Recurrent evolution of ligand-binding domain multiplicity fine-tunes TGFβ signaling in vertebrates

Petra Knaus

petra.knaus@fu-berlin.de

Freie Universitaet Berlin https://orcid.org/0000-0003-2492-1667

Jerome Jatzlau

Freie universitaet Berlin

Michael trumpp

Freie Universität Berlin

Julia Kühlwein

Freie Universität Berlin

Leon Obendorf

Freie Universität Berlin

Yao Le

National university Singapore

Heiner Kuhl

Leibnitz Institute

Marco Preussner

FU Berlin https://orcid.org/0000-0001-5155-0844

Paul Mendez

Freie Universität Berlin

Hendrik Burkert

Freie Universität Berlin

Wiktor Burdzinski

Freie Universität Berlin

Stefan Mundlos

Max Planck Institute for Molecular Genetics https://orcid.org/0000-0002-9788-3166

Christoph Winkler

University of Singapore https://orcid.org/0000-0003-4688-6241

Matthias Stöck

Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB) https://orcid.org/0000-0003-4888-8371

Biological Sciences - Article

Keywords:

Posted Date: January 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8406929/v1

Additional Declarations: There is **NO** Competing Interest.

Recurrent evolution of ligand-binding domain multiplicity fine-tunes TGFβ signaling in vertebrates

Jerome Jatzlau[1, *, ✉], Michael Trumpp[1,2, *], Julia Kühlwein[1], Leon Obendorf[1,2], Yao Le[6], Heiner Kuhl[5,7], Marco Preussner[1], Paul Mendez[1], Hendrik Burkert[1], Wiktor Burdzinski[1,3], Stefan Mundlos[8], Christoph Winkler[6], Matthias Stöck[5], Petra Knaus[1,3, ✉]

[1] Freie Universität Berlin, Institute of Chemistry and Biochemistry - Biochemistry, Berlin, Germany.

[2] International Max Planck Research School for Biology and Computation, Berlin, Germany

[3] Berlin-Brandenburg School for Regenerative Therapies (BSRT), Berlin, Germany.

[4] Leibniz-Forschungsinstitut für Molekulare Pharmakologie, Berlin, Germany.

[5] Department of Fish Biology, Fisheries and Aquaculture, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany.

[6] Department of Biological Sciences and Centre for Bioimaging Sciences, National University of Singapore, Singapore, Singapore.

[7] Ecotoxicological Laboratory, German Environment Agency (UBA), Berlin, Germany.

[8] Max Planck Institute for Molecular Genetics, Berlin, Germany.

✉ Corresponding author

* Equal contributing authors

Abstract

From sponges to mammals, TGFβ-signaling has evolved as one of the key pathways governing body plans, organ and cell type differentiations as well as homeostatic balance. The extensive conservation of the receptor core architecture across metazoans reflects strong evolutionary constraints, consistent with the severe developmental phenotypes caused by TGFβ receptor mutations in humans. Here, we identify unexpected evolutionary divergence within a key receptor structure: 12 independent multiplications of the ligand-binding domain across three receptor classes (ACVR1, BMPR2, TGFBR2) in a diverse set of vertebrate lineages, including fish, amphibians, birds, and mammals. These ligand binding domain duplications indicate structural and functional plasticity of TGFβ receptors arising from domain-level innovation, a feature that has remained unnoticed in established model organisms including zebrafish, the African clawed frog, and chicken. Two principal effects emerged. First, recently diverged lineages display strong conservation of the membrane-proximal ligand binding domain (LBD) relative to the membrane-distal LBD, including the ligand-interacting residues. In species such as chicken and salmon, this pattern is associated with enhanced ligand binding. Second, more ancient lineages, such as zebrafish, exhibit elevated evolutionary rates in the membrane-distal LBD, corresponding to the acquisition of an inhibitory function. Our findings reveal LBD multimerization as a recurring, lineage-independent broader mechanism to diversify or fine-tune TGFβ-signaling. This adds a novel regulatory dimension to one of the best-examined conserved and essential pathways in metazoan biology.

Introduction

Signaling through ligand-binding receptors has been a cornerstone of multicellular life since the early evolution of metazoans[1]. Multiplication of single exon-coded domains shaped the extracellular architectures of many growth factor receptors. Some, such as EGF-, FGF-, and VEGF receptors, contain repeated ligand-binding domains (LBDs), yet the number and arrangement of these repeats have remained conserved from early metazoans to vertebrates[2-5]. By contrast, other key ligand-binding receptors such as those of the TGFβ receptor family have been supposed to contain only a single LBD, conserved from invertebrates to mammals. Remarkably, we found an exception in this otherwise highly conserved pathway: in medaka (*Oryzias latipes*), a single exon encoding the ligand-binding domain in *ACVR1* underwent triplication, producing a receptor with three tandem LBDs[6].

The TGFβ-receptor family, comprising TGFβ and bone morphogenetic protein (BMP) receptors, represents a highly conserved signaling system, essential for regulating developmental and homeostatic processes across all vertebrates[7]. Their roles span from orchestrating early embryonic tissue formation to fine-tuning cellular identities across diverse cell lineages[8,9] These receptors share a characteristic organization, an extracellular LBD followed by a transmembrane (TM) and an intracellular kinase domain (KD)[10,11], that is preserved from invertebrates to vertebrates.

Following two rounds of whole-genome duplication (WGD) events, the number of TGFβ receptor genes expanded from five in invertebrates[12] to thirteen in vertebrates[13], of which twelve are retained in mammals, including five type II and seven type I receptors[14]. This expanded receptor repertoire co-evolved with more than thirty TGFβ ligand family members, forming highly specific ligand–receptor pairs[14]. Upon ligand binding to their high-affinity receptors, a tetrameric complex of two type I and two type II receptors assembles, enabling transphosphorylation and activation of the type I receptor kinase by a constitutively active type II kinase. Activated type I receptors phosphorylate intracellular mediators, most commonly SMAD transcription factors, which regulate the expression of TGFβ/BMP target genes in a cell type- and context-dependent manner to control key developmental decisions[8].

In this study we investigated whether LBD multiplications within the TGFβ receptor family are more widespread as previously recognized and how such structural innovations affect receptor signaling. Through a comprehensive analysis across vertebrate genomes, we identified 12 independent cases of convergent LBD multiplication in three TGFβ receptor genes, spanning a broad phylogenetic range from fish to mammals. These events were detected in multiple species, including elephant-nose fish, carps, clawed frogs, chicken and horses, highlighting that LBD multimerization is a recurrent evolutionary phenomenon across vertebrate lineages. Our findings reveal that these can substantially modulate receptor signaling properties. Thus, LBD multiplication represents a previously unknown mode of molecular innovation within the highly conserved TGFβ signaling system, adding a new layer of complexity and regulatory potential to this essential developmental pathway.

