). Material-driven fibronectin assembly rescues matrix defects due to mutations in collagen IV in fibroblasts. *Biomaterials* **252**, 120090.

Parkin, J. D., San Antonio, J. D., Pedchenko, V., Hudson, B., Jensen, S. T. and Savige, J. (2011). Mapping structural landmarks, ligand binding sites, and missense mutations to the collagen IV heterotrimers predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes. *Hum Mutat* **32**, 127-43.

Poschl, E., Schlotzer-Schrehardt, U., Brachvogel, B., Saito, K., Ninomiya, Y. and Mayer, U. (2004). Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development. *Development* **131**, 1619-28.

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ricard-Blum, S. (2011). The collagen family. *Cold Spring Harb Perspect Biol* **3**, a004978.

Rødahl, E., Knappskog, P. M., Majewski, J., Johansson, S., Telstad, W., Kråkenes, J. and Boman, H. (2013). Variants of anterior segment dysgenesis and cerebral involvement in a large family with a novel COL4A1 mutation. *Am J Ophthalmol* **155**, 946-53.

Sakai, T., Li, S., Docheva, D., Grashoff, C., Sakai, K., Kostka, G., Braun, A., Pfeifer, A., Yurchenco, P. D. and Fassler, R. (2003). Integrin-linked kinase (ILK) is required for polarizing the epiblast, cell adhesion, and controlling actin accumulation. *Genes Dev* **17**, 926-40.

Samuelsson, A. R., Belvindrah, R., Wu, C., Muller, U. and Halfter, W. (2007). Beta1-integrin signaling is essential for lens fiber survival. *Gene Regul Syst Bio* **1**, 177-89.

Sen, S. and Churchill, G. A. (2001). A statistical framework for quantitative trait mapping. *Genetics* **159**, 371-87.

Shah, S., Ellard, S., Kneen, R., Lim, M., Osborne, N., Rankin, J., Stoodley, N., van der Knaap, M., Whitney, A. and Jardine, P. (2012). Childhood presentation of COL4A1 mutations. *Dev Med Child Neurol* **54**, 569-74.

Sibon, I., Coupry, I., Menegon, P., Bouchet, J. P., Gorry, P., Burgelin, I., Calvas, P., Orignac, I., Dousset, V., Lacombe, D. et al. (2007). COL4A1 mutation in Axenfeld-Rieger anomaly with leukoencephalopathy and stroke. *Ann Neurol* **62**, 177-84.

Teo, Z. L., McQueen-Miscamble, L., Turner, K., Martinez, G., Madakashira, B., Dedhar, S., Robinson, M. L. and de Iongh, R. U. (2014). Integrin linked kinase (ILK) is required for lens epithelial cell survival, proliferation and differentiation. *Exp Eye Res* **121**, 130-42.

Thaung, C., West, K., Clark, B. J., McKie, L., Morgan, J. E., Arnold, K., Nolan, P. M., Peters, J., Hunter, A. J., Brown, S. D. et al. (2002). Novel ENU-induced eye mutations in the mouse: models for human eye disease. *Hum Mol Genet* **11**, 755-67.

Vahedi, K., Kubis, N., Boukobza, M., Arnoult, M., Massin, P., Tournier-Lasserve, E. and Bousser, M. G. (2007). COL4A1 mutation in a patient with sporadic, recurrent intracerebral hemorrhage. *Stroke* **38**, 1461-4.

Vandenberg, P., Kern, A., Ries, A., Luckenbill-Edds, L., Mann, K. and Kuhn, K. (1991). Characterization of a type IV collagen major cell binding site with affinity to the alpha 1 beta 1 and the alpha 2 beta 1 integrins. *J Cell Biol* **113**, 1475-83.

Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. and Ng, P. C. (2016). SIFT missense predictions for genomes. *Nat Protoc* **11**, 1-9.

Vieira, N. M., Elvers, I., Alexander, M. S., Moreira, Y. B., Eran, A., Gomes, J. P., Marshall, J. L., Karlsson, E. K., Verjovski-Almeida, S., Lindblad-Toh, K. et al. (2015). Jagged 1 Rescues the Duchenne Muscular Dystrophy Phenotype. *Cell* **163**, 1204-1213.

Vilain, C., Van Regemorter, N., Verloes, A., David, P. and Van Bogaert, P. (2002). Neuroimaging fails to identify asymptomatic carriers of familial porencephaly. *Am J Med Genet* **112**, 198-202.

White, E. S., Baralle, F. E. and Muro, A. F. (2008). New insights into form and function of fibronectin splice variants. *J Pathol* **216**, 1-14.

