1 Genetic association and characterization of *FSTL5* in isolated clubfoot

2

3 Anas M Khanshour,[1] Yared Kidane,[1] Julia Kozlitina,[2] Reuel Cornelia,[1] Alexandra

4 Rafipay,[3] Vanessa De Mello,[3] Mitchell Weston,[4] Nandina Paria,[1] Aysha Khalid,[1]

5 Jacqueline T Hecht,[5] Matthew B Dobbs,[6] B. Stephens Richards,[7,8] Neil Vargesson,[3] F

6 Kent Hamra,[9] Megan Wilson,[4] Carol Wise,[1,2,7,8,10] Christina A Gurnett,[11] Jonathan J

7 Rios[1,2,7,8,10,*]

8

9

10 [1]Center for Pediatric Bone Biology and Translational Research, Scottish Rite for

11 Children, Dallas, TX, 75219, USA

12 [2]McDermott Center for Human Growth and Development, UT Southwestern Medical

13 Center, Dallas, TX, USA

14 [3]School of Medicine, Medical Sciences & Nutrition, Institute of Medical Sciences,

15 University of Aberdeen, Foresterhill, Aberdeen, UK

16 [4]Department of Anatomy and Structural Biology, University of Otago, Dunedin 9016,

17 New Zealand

18 [5]Department of Pediatrics, McGovern Medical School, University of Texas Health,

19 Houston, TX, USA

20 [6]Paley Orthopedic and Spine Institute, West Palm Beach, FL, USA

21 [7]Department of Orthopaedics, Scottish Rite for Children, Dallas, TX 75219, USA

22 [8]Department of Orthopaedic Surgery, UT Southwestern Medical Center, Dallas, TX,

23 USA

24 [9]Department of Obstetrics and Gynecology, Cecil H. & Ida Green Center for
25 Reproductive Biology Sciences, UT Southwestern Medical Center, Dallas, TX, USA
26 [10]Department of Pediatrics, UT Southwestern Medical Center, Dallas, TX, USA
27 [11]Department of Neurology, School of Medicine, Washington University, St. Louis, MO,
28 USA
29 *Corresponding author: Jonathan Rios
30 (P) 214-648-7658
31 (F) 214-559-7872
32 (E) Jonathan.Rios@tsrh.org
33
34 **Running title**: *FSTL5* associated with isolated clubfoot
35 **Keywords**: genome-wide association study, clubfoot, FSTL5, imputation
36 Word Count: 7,378
37 Figures: 4
38 Tables: 0
39
40
41

42 **ABSTRACT**

43 Talipes equinovarus (clubfoot, TEV) is a congenital rotational foot deformity
44 occurring in 1 per 1,000 births with increased prevalence in males compared to females.
45 The genetic etiology of isolated clubfoot (iTEV) remains unclear. Using a genome-wide
46 association study, we identified a locus within *FSTL5*, encoding *follistatin-like 5,*
47 significantly associated with iTEV. *FSTL5* is an uncharacterized gene whose potential
48 role in embryonic and post-natal development were previously unstudied. Utilizing
49 multiple model systems, we found that *Fstl5* was expressed during later stages of
50 embryonic hindlimb development, and, in mice, expression was restricted to the
51 condensing cartilage anlage destined to form the limb skeleton. In the post-natal growth
52 plate, *Fstl5* was specifically expressed in pre-hypertrophic chondrocytes. As *Fstl5* knock-
53 out rats displayed no gross malformations, we engineered a conditional transgenic mouse
54 line (*Fstl5*^{LSL}) to over-express *Fstl5* in skeletal osteochondroprogenitors. We observed
55 that hindlimbs were slightly shorter and that bone mineral density was reduced in adult
56 male, but not female, *Prrx1*-cre;*Fstl5*^{LSL} mice compared to control. No overt clubfoot-
57 like deformity was observed in *Prrx1*-cre;*Fstl5*^{LSL} mice, suggesting FSTL5 may function
58 in other cell types to contribute to iTEV pathogenesis. Interrogating published mouse
59 embryonic single-cell expression data showed *Fstl5* was expressed in cell lineage sub-
60 clusters whose transcriptomes were associated with neural system development.
61 Moreover, our results suggest lineage-specific expression of the *Fstl* genes correlates
62 with their divergent roles as modulators of TGF-ß and BMP signaling. Results from this
63 study associate *FSTL5* with iTEV and suggest an potential sexually dimorphic role for
64 *Fstl5 in vivo*.

65 **INTRODUCTION**

66 Talipes equinovarus (clubfoot, TEV) is a congenital rotational foot deformity
67 occurring in ~1 per 1,000 live births and affecting either or both limbs (1) (**Fig. 1A**).
68 Infants are often treated with non-operative methods to gradually manipulate the foot to a
69 corrected position, though relapse of the deformity may occur and require surgical
70 intervention (2, 3). Syndromic TEV coincides with other diagnoses, such as cerebral
71 palsy or spina bifida, and may be caused by single-gene mutations or large chromosomal
72 abnormalities (1, 4). Most often, however, clubfoot is isolated (iTEV), presenting as a
73 distal limb malformation without significant co-occurring sequelae. The prenatal
74 developmental etiology of iTEV remains poorly understood, though it is generally
75 accepted that bone, muscle, nerve, and tendon may all contribute to its development.

76 Mendelian inheritance of iTEV has been described in multi-generational families
77 (5-7). By linkage analysis and resequencing, a novel dominant-negative mutation in
78 *PITX1* segregated with clubfoot in a multi-generational family (8). Subsequent analyses
79 identified rare copy number variations including the *PITX1* or *TBX4* genes, among others,
80 in clubfoot, (9-12), though these account for only a small fraction of clubfoot cases.
81 While *PITX1* haploinsufficiency was implicated in clubfoot, mis-expression of *PITX1*
82 from deletion or structural rearrangement of the 5' regulatory region was shown to cause
83 Liebenberg syndrome, which is characterized by partial arm-to-leg transformation (13).
84 Therefore, alterations in these master regulators of limb development, through genetic
85 mutation, may lead to pleiotropic limb malformations.

86 Genetically-engineered mouse models have shown haploinsufficiency or mis-
87 expression of *Pitx1* results in phenotypes resembling clubfoot and Liebenberg syndrome,

88 respectively (9, 13). More recently, a spontaneous mutation in *Limk1* was identified in the
89 "*pma*" mouse, which develops a fully penetrant recessive clubfoot-like deformity (14). In
90 homozygous *pma*/*pma* mice, increased embryonic hindlimb expression and activation of
91 Limk1 was associated with inactivation of Cofilin leading to defects in hindlimb muscle
92 innervation. Results from these and other mouse models (15) suggest multiple
93 developmental origins of clubfoot.