Results

1. WGD-independent LBD multiplication events in TGFβ receptors in vertebrate evolution

Previously, we identified and verified that the medaka *ACVR1* paralogs, derived from teleost-specific whole genome duplication, differ in their number of LBDs[6]. Whereas the medaka *ACVR1* carries three LBDs, the *ACVR1L* gene encodes for a BMP type I receptor with only 1 LBD. In contrast, zebrafish (*Danio rerio)* has lost the *ACVR1* gene and retained only *ACVR1L* with 1 LBD. To trace the origin of the ACVR1 LBD triplication, we compared available genome and transcriptome data across various ray-finned fish (Actinopterygii) lineages, supplementing missing data with RNA sequencing of heart tissues from selected representatives of individual families (**Fig. 1**). We found that an initial duplication of the *ACVR1* LBD occurred in the early ancestors of Osteoglossocephalai, followed by a secondary triplication event common to all Clupeocephala. Interestingly, individual LBDs were subsequently lost within herrings, anchovies and shads (Clupeoidei*)*, and the entire *ACVR1* gene was lost in all carp-like fishes (Cyprinoidei). We next asked whether LBD multiplication is specific to the *ACVR1* locus, or if it is a common feature across all teleost genomes following evolutionary adaptation after the teleost whole genome duplication event[15]. Surprisingly, we found an LBD triplication event in the *ACVR1* paralog *ACVR1L*, which is unique to the bonytongues (Osteoglossidae). We uncovered six additional LBD multiplication events in the type II receptor genes *BMPR2* and *TGFBR2*, indicating a broader pattern beyond the *ACVR1* locus. Notably, the *BMPR2* gene underwent LBD triplication twice, once within the bichirs (Polypteridae) and once the featherbacks and knifefishes (Notopteroidei). These events appear to be independent of *BMPR2* paralogy, as Polypteridae diverged before the teleost whole genome duplication. The *TGFBR2* locus underwent four independent LBD multiplication events in Clupeoidei, Cyprinoidei, pikes and mudminnows (Esociformes), and salmonids (Salmoninae). While the carp-specific *TGFBR2* LBD duplication preceded the carp-specific whole genome duplication (WGD) (**Supp Fig. 3A**), the *TGFBR2* LBD multimerization in Salmoninae occurred after the salmon-specific WGD. Finally, we extended our analysis to all jawed vertebrates (Gnathostomata) (**Fig. 2**) to test whether TGFβ-superfamily receptor LBD multiplication depends on preceding WGD. We identified four additional TGFBR2 LBD multiplication events: one in the *TGFBR2* paralog *TGFBR2L* within the african lungfish (Protopteridae), and three in *TGFBR2* within amphibians (Amphibia), chicken (Gallus)[16], and odd-toed hoofed mammals (Perissodactyla). Collectively, we identified 12 convergent LBD multiplication events across three TGFβ-family receptor genes, *ACVR1/ACVR1L*, *BMPR2*, and *TGFBR2/TGFBR2L,* in species that had undergone WGD and retained a backup paralog (e.g., *ACVR1* in all teleosts, or *TGFBR2* in Salmoninae). However, we also observed LBD multiplication in species with only a single gene copy, such as BMPR2 in Polypteridae and *TGFBR2* in Perissodactyla, suggesting that WGD is not a prerequisite for LBD multimerization. Interestingly, type I receptor

LBDs are encoded by a single exon, whereas type II receptor LBDs are split across two exons, highlighting that, in the case of *BMPR2* and *TGFBR2*, duplication of an entire exon pair was required (**Supp Fig. 4A**). Since the *TGFBR2* locus exhibited the highest number of LBD duplication events, we next investigated whether repetitive elements flanking exons 2 and 3 of *TGFBR2* are conserved across species, potentially facilitating repeated independent duplications. Whereas in the zebrafish *Tgfbr2a* locus, retroelement derived GT-rich sequences are framing the duplicated exon pair, these or other repetitive sequences are absent in the chicken (*Gallus gallus*) and horse (*Equus caballus*) *TGFBR2* loci at duplication sites (**Supp Fig. 4B**). The duplication sizes range widely, from ≥2.1 kb in zebrafish to ≥18.7 kb in horse, indicating that these exon-pair duplications arose through separate genetic events rather than from a shared ancestral duplication (**Supp Fig. 4B**).

2. The membrane-proximal LBD conveys ligand binding and signaling competences in multimerized elephant-nose fish gnBmpr2a

To assess the molecular function of convergently multimerized LBDs, we first analysed the sequence conservation among 3 LBDs within individual BMPR2 proteins of the Notopteroidei and Polypteridae lineages (**Fig. 3A, Supp. Fig. 1B**). In both groups, the LBDs exhibited intermediate levels of sequence identity (50-70%) relative to the membrane-proximal LBD (LBD^{in}), suggesting partial functional divergence within the repeats (**Fig. 3A**). BMPR2 functions as a type II receptor for a wide range of BMP and Activin ligands[17,18]. Whereas BMP binding to BMPR2-containing receptor complexes requires the presence of a high affinity type I receptor, BMPR2 can directly bind Activin A through a specific set of hydrogen bonds and hydrophobic interactions primarily involving residues located between fingers 2 and 3 (**Fig. 3B, C**)[19].

To determine the conservation of these critical interaction sites, we compared the BMPR2 LBD residues involved in ligand binding across the 3 domains. In both lineages, the inner LBD showed the highest degree of sequence conservation with human BMPR2, particularly at residues forming the Activin-specific binding interface (**Fig. 3C**). Interestingly, the inner LBDs, of Notopteroidei species retained 85% of the key residues involved in A-loop contacts in finger 2, while Polypteridae retained 69% (**Fig. 3C, D**). Consistent with this, members of the Notopteroidei also exhibited minimal loss of conserved interaction-mediating residues on finger 3, consistent with a maintained or partially preserved capacity of Activin binding in LBD^{in} (**Fig. 3D, Supp. Fig. 5A**).