Yoneda, Y., Haginoya, K., Kato, M., Osaka, H., Yokochi, K., Arai, H., Kakita, A., Yamamoto, T., Otsuki, Y., Shimizu, S. et al. (2013). Phenotypic spectrum of COL4A1 mutations: porencephaly to schizencephaly. *Ann Neurol* **73**, 48-57.

Zagaglia, S., Selch, C., Nisevic, J. R., Mei, D., Michalak, Z., Hernandez-Hernandez, L., Krithika, S., Vezyroglou, K., Varadkar, S. M., Pepler, A. et al. (2018). Neurologic phenotypes associated with COL4A1/2 mutations: Expanding the spectrum of disease. *Neurology* **91**, e2078-e2088.

Zollinger, A. J. and Smith, M. L. (2017). Fibronectin, the extracellular glue. *Matrix Biol* **60-61**, 27-37.

FIGURE LEGENDS

Figure 1. A genome-wide screen identified a single modifier locus on Chr 1 for ASD and myopathy. (**A**) Schematic representation of the (CAST X B6) F_2 cross. One hundred ninety two mutant (CAST X B6-*Col4a1*$^{+/\Delta ex41}$) F_2 progeny were genotyped and phenotyped for ASD. (**B**) Representative slit-lamp images illustrating scoring of ASD severity: mild, score = 1; moderate, score = 2; severe, score = 3. (**C**) A one-dimensional genome scan identified a locus on Chr 1 for ASD (max. LOD score = 11.2, 99% Bayesian confidence interval extending from 51.3 - 73.0 Mb, Ensembl GRCm38.p6). Solid horizontal line, 5% false positive threshold; dashed horizontal line, 10% threshold. (**D**) A genome-wide scan conditioned on the Chr 1 modifier locus suggests that no other single locus has a strong effect on ASD. Dashed horizontal line: 20% false positive threshold. (**E**) Representative images of cross sections from quadriceps stained with H&E showing variable myopathy severity in *Col4a1*$^{+/\Delta ex41}$ mice with different genetic backgrounds. Arrows indicate muscle fibers with NPN. Scale bar = 20 μm. (**F**) To identify skeletal muscle modifier loci, a subset (49) of the 192 (CAST X B6) F_2 progeny was assessed for myopathy. One dimensional genome scan identified a myopathy modifier at the same position as the ASD modifier (Max. LOD score = 5.61).

Figure 2. An incipient congenic strain containing the CAST-derived *MoGS1* locus suppressed ASD and myopathy, but not ICH. (**A**) Evaluation of eyes from the incipient congenic mice at N5F2 validates *MoGS1* as a genetic suppressor of ASD. The bar graph shows the percentage of eyes at each ASD severity level from *Col4a1*$^{+/\Delta ex41}$ mice that are homozygous for B6 alleles (B/B), heterozygous or homozygous for CAST alleles (C/B and C/C, respectively) in the congenic interval. ASD for *Col4a1*$^{+/\Delta ex41}$ mice on the pure B6 and CASTB6F1 backgrounds are also shown. *$p < 0.05$; ****$p < 0.0001$, by Kruskal-Wallis test followed by Dunn's multiple comparison test. (**B**) Evaluation of muscles from the N5F2 congenic mice validates *MoGS1* as a genetic suppressor of skeletal myopathy. Data are presented as mean ± SEM. *$p < 0.05$, by one-way ANOVA followed by Tukey's multiple comparison test. (**C**) Penetrance of porencephaly for *Col4a1*$^{+/\Delta ex41}$ mice with different genetic backgrounds was determined by brain histology. Top representative images show absence or presence of porencephaly. * marks brain area with porencephaly. (**D**) ICH in mice was assessed using Perl's Prussian blue staining. Data for each mouse are shown as a percentage of brain area with Prussian blue staining averaged over 28 sections from each brain. Representative images of Prussian blue-stained brain sections and respective Prussian blue staining quantification are shown above the graphs. Sample sizes are indicated in parentheses. Data are presented as mean ± SEM. *$p < 0.05$, Student's t-test for comparison between B6 and CASTB6F1; One-way ANOVA followed by Tukey's test for multiple comparisons between B/B, C/B, and C/C.

