Supplementary Information

Supplementary Text

Evolution of the *Acvr1l* gene in vertebrates

Our experiments in zebrafish establish the BMP receptor type 1 gene *activin A receptor type 1-like* (*acvr1l*) as major target of Gli3 during pectoral girdle development (Figure 3). Due to the presence of *Acvr1l* in both actinopterygians such as zebrafish as well as in the sarcopterygian coelacanth, but not in tetrapods, it was previously concluded that this gene was present in the common ancestor of all bony vertebrates. Based on limited species coverage at the time, it was proposed that *Acvr1l* got lost specifically in tetrapods in association with the fish-to-tetrapod transition[1]. Here, we reconstruct the Activin receptor type 1 gene family history to reassess the evolution of the *Acvr1l* gene among vertebrates using an expanded taxon sampling (Extended Data Table 4).

Our phylogenetic and conserved synteny analyses show that three members of the gene family exist in jawed vertebrates: *Acvr1*, *Acvr1l*, and *Acvrl1* (Extended Data Fig. 4A, B; Extended Data Table 4). Within the gene family phylogeny, *Acvr1* and *Acvr1l* share a closer relationship with each other (Extended Data Fig. 4A). Orthology assignments to the three groups are furthermore supported by conserved syntenies (Extended Data Fig. 4B). Mapping the three members of the gene family onto previously established, ancestral chordate linkage groups (CLGs)[2], this gene family is likely derived from CLGB. The ancestral gene on CLGB amplified through the two genome duplications before the emergence of the gnathostome ancestor into four members, of which one was eventually lost, leaving a three member gene family.

Acvr1 is present in all major gnathostome lineages including cartilaginous fishes (i.e., little skate, ghost shark), actinopterygians (holosteans, teleosts), sarcopterygian fishes (coelacanth, lungfish) and tetrapods. Note though that zebrafish *acvr1* was not detected here, an absence that has been observed before by others, who therefore used zebrafish *acvr1l* as functional surrogate for human *ACVR1* in biomedical studies[3]. *Acvrl1* is also found in all major jawed vertebrate groups (Extended Data Fig. 4A, B; Extended Data Table 4).

Acvr1l, our focal gene family member, in contrast, shows a unique phylogenetic distribution (Extended Data Fig. 4C; Extended Data Table 4). In line with previous analyses[1,4], *Acvr1l* is present in all fish lineages, i.e., in cartilaginous fishes such as little skate and ghost shark, in ray-finned fishes including holosteans (gar, bowfin) as well as teleosts (e.g., zebrafish, medaka), and in lobe-finned fishes (i.e., coelacanth and lungfish). Our genomic surveys for *Acvr1l* furthermore support an almost complete absence of this gene from tetrapods. Strikingly though, we found one exception, an *Acvr1l* ortholog specifically retained in the genomes of crocodilians (i.e., American alligator, saltwater crocodile). Orthology of the crocodilian *Acvr1l* gene with the *Acvr1l* in fishes is supported by its grouping within the phylogeny (Extended Data Fig. 4A) and by conserved synteny with the *Acvr1l* locus in cartilaginous fishes (Extended Data Fig. 4B.1). On the microsyntenic scale, the gene content directly adjacent to *Acvr1l* is conserved in alligator and cartilaginous fishes: *Col15a1> Tgfbr1> Acvr1l> Reck>* in alligator and *Reck> Acvr1l> Tgfbr1l> <Col15a1* in little skate and ghost shark. In coelacanth, the situation is similar with *Reck> Acvr1l>* (Extended Data Fig. 4B.1). In addition, in chicken and spotted gar, the gene orders *COL15A1> TGFBR1> RECK>* and *col15a1> reck> tgfbr1>*, respectively (Extended Data Fig. 4B.2), support that the last common ancestor of jawed vertebrates had a genomic region consisting of *col15a1, tgfbr1*, *acvr1l*, *reck*.

This analysis further suggests that the *Acvr1l* gene took different evolutionary paths in ray-finned fishes and tetrapods: In the ray-finned lineage, the *acvr1l* gene got translocated out of its ancestral genomic context into a new genomic environment on a different chromosome which is conserved among extant ray-finned fishes (Extended Data Fig. 4B.3). For tetrapods, the presence of the *Acvr1l* gene in alligator in a region with ancestrally conserved synteny to cartilaginous fishes leads us to revise the previous model for the loss of tetrapod *Acvr1l*. Instead of a single gene loss in the tetrapod ancestor as previously suggested[1], the *Acvr1l* gene got lost multiple times within the tetrapod lineage, including in amphibians, mammals, birds, turtles, and lepidosaurs (Extended Data Fig. 4C). We therefore propose that the importance of *Acvr1l* gradually faded in the tetrapod lineage towards becoming a non-essential gene, concomitantly with the dermal-to-endochondral skeletal shift of the pectoral girdle from fish to tetrapods. However, full loss of tetrapod *Acvr1l* occurred only later, independently and lineage-specifically, within all most all major extant tetrapod branches, with the notable exception of crocodilians (Extended Data Fig. 4C). Future analyses will need to clarify the function of the retained *Acvr1l* gene in the crocodilian lineage, what role it may play in girdle formation, and/or whether it obtained a novel function.

Supplementary references

1 Amemiya, C. T. *et al.* The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**, 311-316 (2013). https://doi.org:10.1038/nature12027

2 Simakov, O. *et al.* Deeply conserved synteny resolves early events in vertebrate evolution. *Nat Ecol Evol* **4**, 820-830 (2020). https://doi.org:10.1038/s41559-020-1156-z

3 Allen, R. S., Tajer, B., Shore, E. M. & Mullins, M. C. Fibrodysplasia ossificans progressiva mutant ACVR1 signals by multiple modalities in the developing zebrafish. *Elife* **9** (2020). https://doi.org:10.7554/eLife.53761

4 Zheng, S., Long, J., Liu, Z., Tao, W. & Wang, D. Identification and Evolution of TGF-beta Signaling Pathway Members in Twenty-Four Animal Species and Expression in Tilapia. *Int J Mol Sci* **19** (2018). https://doi.org:10.3390/ijms19041154

5 Long, W. L. & Ballard, W. W. Normal embryonic stages of the longnose gar, Lepisosteus osseus. *BMC Dev Biol* **1**, 6 (2001). https://doi.org:10.1186/1471-213x-1-6

6 Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137-R137 (2008). https://doi.org:10.1186/gb-2008-9-9-r137