To explore the functional impact of LBD triplication in BMPR2 receptors from Notopteroidei, we performed computational protein-protein docking using the Rosetta Commons suite[20]. Docking simulations were conducted using the triplicated LBD of BMPR2a from elephant-nose fish (*Gnathonemus petersii;* member of the Notopteroidei*),* and human Activin A as the ligand (**Fig. 3E, Supp. Fig. 5B**). Additionally, lineage-specific docking analyses were performed using the appropriate Activin A orthologs for both Notopteroidei and Polypteridae species (**Supp. Fig. 5E-G**). Consistent with the LBD conservation analysis (**Fig. 3A-C**), the inner LBD showed the lowest predicted interface energy, indicating the highest ligand-binding affinity (**Fig. 3E**). To experimentally test the functionality of the individual LBDs, we generated expression constructs of elephant-nose fish BMPR2a, including full-length receptors, deletion mutants containing only one or two LBD ($LBD^{mid\text{-}in}$). Binding of fluorescently-labelled human Activin A-

Cy5 was assed using ligand surface binding assay (LSBA, **Fig. 3F; Supp. Fig. 5C,6**)[21], and downstream signaling was quantified via a $CAGA_{12}$-driven dual luciferase assay, which reports SMAD2/3 activity (**Fig. 3G; Supp. Fig. 5D**).

In agreement with the docking predictions, only the inner, membrane-proximal LBD bound Activin A with high affinity. Neither the outer, nor the middle LBD alone showed detectable binding, while LBD^{in} and $LBD^{mid\text{-}in}$ exhibited comparable binding to the full-length receptor (**Fig. 3F**). This binding pattern was mirrored in the functional signaling output: constructs containing only the outer or middle LBD showed minimal SMAD2/3 activation, whereas LBD^{in} displayed the highest signaling competence (**Fig. 3G**). Notably, deletion of the outer and middle LBDs enhanced signaling efficiency, indicating that these LBDs are functional but act repressively, whereas only the membrane-proximal inner LBD contributes positively to Activin A responsiveness in the LBD triplicated receptor (**Fig. 3G**).

3. Multimerized TGFBR2 LBDs follow divergent evolutionary adaptations to fine-tune signaling

Next, we examined the TGFBR2 gene, which independently acquired LBD multiplications in seven distinct lineages, followed by additional expansions in salmonids and amphibians, as seen in atlantic salmon (*Salmo salar;* 3 LBDs) and lake trout (*Salvelinus namaycush;* 5 LBDs), gaboon caecilian (*Geotrypetes seraphini;* 3 LBDs), and african clawed frog (*Xenopus laevis;* 3 LBDs) (**Fig. 1,2**). To predict the molecular function of the multimerized LBD receptors, we first compared the LBD sequence homology within one protein (**Fig. 4A**). We observed that the degree of conservation varied among the seven families. While the carp (*Cyprinus carpio*) and clawed frog TGFBR2 LBDs showed the lowest conservation (25–50%) relative to the membrane-proximal LBD (LBD^{in}), pike (*Esox lucius*) and herring (*Clupea harengus*) exhibited intermediate conservation (50–90%), and the salmons, chicken, and horse displayed high conservation levels (90–100%). (**Fig. 4A**). TGFBR2 is known for its high affinity binding to TGFβ-ligands via polar and hydrophobic interactions of residues mainly residing in finger 1 of the LBD (**Fig. 4B, C**)[22]. When comparing the amino acids, which form the interaction epitope for human TGFβ1 within TGFBR2, these are highly conserved in the LBD^{in} domain of all tested species, whereas the middle or outer LBD domains show a lower degree of conservation (**Fig. 4C, Supp. Fig 7**). Interestingly, consistent with its overall low degree of conservation, the zebrafish Tgfbr2a LBD^{out} shows no conserved TGFβ1-binding amino acid residues (**Fig. 4C, D**). In contrast, the chicken LBDs exhibit 98% overall similarity and share an identical TGFβ1-binding epitope (**Fig. 4C, D**). Lastly, the horse and tapir LBD^{out} show a high level of overall conservation (93.1%, 96%), with three or one amino acid substitutions within the TGFβ1-binding epitope, respectively (**Fig. 4C, D; Supp. Fig 7**). In contrast, to the varying degree of conservation on the side of the LBD interface, all TGFβ1 ligand orthologs show a complete conservation of amino acids forming the receptor interacting interface (**Supp. Fig 7**). Finally, to evaluate whether the extent of duplicated LBD sequence divergence corresponds to the evolutionary time separating each lineage, we calculated molecular clocks using orthologous LBD domains. The resulting lineage-specific clocks indicated that the LBD duplications in horse, chicken, and salmon evolved at rates consistent with their respective orthologous LBDs, indicating that these domains remain highly conserved under a clock-like model of evolution. However, the zebrafish LBD duplication strongly deviated from this pattern, exhibiting an exceptionally high substitution rate, far exceeding that of the clade-specific (otomorpha) LBD molecular clock, indicative of accelerated evolution (**Supp. Fig 8**).

To functionally predict the effect of TGFBR2 LBD duplication on ligand binding capabilities we again performed protein-protein docking studies (**Fig. 5A**) using human TGFβ1 and the respective animal ortholog TGFβ1, which exhibit a high degree of conservation (**Fig. 5**, **Supp. Fig. 9,10,11**). Furthermore, we cloned zebrafish, carp, clawed frog, chicken and horse TGFBR2 expression constructs, and generated single LBD mutants, to separately test the functionality of individual LBDs. Using fluorescently labeled human TGFβ1-SiRd12, we performed LSBA[21] to verify the predicted docking data (**Fig. 5B**) and tested the downstream signaling response using the SMAD2/3-sensitive $CAGA_{12}$-dual luciferase assay (**Fig. 5C**). In line with the low degree of LBD conservation in both zebrafish and carp Tgfbr2a, the LBD^{out} exhibits almost no TGFβ1 binding, whereas the LBD^{in} domain is capable of TGFβ1 binding and transmitting signals (**Fig. 5D-F**, **Supp. Fig. 9A-C, 12, 13**). Strikingly, only in the zebrafish dTgfbr2a the presence of the LBD^{out} is inhibitory on TGFβ1-binding to the LBD^{in}, as seen by diminished ligand binding in the LSBA (**Fig. 5E, Supp. Fig. 12**), and reduced SMAD2/3 activity as seen in the reporter gene assay (**Fig. 5F**).

Similarly, in horse and frog TGFBR2 orthologs, the membrane-proximal domain LBD^{in} shows the highest predicted and experimentally confirmed binding to TGFβ1, while the outer domain (LBD^{out}) exhibits reduced or no binding and lacks signaling competence. Importantly, the presence of the LBD^{out} does not impair binding or signaling function of the full length TGFBR2 for either species (**Fig. 5G-I**, **Supp. Fig. 9D-F, 14, 15**).