Figure 3. Fine mapping of the *MoGS1* locus. (**A**) Schematic illustration of subcongenic lines containing different intervals of the CAST modifier locus backcrossed onto the B6 background for 5 generations. Subcongenic Line 2 (68723089 – 72973981bp on Chr 1, Ensembl GRCm38.p6) ameliorated ASD and myopathy severity. Black line, chromosome fragments from B6; grey bar, chromosome fragments from CAST. CI, critical interval. (**B**) Analyses of ASD severity in mice homozygous for B6 alleles (B/B), heterozygous or homozygous for CAST alleles (C/B and C/C, respectively) at the refined distal interval. Percentage of eyes at each ASD severity level was shown. ***p<0.001 by Kruskal-Wallis test followed by Dunn's multiple comparison test. (**C**) Analysis of myopathy by NPN showing severity in mice with different genotypes at the refined distal interval. Sample size indicated in parentheses. Data are presented as mean ± SEM.*p < 0.05, one-way ANOVA followed by Tukey's multiple comparison test.

Figure 4. Evaluating FN1 as a candidate modifier gene. (**A**) Quantitative qPCR analysis showed increased expression of *Fn1* mRNA in P0 eyes from *Col4a1*$^{+/+}$ mice homozygous for the CAST interval at the refined *MoGS1* locus (*+/+;C/C*) compared to mice homozygous for the B6 interval (*+/+;B/B*). *n* = 4 per genotype. Data are presented as mean ± SEM. *p < 0.05, by Student's t-test. (**B-C**) Representative images (B) and quantification (C) of Western-blot analyses of P10 quadriceps showing a trend towards increased FN1 protein levels in mice homozygous for CAST alleles at the refined *MoGS1* locus *(+/+;C/C* and *+/Δex41;C/C*) compared to mice homozygous for B6 alleles (*+/+;B/B* and *+/Δex41;B/B*). ILK levels were increased in *Col4a1*$^{+/Δex41}$ mice irrespective of the genotype at the *MoGS1* locus (*+/Δex41;B/B* and *+/Δex41;C/C*, *p = 0.06* for *MoGS1*$^{C/C}$) compared to their corresponding *Col4a1*$^{+/+}$ controls (*+/+;B/B, and +/+;C/C,* respectively). Ratios of pFAK/FAK, were significantly elevated in *Col4a1*$^{+/Δex41}$ mice homozygous for *MoGS1* B6 alleles *(+/Δex41;B/B*) but not CAST alleles (*+/Δex41;C/C*) (*p* = 0.35) compared to their corresponding *Col4a1*$^{+/+}$ controls (*+/+;B/B, and +/+;C/C,* respectively) suggesting altered integrin signaling that is partially restored by the *MoGS1* locus. Levels of ITBG1 did not differ between genotypes. *n* = 4, 5, 5, 5 for *+/+;B/B*, *+/Δex41;B/B*, *+/+;C/C* and *+/Δex41;C/C* respectively. Data are presented as mean ± SEM. *p < 0.05, **p < 0.01 by one-way ANOVA followed by Sidak's multiple comparison test.

Figure S1. Subcongenic Line 1 did not confer the modifier effect. (**A-B**) A ~10 Mb proximal portion of the CAST *ModGS1* locus (52418578-61808145bp on Chr 1, Ensembl GRCm38.p6) showed no protective effect for ASD (A) or skeletal myopathy (B) phenotypes seen in *Col4a1*$^{+/Δex41}$ mice. Sample

size indicated in parentheses. Data are presented as mean ± SEM. one-way ANOVA followed by Tukey's multiple comparison test.

Table S1. List of protein-coding and non-coding genes in the *ModGS1* locus.

Table S2. SNPs, small InDels and structural variants in *Fn1*.

Table S3. Primer sequences for quantitative PCR.