94 iTEV is considered a complex genetic disorder likely impacted by a spectrum of
95 associated genetic loci, and the prevalence varies greatly between ethnic populations and
96 between genders (1, 5, 16). Here, we report results from a genome-wide associated study
97 (GWAS) of iTEV, which identified an associated locus within *FSTL5*. We show *Fstl5* is
98 expressed in the embryonic hindlimb and the post-natal growth plate, and conditional
99 over-expression of *Fstl5* in osteochondroprogenitors resulted in sexually dimorphic
100 differences in skeletal development. Unlike other *Fstl* genes, our results suggest *Fstl5*
101 likely does not modulate TGF-ß and BMP signaling. Finally, using results from publicly-
102 available mouse embryonic single-cell expression profiling, we provide evidence for a
103 neurogenic role of *Fstl5* in clubfoot.
104

105 **RESULTS**

106 ***FSTL5* is associated with iTEV**

107 To identify novel loci associated with iTEV, we performed a GWAS with 399 Caucasian
108 iTEV subjects (previously described (17)) and 7,820 ethnicity-matched controls from the
109 population-based Atherosclerosis Risk in Communities study (dbGAP: phs000280.v3.p1)
110 (18) (**Table S1, Fig. S1A**). Following imputation and quality control filtering, 7,794,536

111 autosomal variants with overall minor allele frequency >0.02 and imputation quality
112 (Rsq) >0.3 were tested for association with iTEV using logistic regression, adjusting for
113 gender and ten principal components. A single locus on chromosome 4 exceeded
114 genome-wide significance (**Fig. 1B, Supplementary Figure 1B,C**). The top-associated
115 variant (rs76973778, Rsq= 0.87, p= 1.92 x10^{-8}; OR= 0.35 [0.26-0.60]) was located within
116 intron 3 of the *Follistatin-like 5* (*FSTL5*) gene (**Fig. 1C**). Because clubfoot is twice as
117 frequent in males compared to females (6), we investigated the association of *FSTL5* with
118 iTEV in males and females separately. The *FSTL5* association was more strongly evident
119 in males (P=4.41e^{-8}, OR= 0.26 [0.19-0.50]) compared to females (P=0.02, OR= 0.51
120 [0.33-0.92]), which was attributed to reduced statistical power in female-only analyses
121 (**Fig. 1D,E; Supplementary Table 2**). To confirm the association with *FSTL5*,
122 rs76811724, which was in high linkage-disequilibrium with rs76973778 (R^2=0.94;
123 D'=1), was genotyped in an independent ethnicity-matched cohort (**See Methods, Table**
124 **S1**). The rs76811724 variant showed a highly significant and directionally concordant
125 association with iTEV in the discovery cohort (Rsq= 0.89, p= 2.06 x10^{-7}; OR= 0.39
126 [0.26-0.58]) and the replication cohort (p= 2.70e^{-3}, OR=0.62 [0.44-0.87]). After
127 combining both the discovery and replication cohorts, the association of rs76811724
128 exceeded genome-wide significance (p= 2.99 x10^{-9}; OR= 0.49 [0.38-0.64]) (**Fig. 1F,**
129 **Supplementary Table 3**), providing further evidence for an association of *FSTL5* with
130 iTEV.
131
132 **Embryonic expression of *Fstl5***

133 Little is known regarding the expression of *Fstl5* in the hindlimb, and the top-
134 associated rs76973778 was not positioned within an annotated regulatory element
135 (**Supplementary Figure 3A**) and was not previously annotated as an expression
136 quantitative trait locus in human tissues. Therefore to assess whether Fstl5 may regulate
137 hindlimb development *in vivo*, we first examined *Fstl5* expression in multiple diverse
138 species. In the developing embryonic bat limb, an emerging species to study conserved
139 regulatory mechanisms of limb development (19), expression profiling showed *Fstl5* was
140 expressed throughout multiple developmental timepoints, albeit at lower levels compared
141 to *Fstl1* (**Fig. 2A**). In developing chick embryos, *Fstl5* expression was restricted to the
142 brain early, then was later present in the developing limbs (**Fig. 2B**).
143 In mice, RNA-seq analysis of embryonic limb buds detected increasing *Fstl5*
144 expression from E12.5 to E13.5 stages of development (**Fig. 2C**). We then evaluated
145 expression of multiple *Fstl* genes in mouse E15.5 forelimb and hindlimb by RT-qPCR.
146 *Fstl1* had the highest expression; however, only *Fstl5* was differentially expressed
147 between forelimb and hindlimb, with significantly higher expression in embryonic mouse
148 hindlimb (**Fig. 2D, Supplementary Figure 2**). In all assays, *Fstl5* was expressed at lower
149 levels than *Fstl1*, which is strongly expressed throughout the entire limb (20). Therefore,
150 we hypothesized *Fstl5* expression may either be expressed at low levels throughout the
151 entire developing hindlimb or show a spatially restricted expression pattern, leading to
152 overall low levels of expression. *Fstl5* mRNA expression was detected by *in situ*
153 hybridization in multiple tissues of the developing E15.5 mouse, including the limb,
154 forebrain, and heart (**Fig. 2E**). Closer examination of the mouse hindlimbs showed *Fstl5*
155 was strongly and specifically expressed within the condensing cartilage anlage, the

156 primary source for chondrocytes during endochondral ossification (**Fig. 2F**). Together,
157 these results suggest a potential role for *Fstl5* during later stages of embryonic limb
158 development, following specification and outgrowth.
159
160 ***Fstl5* is expressed in the post-natal growth plate**
161 To evaluate post-natal expression of *Fstl5 in vivo*, we utilized Sprague-Dawley rats
162 harboring a beta-galactosidase insertion (*Fstl5*$^{\beta geo}$) within intron 3 of *Fstl5* that resulted in
163 a premature truncation and significantly reduced expression (**Fig. 3A,B)**. Heterozygous
164 *Fstl5*$^{+/\beta geo}$ and knock-out *Fstl5*$^{\beta geo/\beta geo}$ rats were born at expected Mendelian (p= 0.38)
165 and male:female (Binomial p=0.82) ratios with no obvious skeletal or other
166 developmental abnormalities (**Supplementary Tables 4, 5**). Using the endogenous
167 *Fstl5*$^{\beta geo}$ reporter, we found □-galactosidase expression was restricted to the pre-
168 hypertrophic chondrocytes of the femur growth plate and, to a lesser extent, in
169 hypertrophic chondrocytes in 2-week old rats (**Fig. 3C**). During endochondral
170 ossification, the embryonic cartilage condensate gives rise to chondrocytes whose
171 proliferation and maturation in the growth plate regulate subsequent bone formation (21,
172 22). Thus, the restricted expression of *Fstl5* in the embryonic cartilage condensate and
173 post-natal pre-hypertrophic chondrocytes as well as the lack of gross skeletal
174 malformations in *Fstl5*$^{\beta geo/\beta geo}$ rats suggest a non-essential role for *Fstl5* in chondrocyte
175 maturation (23).
176
177 **Sexually dimorphic skeletal development in *Prrx1*-cre;*Fstl5*^{LSL} mice**

178 We sought to test the impact of conditional *Fstl5* over-expression in the
179 osteochondral cell lineage. Conditional over-expression of *Fstl5* was tested in mice for
180 two reasons. First, loss of *Fstl5* expression did not result in any observable skeletal
181 phenotype in *Fstl5*$^{\beta geo/\beta geo}$ rats. Second, the orthologous position of the clubfoot-
182 associated rs76973778 allele was within a short segment deleted from the mouse genome,
183 preventing generation of allele-specific knock-in mice (**Supplementary Figure 3A**).
184 Therefore, we engineered *Fstl5*^{LSL} mice with the *Fstl5* coding sequence under the control
185 of a CAG promoter flanked by LoxP sites (**Fig. 3D**). The *Fstl5*^{LSL} expression cassette
186 was introduced into the ROSA locus using CRISPR/Cas9, and resultant founders were
187 out-crossed and subsequently inter-crossed with *Prrx1*-cre mice to generate heterozygous
188 *Prrx1*-cre;*Fstl5*$^{+/LSL}$ breeders. Heterozygous mice were crossed to produce control
189 (without *Prrx1*-cre or *Fstl5*$^{+/+}$), heterozygous, and homozygous (*Prrx1*-cre;*Fstl5*$^{LSL/LSL}$)
190 mice. To confirm Cre-dependent over-expression of *Fstl5*, bone marrow from 5-month
191 old mice were flushed, and plastic-adherent bone stromal cells (BSCs) were expanded in
192 culture and tested for Cre-dependent *Fstl5* expression. *Fstl5* was not detected in control
193 BSCs; however, Cre-dependent *Fstl5* was robustly expressed in BSCs from *Prrx1*-
194 cre;*Fstl5*^{LSL} mice (**Fig. 3E,F, Supplementary Figure 3B**). *Prrx1*-cre;*Fstl5*^{LSL} mice
195 showed no overt malformations and were visually indistinguishable from litter-mate
196 control mice. We evaluated skeletal development using X-ray and Dual Energy X-Ray
197 Absorptiometry (DEXA) imaging. No differences in tibia and femur length or bone
198 mineral density (BMD) were evident in 3-month old female *Prrx1*-*Fstl5*^{LSL} mice
199 compared to control (**Fig. 3G**). In contrast, tibia and femur lengths were slightly, though
200 significantly, reduced in 3-month old male *Prrx1*-cre;*Fstl5*^{LSL} mice compared to control