In contrast, chicken TGFBR2 shows a different pattern. Both LBD^{in} and LBD^{out} share a conserved TGFβ1 binding epitope and 98% sequence identity, resulting in similar predicted and measured TGFβ1 binding. Each domain alone is sufficient to support efficient ligand binding and downstream signaling (**Fig. 5J-L**, **Supp. Fig. 16**). Interestingly, the native chicken receptor containing both LBDs binds TGFβ1 more strongly than either single domain receptor although signaling output remains comparable regardless of the number of LBDs (**Fig. 5K-L)**.

To model the functional impact of a recent, fully conserved LBD duplication, we engineered a synthetic human TGFBR2 variant with two identical LBDs and assessed its ability to bind human TGFβ1. Consistent with the chicken receptor, LBD duplication significantly enhanced ligand binding (**Fig. 6A-B**, **Supp. Fig. 17**). Whereas $CAGA_{12}$-luc activity after 24 hours was comparable for both TGFBR2 variants (**Fig. 6C**), short term activation of SMAD2 was reduced in the presence of TGFBR2 2 LBD compared to TGFBR2 expression (**Fig. 6D**). Collectively, this highlights that 2 LBD TGFBR2 receptors remain functional but limit the downstream SMAD activation through excessive ligand binding (**Fig. 6E**).

These findings suggest that LBD duplication in TGFBR2 is a recurrent evolutionary mechanism capable of fine-tuning ligand binding without compromising signaling output, thereby offering a potential selective advantage under specific physiological or developmental contexts.

4. LBD duplication generates a low-signaling dTgfbr2a variant with tissue-specific roles in zebrafish

Notably, TGFBR2 two-LBD variants from different species exhibited different signaling capabilities, which raises the question how these integrate into the critical fine-tuning of balanced BMP and TGFβ signaling in the respective species. To answer this, we analyzed publicly available RNA-seq datasets to screen for tissue-specific expression patterns in chicken, horse and zebrafish, which were characterized by different TGFBR2 variants. In both chicken and horse, the TGFBR2 two-LBD variant was highly expressed in the cardiovascular and reproductive systems (**Supp. Fig. 18B**). Interestingly, the high degree of amino acid

homology between both LBDs is equally reflected in the coding sequence of exon 3 and exon 3' for both TGFBR2 ortholog variants (**Supp. Fig. 18A**). Exon duplication events have been associated with alternative splicing in multiple gene families, e.g., immunoglobulin superfamily members[23] and tropomyosin[24]. Indeed, in both chicken and horse, we observed single LBD variants due to alternative splicing but at a marginal rate, indicating that the TGFBR2 two-LBD variant is the predominant receptor (**Supp. Fig. 18B-C**).

Similarly, zebrafish express only a two-LBD variant of dTgfbr2a and do not exhibit alternative splicing at this locus. However, unlike chicken and horse, zebrafish possess a second paralog, Tgfbr2b, which encodes a receptor variant containing a single LBD (**Fig. 7A**). To compare the functional properties of these paralogs, we assessed their ligand-binding affinity and signaling capacity. The LBD of dTgfbr2b displayed a modestly higher affinity for TGFβ1 relative to the inner LBD of dTgfbr2a (**Fig. 7B**). Notably, dTgfbr2b also showed markedly increased ligand sequestration at the cell surface, likely attributable to the inhibitory effect of the outer LBD in dTgfbr2a (**Fig. 7C**). This was reflected in reduced downstream signaling competence of dTgfbr2a *vs*. dTgfbr2b, suggesting potential for fine tuning TGFβ signaling by dTgfbr2a compared to dTgfbr2b, indicating that the two paralogs may enable nuanced regulation of TGFβ signaling in zebrafish (**Fig. 7D**).

To explore potential tissue-specific functions of the two zebrafish paralogs, we analyzed publicly available single-cell RNA-sequencing data from developing zebrafish embryos. In contrast to chicken and horse, where only the two-LBD variant is present, the single-LBD dTgfbr2b paralog was broadly expressed across all tissues (**Fig. 7E**). By contrast, the two-LBD dTgfbr2a paralog displayed selective enrichment in the intermediate mesoderm and hematopoietic lineages, with its expression increasing relative to dTgfbr2b during development (**Fig. 7F-G**).

Comparative transcriptomic profiling of cells expressing either Tgfbr2 paralog revealed largely non-overlapping gene expression signatures, suggesting distinct biological roles (**Fig. 7H**). Gene Ontology (GO) enrichment analysis revealed that genes specifically upregulated in dTgfbr2a^{+} cells were predominantly associated with cytoskeletal remodeling, small GTPase signaling and migration, particularly within immune and cardiovascular contexts (**Fig. 7I**). Given the well-established role of TGFβ-signaling in driving cellular plasticity through epithelial-to-mesenchymal-transition (EMT)[25-28], we next interrogated the expression of canonical EMT transcription factors. Intriguingly, EMT drivers such as Zeb1, Snai, Slug and Twist were markedly enriched in dTgfbr2b^{+} cells (**Fig. 7J**), consistent with its higher signaling competence relative to dTgfbr2a (**Fig. 7D**).

To validate these differences *in vivo*, we transiently expressed each receptor in developing zebrafish embryos. Expression of dTgfbr2b robustly induced phenotypes associated with imbalanced TGFβ/BMP-signaling[29,30], including increased incidence of cyclopia and ventralization compared to dTgfbr2a (**Fig. 7K, Supp. Fig. 19**). Notably, deletion of the outer LBD in dTgfbr2a restored signaling activity, phenocopying the higher incidences of ventralization and cyclopia observed with dTgfbr2b, and indicating a structural basis for reduced signaling capacity of dTgfbr2a (**Fig. 7K**). Conversely, the isolated expression of dTgfbr2a LBD^{out} enhanced cyclopia incidence relative to full-length dTgfbr2a, but not ventralization. Overall, the elevated ventralization observed *in vivo* is consistent with the higher signaling output measured *in vitro*, suggesting that increased dTgfbr2 signaling capability contributes to the ventralizing effect (**Fig. 7D, K**).

Together, these findings reveal that LBD duplication in paralogous TGFBR2 receptors can drive neofunctionalization, yielding a receptor variant with attenuated signaling capacity. This structural innovation likely enables tissue-specific modulation of TGFβ-signaling output,

offering a previously unrecognized mechanism by which vertebrates can diversify and fine-tune key developmental and homeostatic pathways.