700

Figure 1

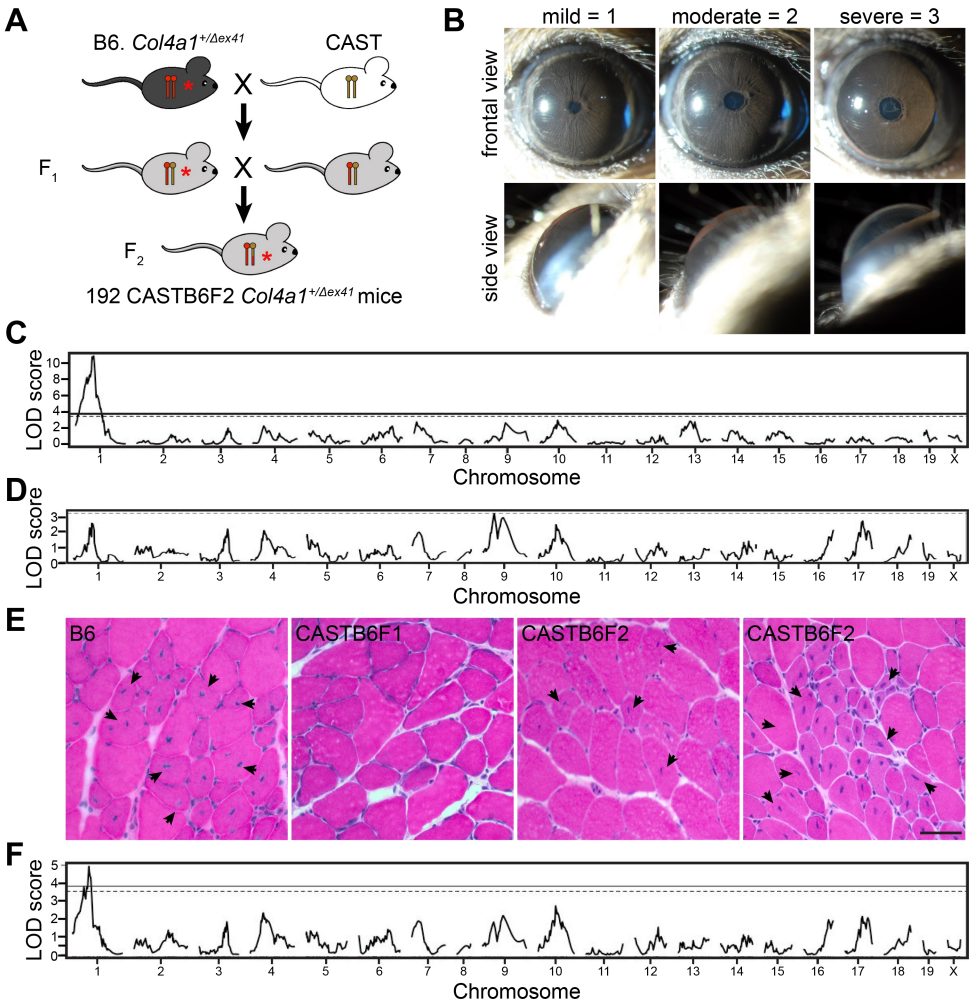

Figure 2

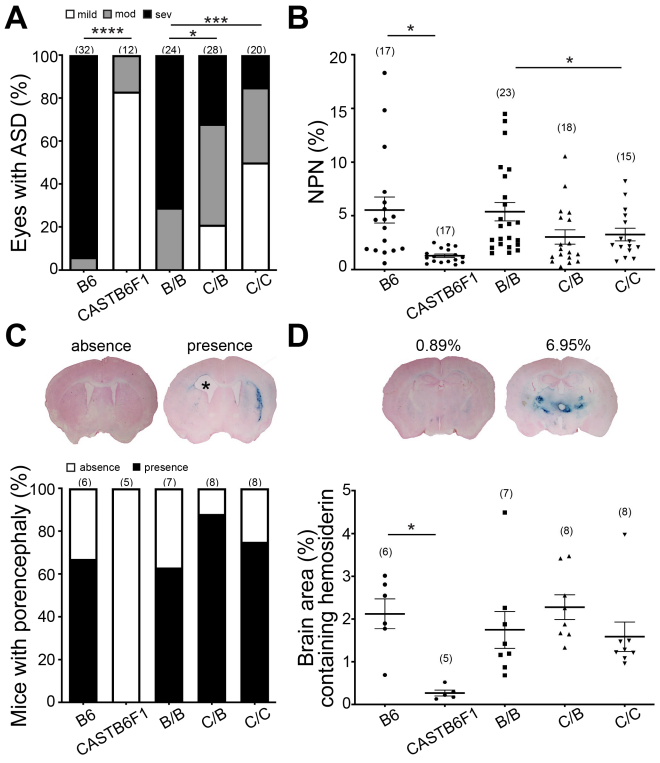

Figure 3

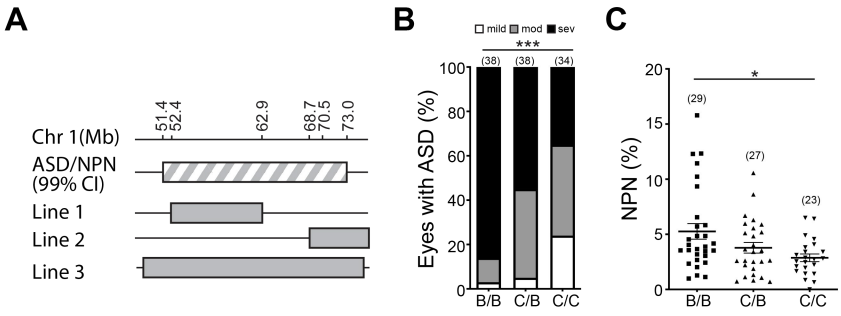

Figure 4