201 (**Fig. 3H**). Additionally, both tibia and femur BMD were significantly reduced in male
202 *Prrx1*-cre;*Fstl5*^{LSL} mice compared to control (**Fig. 3H**).
203
204 ***Fstl5*-expressing cell lineages are enriched in neuropathic gene signatures**
205 As no overt limb malformations were evident in *Fstl5*$^{\beta geo}$ knock-out rats or in
206 *Prrx1*-cre;*Fstl5*^{LSL} conditional over-expression mice, we sought putative functional
207 pathways through which Fstl5 may influence hindlimb morphogenesis and development.
208 For this, we evaluated *Fstl5* expression using available mouse embryonic single-cell
209 transcriptomes (24) (**Fig. 4A**). Consistent with our *in situ* expression results (**Fig. 2E**),
210 highest levels of *Fstl5* expression were evident in neuronal cell lineages, with lower
211 expression in skeletal lineages, such as "*Limb Mesenchyme*", "*Osteoblast*", "*Chondrocyte*
212 *Progenitor*", and "*Chondrocyte/Osteoblast*" (**Supplementary Figure 4**). For each
213 skeletal lineage, expression of *Fstl5* was significantly enriched in a single sub-cluster of
214 cells (**Fig. 4A**). As the transcriptional landscape of lineage sub-clusters may help to
215 broadly inform future cell fates (25), we extracted all genes with significantly enriched
216 expression within *Fstl5*-expressing sub-clusters and performed pathway analyses. *Fstl5*-
217 expressing sub-clusters were significantly enriched for genes involved in neural
218 development and synaptic transmission (**Fig. 4B**). Neural system genes were significantly
219 enriched in the *Fstl5*-expressing sub-clusters of the *Chondrocyte progenitor* ($p=6.60e^{-8}$),
220 *Chondrocyte and Osteoblast* ($p=3.46e^{-10}$), and *Osteoblast* ($p=1.71e^{-11}$) cell lineages.
221 These results implicate a potential role for *Fstl5* in the neuropathogenesis of iTEV.
222
223 **Lineage-specific gene expression suggests functional differences in Fstl proteins**

224 Fstl proteins were originally named for having follistatin-like domains, suggesting
225 potential roles in regulating Tgf-ß and BMP signaling, as has been described for Fstl1
226 (26-28). However, other Fstl family members are less well characterized. Prior
227 phylogenetic protein analyses identified sequence similarities between Fstl4 and Fstl5,
228 which clustered most closely with Fstl1, while Fstl3 and Follistatin (Fst) were both
229 unique (20, 29, 30). We queried expression of *Fst* and *Fstl* genes across the 38 mouse
230 embryonic cell lineages defined by single-cell sequencing of mouse embryos (24).
231 Surprisingly, hierarchical clustering distinguished patterns of expression between *Fstl*
232 genes that resembled phylogenetic analyses based on protein structure (20, 30) (**Fig. 4C**).
233 Both *Fstl4* and *Fstl5* clustered together with highest expression in neural cell lineages.
234 Consistent with skeletal abnormalities in *Fstl1* knock-out mice (27), *Fstl1* expression was
235 highest in mesenchymal and skeletal cell lineages. *Fstl3* showed highest expression in the
236 cardiac cell lineage, among others, and mice lacking *Fstl3* were previously shown to
237 develop cardiac abnormalities resulting in hypertension (31). These results suggest
238 lineage-specific embryonic expression of *Fstl* genes in mice correlates with proposed
239 functional differences among the Fstl family of proteins.
240
241 **DISCUSSION**
242 We identified a locus within a previously uncharacterized gene, *FSTL5*,
243 associated with isolated clubfoot. Analysis of *Fstl5* in multiple vertebrate embryos
244 identified conserved expression in the embryonic hindlimb mesenchyme and neural
245 tissues. Conditional over-expression of *Fstl5* in the embryonic limb mesenchyme (using
246 *Prrx1*-cre) resulted in reduced BMD in male, but not female, mice, though no gross

247 malformations were evident. Finally, we show that evolutionary divergence in the
248 essential roles of Fstl proteins resembles differences in their lineage-specific expression
249 in developing mouse embryos (24).
250 Follistatin-like 5, the protein encoded by *FSTL5*, is a member of the secreted
251 protein acidic rich in cysteines (SPARC) family of proteins (30). Follistatin is a secreted
252 TGF-ß and BMP antagonist, and because Fstl1 harbors a domain with modest sequence
253 homology to Follistatin, Fstl proteins were hypothesized to also regulate TGF-ß and BMP
254 signaling (20, 32, 33). Consistent with this, *Fstl1* knock-out mice developed lung and
255 skeletal defects and died shortly after birth from respiratory insufficiency (26, 27).
256 Skeletal abnormalities included defects in axial skeleton patterning, long bone dysplasia,
257 and a hindfoot rotational deformity caused by malposition of the distal fibula (27).
258 Moreover, *Fstl1* was shown to promote chondrogenic lineage differentiation of
259 mesenchymal stromal cells *ex vivo* (34). However, functional similarities between Fstl1
260 and other Fstl proteins, such as Fstl5, was unclear.
261 While the function of FSTL5 was largely unstudied, its mouse homolog (Fstl5)
262 shared highest sequence similarity to Fstl4, another uncharacterized protein, and Fstl1
263 (20). *Fstl4* expression was localized to neural tissues, and though neurologic differences
264 were observed in *Fstl4* knock-out mice, a lack of gross developmental deformities
265 suggested Fstl4 likely does not regulate BMP signaling (35, 36). Like *Fstl4*, *Fstl5* was
266 expressed more highly in neural tissues, and rats lacking *Fstl5* showed no gross
267 developmental malformations. Our results suggest that Fstl5, like Fstl4, does not
268 modulate Tgf-ß and BMP signaling and is not essential for embryonic skeletal
269 development, unlike Fst and Fstl1 (26-28, 37). Moreover, lineage-specific expression of

270 *Fstl* genes in mouse embryos is supported by the functional differences between Fstl
271 proteins and differences in developmental phenotypes among genetically-engineered
272 mouse models. Expression levels of both *Fstl4* and *Fstl5* were highest in neural tissues,
273 while *Fstl3* and *Fstl1* were expressed highest in cardiac and osteochondroprogenitor
274 lineages, respectively.