Discussion

In this study, we identify a striking evolutionary innovation: independent duplication of ligand-binding domains across three receptor classes (ACVR1, BMPR2, TGFBR2) in at least 12 vertebrate lineages. Our findings reveal that LBD multiplication serves as an evolutionary platform for modulating TGFβ-signaling at two distinct levels, by influencing ligand-binding capacity (e.g., chicken) or signal transmission efficacy (e.g., zebrafish). Moreover, we uncover recurrent instances of secondary LBD trimerization, notably within amphibians and salmonids, suggesting a sustained selective pressure for structural innovations in receptor design. These repeated, lineage specific adaptations point to a broader strategy by which vertebrates fine-tune pathway responsiveness through domain architecture. Given the association of TGFβ receptor mutations with numerous human developmental disorders[31-33], our findings underscore the evolutionary constraints on receptor core architecture, while simultaneously revealing unexpected structural plasticity through domain-level innovation.

The evolutionary roots of the TGFβ-receptors trace back to early metazoans such as sponges and placozoans, which encoded a minimal complement of two type I and one type II receptor genes[14]. With the emergence of bilateral symmetry, the TGFβ ligand repertoire expanded in parallel, reflecting the increasing demand for finely tuned BMP and TGFβ signaling during axis specification and early morphogenesis[34]. By contrast, vertebrates exhibit a markedly expanded TGFβ receptor repertoire, reflecting two rounds of whole-genome duplications, with additional lineage-specific duplications in teleosts, *Xenopus laevis*[35-37] and more recently in carps and salmonids[14,38,39]. This expansion supports the complex and context-dependent roles of TGFβ family ligands in vertebrate development, where they direct tissue specification, cell movement, and patterning along dorsal–ventral and anterior–posterior axes[7].

Despite this diversification, vertebrate genomes encode more ligands than receptors, making signaling specificity heavily reliant on selective ligand expression and combinatorial assembly of tetrameric receptor complexes[14,34]. Tight regulatory control is achieved through multiple mechanisms, including soluble antagonists, co-receptors, transcriptional modulators, and both post-translational and epigenetic modifications[40]. The evolutionary significance of maintaining these precise interactions is underscored by the strong sequence conservation observed in co-evolved ligand-receptor pairs[41].

Recurrent vertebrate WGDs generated additional TGFβ receptor paralogs, facilitating the evolutionary refinement of TGFβ signaling. Retention and divergence of paralogous receptor genes allowed one copy to evolve neofunctionalized or modular domain, while the other maintained a canonical receptor form [41,42]. This is exemplified by multiple teleost species, where Y-linked AMHR2 paralogs evolved truncated LBDs and ligand-independent activity, co-opting as sex-determining factors across diverse lineages, including sharks, catfishes, percids, the alligator pipefishes and the common seadragon[43-46].

Here we showed that twelve independent LBD expansions across *ACVR1*, *BMPR2*, and *TGFBR2* arose at different time points in vertebrate evolution. These events demonstrate that ligand-binding domain multimerization has repeatedly evolved as a novel layer of refining TGFβ receptor function. Intriguingly, domain multimerization is not limited to vertebrates. Independent extracellular domain duplications in ancestral TGFBR2 orthologs have been described in invertebrates such as sponges and oysters[47,48], indicating that LBD duplication is an evolutionarily widespread strategy, independent of WGDs.

The initial functional advantage of LBD duplication likely relates to the context-specific assembly of tetrameric TGFβ-receptor complexes[49]. While a second LBD does not alter receptor stoichiometry per se, it could enhance ligand capture, especially for high affinity receptors such as TGFBR2, in tissues with low ligand availability or restricted diffusion. For instance, duplicated LBDs in chicken, tapir and salmon TGFBR2s are nearly identical, potentially facilitating enhanced ligand avidity or interaction with ECM-bound latent ligands, analogous to the role of the co-receptor endoglin in BMP9 signaling[50,51]. In these species the duplicated LBDs show low amino-acid substitution rates, in line with the clade-specific molecular clock, indicating that they are under strong evolutionary constraint and remain highly conserved. This is reflected by enhanced TGFβ1 binding of a synthetic human 2 LBD TGFBR2. Whereas, this receptor exhibits limited SMAD activation, likely due to high ligand sequestration, it remains an open question how 2 LBD receptors integrate into the native receptor complex compositions, including tetrameric complex stoichiometries and modulatory co-receptors. Notably, in lineages such as chicken and horse, alternative splicing can generate receptor variants that selectively include or exclude individual LBDs, suggesting that both duplicated domains remain functionally relevant and may be differentially deployed depending on cellular context.

This contrasts with zebrafish, where the duplicated outer LBD shows markedly higher amino-acid substitution rate, indicating relaxed constraint and suggesting that the domain is drifting or diversifying rather than being maintained for interchangeable function. Such relaxed constraint is likely enabled by the presence of the paralogous receptor (Tgfbr2b) that provides functional redundancy and permits neofunctionalization of the duplicated domains. Consistent with this, we show that the outer LBD of Tgfbr2a attenuates signaling, acting as an inhibitory module and illustrating domain-level neofunctionalization.

Together, these findings reveal that LBD duplication acts as a versatile mechanism for modulating TGFβ-receptor function, enabling the evolution of context-sensitive signaling outputs. Given the pivotal role of TGFβ-signaling in regulating cell fate, plasticity, and migration, such domain-level adaptations likely contributed to the evolutionary diversification of vertebrate body plans. Importantly, these natural strategies may inform the rational design of synthetic TGFβ modulators, such as multivalent ligand traps, with potential applications in regenerative medicine and cancer therapy.

Material and Methods

Ethics and Animal Welfare Statement

Our research complies with all relevant ethical regulations; the relevant animal protection committee of the Leibniz-IGB and the city of Berlin (LaGeSo) approved the animal use protocol where relevant. *Anguilla anguilla* and *Cyprinus carpio* were kept under animal husbandry permit ZH114 (LaGeSo, Berlin) at IGB; three other fish species (*Gnathonemus petersii, Pantodon buchholzi, Erpetoichthys calabaricus*) were obtained from commercial dealers and humanely euthanized using an overdose of buffered Tricaine PHARMAQ 1000 MG/G (MS222; concentration: 500 mg/L), for commercial fish at the day of arrival at IGB. This ensured that no pain, suffering, distress or lasting harm was inflicted on the animals. RNA samples of *Xenopus laevis* were obtained from control groups of an animal experiment approved by the German State of Health and Social Affairs (LaGeSo, Berlin, Germany; G0359/12).

Zebrafish experiments were performed in accordance with protocol BR22-1497 approved by the Institutional Animal Care and Use Committee (IACUC) of the National University of Singapore. Adult zebrafish were housed in recirculating aquaria systems at 28°C under a 14 hour/10 hour light/dark cycle in the fish facility of the Department of Biological Sciences (DBS) at the National University of Singapore.

RNA isolation and cDNA synthesis

Taxon sampling includes selection from *Danio rerio* (fin – adult, whole larvae), *Xenopus laevis* (brain, testis, liver and gonad mixture (adult), larvae), *Cyprinus carpio* (gonad, muscle, brain, heart, fin – adult), *Gnathonemus petersii* (gonad, muscle, brain, heart, fin – adult), *Pantodon buchholzi* (gonad, muscle, brain, heart, fin – adult), *Erpetoichthys calabaricus* (gonad, muscle, brain, heart, fin – adult), *Anguilla anguilla* (gonad, muscle, brain, heart, fin – adult), and *Gallus gallus* (DF1 cell line). Tissues of *Anguilla anguilla, Cyprinus carpio Gnathonemus petersii, Pantodon buchholzi, Erpetoichthys calabaricus* were stored in RNAlater (Thermo Fisher Scientific) for 24 hours at 8°C, then drained and stored at -80°C until further processing. RNAs were extracted using TRIzol Reagent (Thermo Fisher Scientific, Waltham, USA) according to the supplier's recommendation and cleaned using the RNeasy Mini Kit (Qiagen). For synthesis of *Danio rerio* and *Gallus gallus* cDNA, 1 µg of total RNA was reversely transcribed by incubation with random primers (100 pmol μL^{-1}, Invitrogen) and M-MuLV reverse transcriptase enzyme (New England Biolabs) following manufactures recommendations. Synthesis of cDNA of *Xenopus laevis* and *Cyprinus carpio* was performed using iScript RT Supermix for RT-qPCR kit (Biorad); 10 µL of RNA (concentration of 80 ng/µL) were randomly reverse-transcribed following the manufactures recommendations.

RNA sequencing and *de novo* transcriptome assembly

For transcriptomics of *Anguilla anguilla, Gnathonemus petersii, Pantodon buchholzi and Erpetoichthys calabaricus* muscle tissue, RNA and library processing and RNA-seq were carried out by NOVOGENE (Cambridge, UK) on a NovaSeq 6000 PE 150, generating 9.3 – 10.0 Gb of raw data per sample. Raw data of each sample was processed using the fq2fa – merge command from IDBA v1.1.1[52] package to convert fastq to interleaved fasta format. The corresponding files were assembled using idba_tran v 1.1.1 with default parameters. The highest kmer transcript assemblies (kmer 60) were aligned to a custom protein database (consisting of sequences for *ACVR1, ACVRL1, BMPR1A, BMPR1B, ACVR1B, ACVR1C, TGFBR1, ACVR2A, ACVR2B, TGFBR2, BMPR2, AMHR2* and *INHBA* from *Siniperca chuatsi* [53] by blat[54] using parameters (-t=dnax -q=prot). Blat results were sorted by matching score and the best hit transcripts were written to a bed6 formatted table and extracted from the transcriptomes in 5' to 3' orientation using bedtools getfasta (-name -s)[55]. The extracted sequences were further processed as described below.

Phylogenetic analysis

Sequences of TGFβ receptor family orthologs were screened for LBD multimerization using Ensembl gene tree function[56] and NCBI blast (*https://blast.ncbi.nlm.nih.gov/Blast.cgi*) against available vertebrate transcriptomes and/or genomes. RNA-seq was performed for selected species (*Gnathonemus petersii*, *Pantodon buchholzi*, *Erpetoichthys calabaricus*, *Anguilla anguilla)* to complement missing information in respective taxa. In the absence of available transcriptome data, ORFs were predicted in regions identified by nucleotide Blast of closely related species using Augustus web interface (*https://bioinf.uni-greifswald.de/augustus/*)[57].

LBDs were predicted from amino acid sequences using NCBI Conserved Domain Search web interface (*https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi*)[58]. Phylogenetic trees were extracted from NCBI taxonomy CommonTree (*https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi*)[59] and represented using Interactive Tree of Life (iTOL) v6[60]. Species images were extracted from Adobe Stock (ID: #301044215, #327193716, #370603116, #61858943, #115805136, #490808539, #125573996, #730691035, #552043292, #227759813, #574310862, #35897212, #370612679) under the license of the "Max Planck Institut für Molekulare Genetik, Berlin" and edited in Adobe Photoshop Version 26.8.0.

RNA splicing analysis

To analyse *TGFBR2* alternative splicing in horse and chicken, RNA-seq data from various tissues (horse: PRJNA1017964; chicken: PRJEB26695) were aligned to the respective reference genomes (chicken: GCA_016699485.1; horse: EquCab3.0) using STAR v2.7.9a. Unique and multi-mapping reads for *TGFBR2* exons 1–5 were extracted from the SJ.out.tab STAR output files using standard Python scripts. For sashimi plots, merged BAM files were visualized in IGV, and coverage tracks were exported. Junction reads were manually corrected by replacing them with unique and multi-mapping counts from the SJ.out.tab files. Reported values correspond to the sum across all analysed replicates. In chicken samples, no reads corresponding to the E1–E2' junction were detected. In horse, E2 inclusion relative to E2' was calculated using uniquely mapping reads only. For the final table, only samples with at least 40 unique reads for the constitutive E4–E5 junction and tissues with at least two biological replicates were included. For horse, selected tissues known to be regulated by TGFβ-signaling are represented.

RNA single cell analysis in zebrafish

Single-cell RNA-seq data from the zebrahub atlas[61] were downloaded and processed in R using Seurat (v4)[62]. Cells with <500 detected features or genes expressed in <30 cells were excluded. Data were normalized, variable features identified, and expression values scaled prior to PCA (50 components) and UMAP embedding (first 30 PCs). Metadata annotations from the zebrafish anatomy ontology available from the zebrahub atlas were used for cell-type assignments. Gene expression was visualized on UMAPs and dot plots. For the intermediate mesoderm and hematopoietic system, cells were stratified into four groups based on tgfbr2a/tgfbr2b expression (double-positive, tgfbr2a+, tgfbr2b+, or double-negative). Differential expression relative to double-negative cells was assessed using Seurat's FindMarkers function, and pseudobulk-like expression profiles were generated by using Seurat's AverageExpression function. Differentially expressed genes were filtered by a log2 FC ≥ 0.58 and an adjusted p-value ≤ 0.05. Significantly diff. exp. genes relative to double-negative condition were compared using DeepVenn[63] (https://www.deepvenn.com/) and visualized in Photoshop. gProfiler g:GOSt[64] (https://biit.cs.ut.ee/gprofiler/gost) was used for functional enrichment analysis of genes specifically enriched in dTgfbr2a+ or dTgfbr2b+ cells compared to double-negative cells on indicated time points.