275 Taken together, these results implicate a neurogenic mechanism and potential
276 sexually dimorphic role for *FSTL5* in iTEV pathophysiology. A neuropathologic
277 mechanism for clubfoot development is further supported by the recent characterization
278 of a mouse model of peroneal muscular atrophy (14). Regarding *Fstl5*, additional
279 experiments are required to understand what role it plays in the nervous system and how
280 alterations in Fstl5 function or expression may contribute to limb deformity, potentially
281 through the use of the conditional transgenic model reported here. As conditional over-
282 expression in skeletal progenitor cells did not result in significant limb deformity, the
283 impact on limb development following over-expression in neuro-lineage cells remains to
284 be tested. Finally, analysis of *FSTL5* expression in human nervous system tissues is
285 warranted; however, the correlation to clubfoot pathogenesis may be limited since infants
286 with isolated clubfoot mostly undergo non-operative treatments. Our results implicate a
287 neurogenetic pathology to clubfoot pathogenesis.
288

289 MATERIALS AND METHODS

290 Cohort descriptions

291 For the discovery cohort analyses, non-hispanic white clubfoot subjects were
292 recruited from St Louis Children's Hospital and St Louis Shriners Hospital. The study

293 protocol was approved by the Institutional Review Board, and all subjects and/or parents
294 gave informed consent. Patients were diagnosed at infancy with talipes equinovarus
295 (clubfoot) based on the physical examination findings by a single orthopedic surgeon.
296 Exclusion criteria included additional congenital anomalies, developmental delay, or
297 known underlying etiologies such as arthrogryposis, myelomeningocele, or myopathy.
298 Control subjects consisted of an ethnicity-matched subset of the Atherosclerosis Risk in
299 Communities study (ARIC) (18). The ARIC cohort is a prospective community-based
300 recruitment study focused primarily on atherosclerosis and cardiovascular outcomes.
301 Ethnicity was confirmed for both cohorts by principal component analysis compared to
302 HapMap (**Supplementary Figure 1A**).

303 For replication analyses, non-hispanic white clubfoot subjects were recruited from
304 Scottish Rite for Children (Dallas, TX), Shriners Hospital for Children (Houston, TX),
305 the St. Louis Children's Hospital (St. Louis, MO), and St. Louis Shriners Hospital (St.
306 Louis, MO). All subjects provided written informed consent approved by their respective
307 Institutional Review Boards. Inclusion and exclusion criteria were the same as for the
308 discovery cohort. Ethnicity-matched control subjects used for replication were recruited,
309 in part, from Scottish Rite for Children (Dallas, TX). Additional control subjects were
310 included from the Dallas Heart Study (38), a multi-ethnic population-based probability
311 sampling of Dallas County (Dallas, TX), and from a randomized trial with work-place
312 recruitment from the Dallas-Fort Worth metroplex (39). Recruitment was unrelated to
313 clubfoot. The allele frequency of the rs76811724 variant among control subjects of the
314 replication cohort was similar to the allele frequency among non-hispanic whites from the
315 Gnomad consortium (frequency=0.08).

316

317 Genome-wide imputation and association

318 Genome-wide genotyping of iTEV subjects was performed using the Affymetrix

319 6.0 microarray as previously described (17). Control subjects available from the

320 population-based Atherosclerosis Risk in Communities were genotyped on the

321 Affymetrix 6.0 microarray, and genotype data were obtained from dbGAP [dbGAP:

322 phs000280]. Quality-control was performed to exclude individuals with evidence of

323 chromosomal abnormalities, gender inconsistencies, and excessive genotype missingness.

324 Case and control datasets were merged and genotypes harmonized using Plink v1.9 (40).

325 Duplicate and related samples were identified using a subset of linkage disequilibrium

326 (LD)-pruned variants, and only unrelated individuals were included in the imputation.

327 LD-pruned variants were used for principal component analysis including HapMap Phase

328 3 samples for ethnicity verification (**Supplementary Figure 1A**). Prior to imputation,

329 individual single nucleotide polymorphisms (SNPs) were excluded based on missingness

330 rate (>5%), minor allele frequency (<0.01), deviation from Hardy-Weinberg equilibrium

331 (HWE $p<1x10^{-4}$ in controls), non-variant monomorphic SNPs, and SNPs with a

332 differential missingness rate between case and control cohorts. After filtering, 549,589

333 autosomal SNPs were available for imputation.

334 Imputation was performed with MACH v1.0 (41) and Minimac (42) on the

335 merged case-control cohort following pre-phasing with the 1000 Genomes Phase 3

336 reference panel. Imputation accuracy (Rsq) was investigated for different ranges of allele

337 frequency, and a marked decline in the percent of high-quality imputed variants was

338 evident for those with minor allele frequency <2% (**Supplementary Table 6,**

339 **Supplementary Figure 1B**). As a majority of imputed variants were rare or
340 monomorphic and thus poorly imputed, only variants with frequency >2%, Rsq >0.3, and
341 without significant HWE deviation were included for association analysis. Following
342 post-imputation filtering, imputed dosages of 7,794,536 variants were included for
343 logistic regression analysis with *Mach2dat*, which included gender and ten principle
344 components as covariates. Association results were plotted using R and LocusZoom (43).

345 Replication genotyping was performed using allele-discrimination assays
346 (ThermoFisher). The top-associated variant (rs76973778) was genotyped using a
347 commercially-available assay; however, the commercial assay was unable to accurately
348 discriminate between genotypes. Therefore, a custom allele-discrimination assay was
349 designed to genotype the highly correlated rs76811724 (**Supplementary Table 7**). The
350 replication cohort consisted of ethnicity-matched subjects with isolated clubfoot recruited
351 at St. Louis Children's Hospital, Shriners Hospital-St. Louis, Scottish Rite for Children,
352 and Shriners Hospital for Children-Houston. All subjects provided written informed
353 consent approved by the respective Institutional Review Boards. Genotypes were merged
354 with the discovery cohort, and association analysis performed using logistic regression
355 analysis including gender as a covariate.
356
357 Production of *Fstl5*$^{\beta geo}$ rats

358 Generation of *Fstl5*$^{\beta geo}$ rats were previously described (44, 45). Briefly, *Sleeping*
359 *Beauty β-Geo* trap transposons (46) were used to select mutant rat spermatogonial
360 libraries *in vitro* (44). Spermatogonia comprising a selected library were then
361 transplanted into rat testes for production of mutant spermatozoa (47). Recipient males

362 were bred with wildtype females to produce a random panel of donor spermatozoa-
363 derived mutant rat strains enriched with gene traps in protein coding genes (44). Genomic
364 sites of transposon integration were defined in the newly generated mutant rats by
365 splinklerette PCR (44) with sequence analysis alignment on genome build RGSC v3.4
366 (Rn4).
367 For this study, testis cells from a rat harboring a *Sleeping Beauty β-Geo* trap
368 transposon in intron 3 of *Fstl5* (44) were thawed from cryopreservation and used to
369 derive a donor spermatogonial stem cell line that was transplanted into recipient males, as
370 described (48). Recipient males were bred with wildtype females to regenerate the *Fstl5*
371 mutant rat strain. Two founders were established from cryopreservation, both harboring
372 the *Fstl5* intron 3 *β-Geo* trap transposon. *Fstl5* gene-specific PCR primers near *Sleeping*
373 *Beauty* integration sites were used in combination with transposon-specific primers to
374 genotype progeny (**Supplementary Table 8**). *Fstl5* mutant rats were housed in
375 individually ventilated, Lab Products 2100 cages in a dedicated room with atmosphere
376 controls set to 72°F, 45-50% humidity during a 12 hour light/dark cycle (i.e. Light cycle
377 = 6:00am-6:00pm, Central Standard Time adjusted for daylight savings time). Rats were
378 fed Harlan Teklad Irradiated 7912, LM-485 Mouse/Rat Diet, 5% fat Diet with a
379 continuous supply of reverse osmosis water. All rat protocols were approved by the
380 University of Texas Southwestern Medical Center.
381
382 Expression analysis
383 For analysis of $Fstl5^{\beta geo}$ rats, brain tissue was harvested from 3-week old rats and
384 homogenized in 500µl TRIzol reagent (Invitrogen). Samples were vortexed and incubated

385 at room temperature for 5 minutes and 100µl chloroform added. Samples were vortexed
386 and centrifuged at 12,000g for 15 minutes at 4°C. RNA was precipitated using 250µl
387 isopropyl alcohol and incubated at room temperature for 10 minutes followed by
388 centrifugation at 12,000g for 10 minutes at 4°C.The RNA pellet was subsequently
389 washed twice with 500µl 75% ethanol followed by centrifugation at 7,500g for 5 minutes
390 at 4°C. RNA was rehydrated in 50µl nuclease free water and concentrations estimated
391 using a Nanodrop spectrophotometer. *Fstl5* expression was assayed using Sybr-green
392 qPCR together with *Gapdh* for normalization. Relative expression levels were compared
393 using one-sided t-tests compared to *Fstl5*$^{+/+}$ control. Rat qPCR primer sequences are
394 listed in **Supplementary Table 8**. Experimental protocols were approved by the
395 University of Texas Southwestern Medical Center.
396 For mouse *Fstl* limb bud expression, forelimb and hindlimb buds were dissected
397 from E15.5 embryos, RNA extracted using PureLink RNA Mini kit (Ambion), and
398 samples pooled (1 litter per pool). cDNA was generated using iScript (BioRad) and qPCR
399 performed using SybrSelect Mastermix (ThermoFisher). Assays were run in triplicate and
400 expression normalized to *ActinB*. Relative expression levels were compared using two-
401 sided t-tests. Mouse qPCR primer sequences are listed in **Supplementary Table 8**.
402 Experimental procedures were approved by the Otago University Animal Ethics
403 committee.
404
405 <u>Immunohistochemistry</u>
406 Femurs were dissected from 2-week old *Fstl5*$^{\beta geo}$ rats and fixed in 10% formalin
407 followed by decalcification in 14% EDTA in PBS and embedded in paraffin. Four-

408 micron sections were cut and placed on 3-aminopropyltriethoxysaline (APES)-treated
409 slides. Slides were dried at 37°C, deparaffinized, and digested in 0.25% Trypsin in 1mM
410 EDTA for 3 minutes at 37°C incubator. Endogenous peroxidase activity was deactivated
411 using 3% hydrogen peroxide in Methanol for 10 minutes at room temperature and rinsed
412 in PBS with 0.1 % Tween 20 (PBST). Sections were blocked in 10% Normal Goat Serum
413 (Dako) for 30 minutes at room temperature and incubated overnight at 4°C with Anti-
414 Beta Galactosidase Mouse antibody (Promega, 1:500). Following, slides were washed in
415 PBST 3 times and incubated with HRP-conjugated Goat Anti Mouse IgG Antibody,
416 (Millipore) for 1.5 hours and washed in PBST 3x. The DAB substrate –chromogen
417 (DAKO) was applied for 10 minutes, counterstained in Mayer's hematoxylin for 2
418 minutes, and washed in distilled water. Sections were dehydrated in alcohol, cleared in
419 xylene, and mounted with Cytoseal XYL (Thermo Scientific) prior to imaging with a
420 DP73 Olympus microscope using Cell Sens software.
421
422 *In situ* hybridization
423 A 505 bp region of the mouse *Fstl5* mRNA was amplified from embryonic E15.5
424 hindlimb cDNA using oligonucleotide primers (Fstl5F:5'-
425 GGAAGGCTAAGCTCTGCATATT-3', and Fstl5R:5'-
426 GCACTACAGAGAGTGGTTTCAG-3'). The PCR product was cloned into
427 pGEMTEasy (Promega) and sequenced to confirm orientation. Dioxigenin (DIG)-
428 labelled probe synthesis (sense and anti-sense) was carried out as published previously
429 (49).

430 Embryos from time-mated mice were removed and frozen in OCT prior to
431 sectioning (10 μm) for *in situ* hybridization. Thawed sections were air-dried and fixed
432 with 4% paraformaldehyde (PFA)/phosphate buffered saline (PBS) for 10 min, washed
433 with PBT (PBS with 10% Tween-20) three times before proteinase K digestion (1 μg/ml
434 for 10 min). Following, sections were re-fixed with PFA/PBS and acetylated (625 μl
435 acetic anhydride and 295 ml 0.1 M triethanolamine) for 10 min at room temperature.
436 Sections were washed three times with PBT prior to hybridization with the appropriate
437 sense or anti-sense probe in hybridization buffer (50% formamide, 5x saline sodium
438 citrate buffer (SSC), 5x Denhardts, 250 μg/ml yeast RNA, 500 μg/ml herring sperm
439 DNA). Slides were incubated overnight at 65°C in a humidified chamber. Post-
440 hybridization washes were done at 65°C with a series of wash buffers (1xSSC/50%
441 formamide; 2xSSC and 0.2x SSC (twice)) for 20 min each. Sections were washed twice
442 at room temperature with maleic acid buffer with Tween 20 (MABT) before blocking
443 with 20% heat inactivated sheep serum (HISS), 2 % Roche blocking powder in MABT.
444 To detect the DIG-probe, slides were incubated overnight at 4°C with anti-DIG-alkaline
445 phosphatase (Roche) diluted 1:2,500 in 5% HISS/MABT. Unbound antibody was
446 removed by washing with MABT, followed by NTM buffer (100mM NaCl, 100mM
447 Tris.Cl (pH 9.5), 50mM $MgCl_2$) with nitro blue tretrazolium (NBT) and 5-bromo-4-
448 chloro-3-indolyl-phostphate (BCIP) (Roche) to detect alkaline phosphatase. Following
449 the color reaction, slides were washed and fixed with 4% PFA/PBS before mounting for
450 imaging. Sections were imaged on an Olympus AX70 light microscope. All mouse
451 experimental protocols were approved by the Otago University Animal Ethics committee.

452 Chicken *Fstl5* was cloned from a cDNA library from day 4 chicken embryos
453 using PCR primers and subcloned into a pBS vector. In-situ hybridization analysis was
454 carried out using previously published methods (50, 51). Experimental procedures were
455 performed in accordance with UK Home Office Animal licensing and in accordance with
456 University of Aberdeen ethical review committees.
457
458 Generation of the *Fstl5*^{LSL} mouse line
459 A vector containing the CAG promoter flanked by LoxP sites upstream of the
460 Fstl5 coding sequence was designed, produced, and purified using VectorBuilder
461 (https://en.vectorbuilder.com) and Cyagen US (Santa Clara, CA). Mice were engineered
462 using CRISPR/Cas9 via pronuclear injection of 1-cell stage C57BL/6N embryos.
463 Injections included 25ng/uL Cas9 protein (Integrated DNA Technologies), 25ng/uL
464 crRNA (5'-CGCCCATCTTCTAGAAAGAC-3') complexed with tracRNA scaffold, and
465 10ng/uL circular plasmid, diluted in TE buffer. Following injection, embryos were
466 implanted into pseudo-pregnant females. Multiple founder lines were established and
467 maintained by out-crossing to C57BL/6J mice.
468 *Fstl5*^{LSL} mice were crossed to *Prrx1*-cre mice, and resulting *Prrx1*-cre;*Fstl5*$^{+/LSL}$
469 mice were crossed to generate control (no *Prrx1*-cre or *Fstl5*$^{+/+}$), *Prrx1*-cre;*Fstl5*$^{+/LSL}$
470 heterozygous, and *Prrx1*-cre;*Fstl5*$^{LSL/LSL}$ homozygous mice. *Cre* and *Fstl5*^{LSL} genotyping
471 primer sequences are provided in **Supplementary Table 8**. Cre-induced over-expression
472 was validated by qPCR and Western blotting using cultured bone stromal cells from
473 adults *Prrx1*-cre;*Fstl5*^{LSL} mice compared to control.