Molecular cloning

To functionally characterize TGFBR2 or BMPR2 receptors of *Danio rerio*, *Cyprinus carpio*, *Xenopus laevis*, *Gnathonemus petersii*, *Gallus gallus* and *Equus caballus*, we generated an N-terminally Halo-tagged receptor expression plasmid library (**Tab. 1**). For this, we used

various cloning methods including restriction-, Gibson-, blunt end- and deletion mutagenesis cloning, specified for each animal below. All primers used for PCR are listed in (**Tab. 2**). Cloning PCRs were carried out on a Peltier Thermal Cycler PTC-200. The elongation time was adapted according to the product size (Phusion Pol. = 1kb/min). PCR products were resolved by agarose gel electrophoresis and purified using NucleoSpin Gel and a PCR Clean-up kit (Macherey-Nagel), according to the manufacturer's guidelines. Gibson cloning using NEBuilder HiFi DNA Kit (New England BioLabs) was performed according to the manufacturer's protocol. Final products were transformed in chemically competent DH5α *E. coli* bacteria cells. All constructs were confirmed via Sanger sequencing followed by whole plasmid sequencing. Expression is confirmed via Western blot (**Sup. Fig. 20**).

Danio rerio and *Cyprinus carpio*

drTgfbr2a, drTgfbr2b and cTgfbr2a were amplified from respective cDNA and subcloned into a previously generated pcDNA3.1. hTGFBR2-halo plasmid[21] in between *EcoRI* and *NotI* sites, replacing the original receptor ORF resulting in drTgfbr2a-Halo, drTgfbr2b-Halo and cTgfbr2a-Halo expression plasmids. Further, to generate drTgfbr2a-LBD^{in}-Halo, drTgfbr2a-LBD^{out}-Halo, cTgfbr2a-LBD^{in}-Halo and cTgfbr2a-LBD^{out}-Halo constructs, deletion mutagenesis was performed on drTgfbr2a-Halo and cTgfbr2a-Halo templates. Deletion primers were used to flank either the inner or outer ligand binding domain sequences. Subsequently, PCR constructs were ligated via blunt end ligation.

Gnathonemus petersii

Following the RNA-seq analysis of *Gnathonemus petersii* heart tissue, we ordered the synthesis of gnBmpr2a-pTwist-CMV plasmid at "Twist Bioscience" and amplified gnBmpr2a, gnBmpr2a-LBD$^{mid\text{-}in}$ and gnBmpr2a-LBD^{in}. We then subcloned these constructs into hTGFBR2-Halo plasmid in between *EcoRI* and *NotI* sites, replacing the original receptor ORF resulting in gnBmpr2a-Halo, gnBmpr2a-LBD$^{mid\text{-}in}$-Halo and gnBmpr2a-LBD^{in}-Halo expression plasmids. Further, to generate gnBmpr2a-LBD^{out}-Halo and gnBmpr2a-LBD^{mid}-Halo constructs, Gibson cloning was performed on gnBmpr2a-Halo template.

Xenopus laevis

XlTgfbr2L was amplified from respective cDNA and subcloned using Zero Blunt™ PCR Cloning Kit (Thermo Fisher Scientific) according to the manufacturer's protocol into a pCR™-Blunt-vector. The full length xlTgfbr2.L-Halo construct was synthesized by "Genewiz". Consecutively, Gibson-cloning was performed on this vector. By this, we obtained xlTgfbr2.L-LBD^{in}-Halo, xlTgfbr2.L-LBD^{out}-Halo, xlTgfbr2.L-LBD$^{mid\text{-}in}$-Halo.

Equus caballus

Due to high sequence identity between the inner and outer ligand binding domains, we obtained a codon optimized eTGFBR2-pTwist-CMV plasmid from "Twist Bioscience" and amplified eTGFBR2. We then subcloned this construct into hTGFBR2-Halo plasmid in between *EcoRI* and *NotI* sites, replacing the original receptor ORF resulting in eTGFBR2-Halo. Further, to generate eTGFBR2-LBD^{in}-Halo and eTGFBR2-LBD^{out}-Halo constructs, Gibson cloning was performed on eTGFBR2-Halo template.

Gallus gallus

gTGFBR2 (ENSGALT00000018657.6) and gTGFBR2-LBD^{out} (ENSGALT00000037691.5) were amplified from respective cDNA and subcloned into hTGFBR2-halo plasmid in between *EcoRI* and *NotI* sites, replacing the original receptor ORF resulting in gTGFBR2-Halo and gTGFBR2-LBD^{out}-Halo. The subcloned gTGFBR2 ORF contained an additional exon in between the kinase coding exon 6 and exon 7, which was removed by Gibson cloning. Thereby, both gTGFBR2 and gTGFBR2-LBD^{out} variants encoded for the same kinase, making a direct comparison of the duplicated LBD signaling competence possible. Further, to generate gTGFBR2-LBD^{in}-Halo, deletion mutagenesis was performed on gTGFBR2-Halo template. Deletion primers were used to flank the outer ligand binding domain. Subsequently, PCR constructs were ligated via blunt end ligation.

Capped mRNA synthesis and microinjection in zebrafish

To synthesize capped mRNA, the mMESSAGE mMACHINE T7 Transcription Kit (ThermoFisher) was used with *DraIII*-linearized dTgfbr2a-Halo, dTgfbr2a LBD^{in}-Halo, dTgfbr2a LBD^{out}-Halo, and *BfuI*-linearized dTgfbr2b-Halo plasmids as templates, respectively. To synthesize capped *PMT-mEGFP* (membrane bound EGFP) mRNA, the mMESSAGE mMACHINE SP6 Transcription Kit (ThermoFisher) was used with *NotI*-linearized *pcs2+-PMT-mEGFP*[65,66] plasmids as templates. The final concentrations of synthesized mRNAs were measured by Qubit 4 Fluorometer (ThermoFisher) using Qubit™ RNA HS Assay Kit (ThermoFisher). For microinjection, wild type (WT) embryos were collected immediately after spawning and transferred to 0.3X Danieau's solution (17.4 mM NaCl, 0.21 mM KCl, 0.12 mM $MgSO_4$, 0.18 mM $Ca(NO_3)_2$, 1.5 mM HEPES, pH = 7.2). Injected embryos were raised in a 28°C incubator until 22 hpf for screening and imaging. Embryonic stages were defined by hours post fertilization (hpf) at 28°C and morphological features[67]. To compare the respective effects of dTgfbr2a-Halo, dTgfbr2a LBD^{in}-Halo, dTgfbr2a LBD^{out}-Halo and Tgfbr2b-Halo in WT embryos, 170 pg *dTgfbr2a-Halo*, 148 pg *dTgfbr2a LBD^{in}-Halo*, 148 pg *dTgfbr2a LBD^{out}-Halo*, or 148 pg of *dTgfbr2b-Halo* capped mRNAs were co-injected with 10 pg *PMT-mEGFP* at the border between cytoplasm and yolk of 1-cell stage WT embryos, respectively.