474 Mouse protocols were approved by the UT Southwestern Medical Center
475 Institutional Animal Care and Use Committee.
476
477 Cell culture, expression analysis, and western blotting
478 For *ex vivo* bone stromal cell culture, bone marrow was flushed from adult mice
479 and cultured in alpha-MEM (Thermofisher Scientific) with 10% FBS (Sigma) and 1%
480 antibiotics (penicillin/streptomycin) (Thermofisher Scientific) for 5 days at 37°C and 5%
481 CO_2. Following, media was refreshed and non-plastic adherent cells washed away. The
482 cells were then plated in 12-well plates and 100mm plates for RNA and protein
483 extraction, respectively and grown to confluence.
484 For RNA extraction, the cells were harvested in RLT Plus buffer with beta-
485 marcaptoethanol (Sigma). The cells were passed through Qiashredder columns (Qiagen)
486 prior to RNA extraction. RNeasy Plus Mini Kit (Qiagen) was used to extract RNA
487 following manufacturer's recommendations. The samples were quantified using a
488 Nanodrop1000 spectrophotometer (Thermofisher Scientific). cDNA synthesis was
489 performed using the High Capacity RNA to cDNA Kit (Thermofisher Scientific). The
490 qPCR assays were performed using SYBR Green PCR Master Mix (Thermofisher
491 Scientific). Protein was extracted using RIPA buffer with protease inhibitor cocktail
492 (Roche) and concentration was determined using the Pierce BCA Protein Assay
493 (Thermofisher Scientific). Approximately 10µg of protein was loaded on SDS-PAGE.
494 The antibody detection was performed using rabbit anti-Fstl5 (1:2,000) and rabbit anti-
495 Actb (1:5,000) (Cell Signaling Technologies) in 5% BSA in TBST buffer overnight at
496 4°C. After washing, goat anti-rabbit (1:10,000) IRDye secondary antibody (Licor

497 Biosciences) was incubated in 5% nonfat dry milk in TBST buffer for 1 hour at room
498 temperature in the dark. Following washing, images were acquired using the Odyssey
499 CLx system (Licor Biosciences).
500
501 Skeletal assessments
502 For live-animal rodent imaging, mice were anesthetized using 2% inhaled
503 isoflurane. Images were acquired using a Faxitron UltrafocusDXA. Bone densities and
504 bone lengths were measured using Faxitron software and compared to litter-mate controls
505 using two-sided t-tests.

506

507 **ACKNOWLEDGEMENTS**
508 The Atherosclerosis Risk in Communities Study is carried out as a collaborative
509 study supported by National Heart, Lung, and Blood Institute contracts
510 (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C,
511 HHSN268201100008C, HHSN268201100009C, HHSN268201100010C,
512 HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and
513 participants of the ARIC study for their important contributions. Funding for GENEVA
514 was provided by National Human Genome Research Institute grant U01HG004402 (E.
515 Boerwinkle). We thank H. Hobbs and J. Cohen for contributing control samples for
516 replication genotyping, Nadav Ahituv for sharing RNA-seq data for both bat and mouse
517 embryonic limb buds, Tommy Hyatt for designing the custom genotyping assay, and
518 members of the UT Southwestern Transgenic Core facility, including John Ritter, Mylinh
519 Nguyen, and Robert Hammer. Publicly available mouse embryonic expression analysis

results were provided online at https://oncoscape.v3.sttrcancer.org/atlas.gs.washington.edu.mouse.rna/landing (24). The authors acknowledge the contributions and support of the Center for Excellence in Clubfoot Research at Scottish Rite for Children, including Shawne Faulks and Kristhen Atala. *Fstl5* mutant rats were produced by the NIH Mutant Rat Resource at UT Southwestern Medical Center (R24RR03232601, R24OD011108, R01HD036022, and 5R01HD053889). This study was supported by funding from the Scottish Rite for Children Research Fund (J.J.R.), Shriners Hospital for Children (J.T.H), and the National Institutes of Health award R01HD043342 (J.T.H.).

CONFLICTS OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

REFERENCES

1 Wynne-Davies, R. (1972) Genetic and environmental factors in the etiology of talipes equinovarus. *Clin Orthop Relat Res*, **84**, 9-13.

2 Siebert, M.J., Karacz, C.M. and Richards, B.S. (2020) Successful Ponseti-treated Clubfeet at Age 2 Years: What is the Rate of Surgical Intervention After This? *J Pediatr Orthop*, in press.

3 Faulks, S. and Richards, B.S. (2009) Clubfoot treatment: Ponseti and French functional methods are equally effective. *Clin Orthop Relat Res*, **467**, 1278-1282.

4 Brewer, C., Holloway, S., Zawalnyski, P., Schinzel, A. and FitzPatrick, D. (1998) A chromosomal deletion map of human malformations. *Am J Hum Genet*, **63**, 1153-1159.

5 Werler, M.M., Yazdy, M.M., Mitchell, A.A., Meyer, R.E., Druschel, C.M., Anderka, M., Kasser, J.R. and Mahan, S.T. (2013) Descriptive epidemiology of idiopathic clubfoot. *Am J Med Genet A*, **161A**, 1569-1578.

6 Lochmiller, C., Johnston, D., Scott, A., Risman, M. and Hecht, J.T. (1998) Genetic epidemiology study of idiopathic talipes equinovarus. *Am J Med Genet*, **79**, 90-96.

549 7 Beals, R.K. (1978) Club foot in the Maori: a genetic study of 50 kindreds. *N Z*
550 *Med J*, **88**, 144-146.
551 8 Gurnett, C.A., Alaee, F., Kruse, L.M., Desruisseau, D.M., Hecht, J.T., Wise, C.A.,
552 Bowcock, A.M. and Dobbs, M.B. (2008) Asymmetric lower-limb malformations in
553 individuals with homeobox PITX1 gene mutation. *Am J Hum Genet*, **83**, 616-622.
554 9 Alvarado, D.M., McCall, K., Aferol, H., Silva, M.J., Garbow, J.R., Spees, W.M.,
555 Patel, T., Siegel, M., Dobbs, M.B. and Gurnett, C.A. (2011) Pitx1 haploinsufficiency
556 causes clubfoot in humans and a clubfoot-like phenotype in mice. *Hum Mol Genet*, **20**,
557 3943-3952.
558 10 Lu, W., Bacino, C.A., Richards, B.S., Alvarez, C., VanderMeer, J.E., Vella, M.,
559 Ahituv, N., Sikka, N., Dietz, F.R., Blanton, S.H. *et al.* (2012) Studies of TBX4 and
560 chromosome 17q23.1q23.2: an uncommon cause of nonsyndromic clubfoot. *Am J Med*
561 *Genet A*, **158A**, 1620-1627.
562 11 Alvarado, D.M., Aferol, H., McCall, K., Huang, J.B., Techy, M., Buchan, J.,
563 Cady, J., Gonzales, P.R., Dobbs, M.B. and Gurnett, C.A. (2010) Familial isolated
564 clubfoot is associated with recurrent chromosome 17q23.1q23.2 microduplications
565 containing TBX4. *Am J Hum Genet*, **87**, 154-160.
566 12 Alvarado, D.M., Buchan, J.G., Frick, S.L., Herzenberg, J.E., Dobbs, M.B. and
567 Gurnett, C.A. (2013) Copy number analysis of 413 isolated talipes equinovarus patients
568 suggests role for transcriptional regulators of early limb development. *Eur J Hum Genet*,
569 **21**, 373-380.
570 13 Spielmann, M., Brancati, F., Krawitz, P.M., Robinson, P.N., Ibrahim, D.M.,
571 Franke, M., Hecht, J., Lohan, S., Dathe, K., Nardone, A.M. *et al.* (2012) Homeotic arm-
572 to-leg transformation associated with genomic rearrangements at the PITX1 locus. *Am J*
573 *Hum Genet*, **91**, 629-635.
574 14 Collinson, J.M., Lindstrom, N.O., Neves, C., Wallace, K., Meharg, C., Charles,
575 R.H., Ross, Z.K., Fraser, A.M., Mbogo, I., Oras, K. *et al.* (2018) The developmental and
576 genetic basis of 'clubfoot' in the peroneal muscular atrophy mutant mouse. *Development*,
577 **145**.
578 15 Helmbacher, F., Schneider-Maunoury, S., Topilko, P., Tiret, L. and Charnay, P.
579 (2000) Targeting of the EphA4 tyrosine kinase receptor affects dorsal/ventral pathfinding
580 of limb motor axons. *Development*, **127**, 3313-3324.
581 16 Dobbs, M.B. and Gurnett, C.A. (2012) Genetics of clubfoot. *J Pediatr Orthop B*,
582 **21**, 7-9.
583 17 Zhang, T.X., Haller, G., Lin, P., Alvarado, D.M., Hecht, J.T., Blanton, S.H.,
584 Stephens Richards, B., Rice, J.P., Dobbs, M.B. and Gurnett, C.A. (2014) Genome-wide
585 association study identifies new disease loci for isolated clubfoot. *J Med Genet*, **51**, 334-
586 339.
587 18 investigators, T.A. (1989) The Atherosclerosis Risk in Communities (ARIC)
588 Study: design and objectives. The ARIC investigators. *American journal of*
589 *epidemiology*, **129**, 687-702.
590 19 Eckalbar, W.L., Schlebusch, S.A., Mason, M.K., Gill, Z., Parker, A.V., Booker,
591 B.M., Nishizaki, S., Muswamba-Nday, C., Terhune, E., Nevonen, K.A. *et al.* (2016)
592 Transcriptomic and epigenomic characterization of the developing bat wing. *Nat Genet*,
593 **48**, 528-536.

594 20 Adams, D., Larman, B. and Oxburgh, L. (2007) Developmental expression of
595 mouse Follistatin-like 1 (Fstl1): Dynamic regulation during organogenesis of the kidney
596 and lung. *Gene Expr Patterns*, **7**, 491-500.
597 21 Karsenty, G. (2003) The complexities of skeletal biology. *Nature*, **423**, 316-318.
598 22 Kronenberg, H.M. (2003) Developmental regulation of the growth plate. *Nature*,
599 **423**, 332-336.
600 23 Kozhemyakina, E., Lassar, A.B. and Zelzer, E. (2015) A pathway to bone:
601 signaling molecules and transcription factors involved in chondrocyte development and
602 maturation. *Development*, **142**, 817-831.
603 24 Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F.,
604 Mundlos, S., Christiansen, L., Steemers, F.J. *et al.* (2019) The single-cell transcriptional
605 landscape of mammalian organogenesis. *Nature*, **566**, 496-502.
606 25 Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D. and Klein, A.M. (2020)
607 Lineage tracing on transcriptional landscapes links state to fate during differentiation.
608 *Science*, **367**.
609 26 Geng, Y., Dong, Y., Yu, M., Zhang, L., Yan, X., Sun, J., Qiao, L., Geng, H.,
610 Nakajima, M., Furuichi, T. *et al.* (2011) Follistatin-like 1 (Fstl1) is a bone morphogenetic
611 protein (BMP) 4 signaling antagonist in controlling mouse lung development. *Proc Natl
612 Acad Sci U S A*, **108**, 7058-7063.
613 27 Sylva, M., Li, V.S., Buffing, A.A., van Es, J.H., van den Born, M., van der
614 Velden, S., Gunst, Q., Koolstra, J.H., Moorman, A.F., Clevers, H. *et al.* (2011) The BMP
615 antagonist follistatin-like 1 is required for skeletal and lung organogenesis. *PLoS One*, **6**,
616 e22616.
617 28 Matzuk, M.M., Lu, N., Vogel, H., Sellheyer, K., Roop, D.R. and Bradley, A.
618 (1995) Multiple defects and perinatal death in mice deficient in follistatin. *Nature*, **374**,
619 360-363.
620 29 Glusman, G., Kaur, A., Hood, L. and Rowen, L. (2004) An enigmatic fourth runt
621 domain gene in the fugu genome: ancestral gene loss versus accelerated evolution. *BMC
622 Evol Biol*, **4**, 43.
623 30 Sylva, M., Moorman, A.F. and van den Hoff, M.J. (2013) Follistatin-like 1 in
624 vertebrate development. *Birth Defects Res C Embryo Today*, **99**, 61-69.
625 31 Mukherjee, A., Sidis, Y., Mahan, A., Raher, M.J., Xia, Y., Rosen, E.D., Bloch,
626 K.D., Thomas, M.K. and Schneyer, A.L. (2007) FSTL3 deletion reveals roles for TGF-
627 beta family ligands in glucose and fat homeostasis in adults. *Proc Natl Acad Sci U S A*,
628 **104**, 1348-1353.
629 32 Keutmann, H.T., Schneyer, A.L. and Sidis, Y. (2004) The role of follistatin
630 domains in follistatin biological action. *Mol Endocrinol*, **18**, 228-240.
631 33 Massague, J. and Chen, Y.G. (2000) Controlling TGF-beta signaling. *Genes Dev*,
632 **14**, 627-644.
633 34 Chaly, Y., Blair, H.C., Smith, S.M., Bushnell, D.S., Marinov, A.D., Campfield,
634 B.T. and Hirsch, R. (2015) Follistatin-like protein 1 regulates chondrocyte proliferation
635 and chondrogenic differentiation of mesenchymal stem cells. *Ann Rheum Dis*, **74**, 1467-
636 1473.
637 35 Yonehara, K., Shintani, T., Suzuki, R., Sakuta, H., Takeuchi, Y., Nakamura-
638 Yonehara, K. and Noda, M. (2008) Expression of SPIG1 reveals development of a retinal

639 ganglion cell subtype projecting to the medial terminal nucleus in the mouse. *PLoS One*,
640 **3**, e1533.
641 36 Suzuki, R., Fujikawa, A., Komatsu, Y., Kuboyama, K., Tanga, N. and Noda, M.
642 (2018) Enhanced extinction of aversive memories in mice lacking SPARC-related protein
643 containing immunoglobulin domains 1 (SPIG1/FSTL4). *Neurobiol Learn Mem*, **152**, 61-
644 70.
645 37 Welt, C., Sidis, Y., Keutmann, H. and Schneyer, A. (2002) Activins, inhibins, and
646 follistatins: from endocrinology to signaling. A paradigm for the new millennium. *Exp*
647 *Biol Med (Maywood)*, **227**, 724-752.
648 38 Victor, R.G., Haley, R.W., Willett, D.L., Peshock, R.M., Vaeth, P.C., Leonard,
649 D., Basit, M., Cooper, R.S., Iannacchione, V.G., Visscher, W.A. *et al.* (2004) The Dallas
650 Heart Study: a population-based probability sample for the multidisciplinary study of
651 ethnic differences in cardiovascular health. *Am J Cardiol*, **93**, 1473-1480.
652 39 Lakoski, S.G., Xu, F., Vega, G.L., Grundy, S.M., Chandalia, M., Lam, C., Lowe,
653 R.S., Stepanavage, M.E., Musliner, T.A., Cohen, J.C. *et al.* (2010) Indices of cholesterol
654 metabolism and relative responsiveness to ezetimibe and simvastatin. *J Clin Endocrinol*
655 *Metab*, **95**, 800-809.
656 40 Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J.
657 (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets.
658 *Gigascience*, **4**, 7.
659 41 Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using
660 sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet*
661 *Epidemiol*, **34**, 816-834.
662 42 Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. and Abecasis, G.R. (2012)
663 Fast and accurate genotype imputation in genome-wide association studies through pre-
664 phasing. *Nat Genet*, **44**, 955-959.
665 43 Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P.,
666 Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization
667 of genome-wide association scan results. *Bioinformatics*, **26**, 2336-2337.
668 44 Izsvak, Z., Frohlich, J., Grabundzija, I., Shirley, J.R., Powell, H.M., Chapman,
669 K.M., Ivics, Z. and Hamra, F.K. (2010) Generating knockout rats by transposon
670 mutagenesis in spermatogonial stem cells. *Nat Methods*, **7**, 443-445.
671 45 Ivics, Z., Izsvak, Z., Medrano, G., Chapman, K.M. and Hamra, F.K. (2011)
672 Sleeping Beauty transposon mutagenesis in rat spermatogonial stem cells. *Nat Protoc*, **6**,
673 1521-1535.
674 46 Ivics, Z., Li, M.A., Mates, L., Boeke, J.D., Nagy, A., Bradley, A. and Izsvak, Z.
675 (2009) Transposon-mediated genome manipulation in vertebrates. *Nat Methods*, **6**, 415-
676 422.
677 47 Hamra, F.K., Gatlin, J., Chapman, K.M., Grellhesl, D.M., Garcia, J.V., Hammer,
678 R.E. and Garbers, D.L. (2002) Production of transgenic rats by lentiviral transduction of
679 male germ-line stem cells. *Proc Natl Acad Sci U S A*, **99**, 14931-14936.
680 48 Wu, Z., Falciatori, I., Molyneux, L.A., Richardson, T.E., Chapman, K.M. and
681 Hamra, F.K. (2009) Spermatogonial culture medium: an effective and efficient nutrient
682 mixture for culturing rat spermatogonial stem cells. *Biology of reproduction*, **81**, 77-86.

683 49 Yang, Y. and Wilson, M.J. (2015) Lhx9 gene expression during early limb
684 development in mice requires the FGF signalling pathway. *Gene Expr Patterns*, **19**, 45-
685 51.
686 50 Mahony, C. and Vargesson, N. (2013) Molecular analysis of regulative events in
687 the developing chick limb. *J Anat*, **223**, 1-13.
688 51 Vargesson, N. and Laufer, E. (2009) Negative Smad expression and regulation in
689 the developing chick limb. *PLoS One*, **4**, e5173.
690

691 **FIG. LEGENDS**

692 **Fig. 1.** *Fstl5* is associated with isolated clubfoot. **(A)** Clinical image of an infant with
693 bilateral clubfoot. **(B)** Manhattan plot of genome-wide association results. Dashed line
694 represents genome-wide significance ($5x10^{-8}$). Only the *Fstl5* locus exceeded genome-
695 wide significance. **(C-E)** LocusZoom plot showing association of rs76973778 (purple
696 circle) within the *FSTL5* locus in **(C)** the entire cohort, **(D)** males only, and **(E)** females
697 only. **(F)** Variant rs76811724 (purple circle) was genotyped for replication. This variant
698 exceeded genome-wide significance following meta-analysis combining the original and
699 replication cohorts (purple diamond).

700

701 **Fig. 2.** Evolutionarily conserved expression of *Fstl5* in the developing limb. **(A)**
702 Expression of bat *Fstl1*, *Fstl4* and *Fstl5* in developing forelimb and hindlimb at
703 embryonic stages CS15, CS16, and CS17 quantified using RNA-seq. Timepoints
704 correspond to similarly staged mouse limb development shown in panel C. **(B)** Whole-
705 mount *in situ* hybridization showing embryonic expression of chick *Fstl5*. Expression
706 becomes evident during later stages of limb development. **(C)** Expression of mouse *Fstl1*,
707 *Fstl4*, and *Fstl5* in developing forelimb and hindlimb at embryonic stages E12, E13, and
708 E13.5 quantified using RNAseq. Timepoints correspond to similarly staged bat limb
709 development shown in panel A. **(D)** Differential expression of mouse *Fstl5* in E15.5

710 forelimb and hindlimb measured by RT-qPCR (n=3 independent biological replicates).

711 Data represented as mean +/- standard error. Statistically significant differences were

712 determined using two-sided T-test. *******, p<0.001. **(E)** Mouse E15.5 *in situ* hybridization

713 showing *Fstl5* expression throughout the (i) entire embryo, (ii) forebrain, and (iii) heart.

714 **(F)** *In situ* hybridization of mouse embryonic E15.5 distal hindlimb showing localized

715 expression of *Fstl5* to the condensing cartilage mesenchyme.

716

717 **Fig. 3.** *In vivo* characterization of *Fstl5*. **(A)** Schematic diagram of *Sleeping Beauty*

718 insertion of β-galactosidase within intron 3 of rat *Fstl5*, predicted to result in premature

719 truncation. **(B)** Expression of *Fstl5* in brain tissue from control ($Fstl5^{+/+}$), heterozygous

720 ($Fstl5^{+/\beta geo}$), and homozygous ($Fstl5^{\beta geo/\beta geo}$) rats (n=3-5 animals per genotype).

721 Statistically significant differences were determined by a one-sided T-test. Data

722 represented as mean +/- standard error. **(C)** Immunohistochemistry showing (□-

723 galactosidase) expression in the pre-hypertrophic chondrocytes of the femur growth plate

724 of 2-week old $Fstl5^{\beta geo}$ rats. Expression diminishes in the hypertrophic chondrocytes of

725 the growth plate. Similar results were seen in tibia and metatarsal growth plates (data not

726 shown). **(D)** Schematic diagram of the $Fstl5^{LSL}$ conditional over-expression cassette

727 within the mouse *ROSA* locus. **(E and F)** Confirmation of cre-induced over-expression of

728 *Fstl5* in bone stromal cells cultured from *Prrx1*-cre;$Fstl5^{LSL}$ mice compared to control by

729 **(E)** qPCR and **(F)** Western blot. For western blotting, beta-actin is shown as loading

730 control. * indicates background band from Fstl5 antibody. **(G,H)** Skeletal phenotypes of

731 3-month old **(G)** female and **(H)** male *Prrx1*-cre;$Fstl5^{LSL}$ mice (grey) compared to age-

732 and gender-matched littermate controls (black). Data represented as mean +/- standard

733 error. Statistically significant differences were determined using 2-sided T-tests. *,

734 $p<0.05$; **, $p<0.01$.

735

736 **Fig. 4.** Analyses of mouse embryonic single-cell gene expression. **(A)** Schematic of *Fstl5*

737 expression analyses using publicly-available results from mouse embryonic scRNAseq

738 (ref. 28). **(B)** Enrichment of Reactome pathways for *Fstl5*-expressing sub-clusters. **(C)**

739 Bi-clustering of *Fstl* genes and their expression across the various cell lineages.

740

741

742