Fluorescence screening and brightfield imaging of dTgfbr2-Halo injected zebrafish embryos

Injected embryos were incubated in a 28°C incubator and screened for mEGFP fluorescence at 22 hpf. mEGFP-positive embryos were characterized into three categories (normal, cyclopia and ventralized) based on eye, head and trunk morphologies for quantification. Representative embryos were imaged under a stereomicroscope (Nikon, SMZ18).

Cell culture

COS-7 and HEK293T cells were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ) and cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FCS, 2 mM L-glutamine and penicillin (100 units/mL) / streptomycin (100 µg/mL) (DMEM full medium) in a humidified atmosphere at 37 °C and 5% CO_2 (v/v). COS-7 and HEK293T cells were maintained in T175 flasks and split 1:5 or 1:10, depending on need and were kept sub-confluent. For passaging, cells were washed once with PBS before being removed from the flasks surface with trypsin/EDTA (0.05/0.02% in PBS).

SDS-PAGE & Western-blotting

For sodium dodecyl sulphate polyacrylamide gel-electrophoresis (SDS-PAGE), cells transiently expressing Halo-receptors were lysed in 150 µL Laemmli buffer and frozen at -20 °C. To ensure a homogeneous loading, cell lysate was pulled through a 1 ml syringe and boiled for 10 min at 95°C before loading onto 10% polyacrylamide gels. After performed gel-electrophoresis, proteins were transferred onto Methanol-activated PVDF membranes by Western-blotting. Next, membranes were blocked for 1 hour in a solution containing 0.1% TBS-T and 3% w/v bovine serum albumin (BSA), then washed three times in 0.1% TBS-T and incubated with indicated primary antibodies overnight at 4°C. Primary antibodies: anti-Halo (ProMega; #G9211; monoclonal mouse antibody), anti-pSMAD2 Ser465/467 (Cell Signaling; #3108; monoclonal rabbit antibody), and anti-GAPDH (Cell Signaling; #2118; monoclonal rabbit antibody) were used at a 1:1000 dilution in 3% w/v BSA/ TBS-T solution. For HRP-based detection, membranes were incubated with secondary goat-α-mouse or goat-α-rabbit IgG HRP conjugates (± 0.8 mg/ml, Dianova; #111-035-144, #115-035-068) for 1 hour at a dilution of 1:10000. Chemiluminescent reactions were processed using WesternBright Quantum HRP substrate (Advansta Inc.) and documented on a FUSION FX7 digital imaging system.

Confocal microscopy & Ligand Surface Binding Assay (LSBA)

LSBA assay and consecutive image analysis and quantification was previously described in detail for SiR-d12-hTGFβ1 and Cy5-hActivin A[21]. In brief, for visualization of SiR-d12-hTGFβ1 and Cy5-hActivin A ligand binding, 200.000 COS-7 cells / well were seeded in 1 mL DMEM full medium on glass cover slips in a 12 well plate. On the following day, cells were transfected with desired constructs (500 ng DNA per well) using Lipofectamine2000 according to the manufacturer's instructions. The day after, cells were washed with PBS and, additionally to 0.5 mM fluorescent HaloTag-ligand CA-Alexa488 (Alexa Fluor 488, Promega, #G1002), simultaneously incubated with saturating conc. of SiR-d12-hTGFβ1 or Cy5-hActivin A for 30 minutes at 4 °C. Subsequently, cells were washed once with ice-cold PBS before fixation with 100% methanol for 5 minutes. Cells were washed again with PBS and mounted with Fluoromount G (Invitrogen, 00-4958-02). Confocal images were acquired on a commercial expert line Abberior STED microscope, using the Imspector software from Abberior Instruments (Version 16.3) in line-scanning confocal mode; objective lens: 100X NA1.4 (oil) [UPLS], pinhole: 1.0 AU, range: 75 µm x 75 µm, confocal pixel size: 60 nm, pixel dwell time: 5/10 µs. Images were acquired using 485 nm (20% laser power) and 640 nm excitation (20% laser power). The detection windows were set to 500 – 540 nm and 650 – 750 nm. Confocal raw data were post-processed and adjusted for color and contrast (linear adjustments maintained for confocal datasets represented within one figure) using Fiji (ImageJ) software and Adobe Photoshop (Adobe Systems). Surface binding quantification of SiR-d12-hTGFβ1 and Cy5-hActivin A on COS-7 cells transiently expressing Halo-Receptor constructs was performed with Fiji. Per cell, four regions of interest (ROI) (100 µm²) were chosen, and the raw integrated intensity (RawIntDen) of each ROI was measured both in receptor and ligand channels. Per condition, 30 cells were quantified in total in 3 independent experiments. Receptor-ligand binding was calculated relative to intensity values of untransfected COS-7 cells, representing endogenous ligand-receptor binding. Normalized RawIntDen values of receptors and ligands were plotted in GraphPad Prism 10.5 (GraphPad Software Inc.).

Dual Luciferase Reporter Gene Assay

For dual luciferase reporter gene assay, 50.000 HEK293T cells / well were seeded in 200 µL DMEM full medium in a 96 well plate. On the following day, cells were transfected with 50 ng

of corresponding Halo-tagged receptors constructs together with 50 ng of SMAD2/3 sensitive $(CAGA)_{12}$ luciferase reporter. A constitutively expressing construct encoding renilla luciferase (RL-TK; Promega) was co-transfected (30 ng) as internal control. The next day, cells were starved in serum-free medium for 3 h before stimulation with 0.2 nM rhTGFβ1 (PeproTech, Hamburg, Germany) or 0.02 nM rhActivin A (Gift from Marko Hyvönen) for 24 h. Cell lysis was performed using passive lysis buffer (Promega, # E1910) and measurement of luciferase activity was carried out according to manufacturer's instructions using a TECAN Spark plate reader. Experiments testing the TGFBR2 ortholog signaling competence were normalized against the rhTGFβ1 response of hTGFBR2-Halo.

Structure prediction and interface energy calculations

The structure of the receptor and ligand complexes were first predicted using AlphaFold2 multimer[68,69] using the single receptor ECDs and two ligand chains (either hActivin A, hTGFβ1 or the respective animal variants) for the ligand homodimers. In order to obtain a physically more validated structure while simultaneously scoring the interface, first coordinate constrained relaxation protocol was performed to optimize H-bonds based on the Rosetta Energy Function 2015 (ref2015_cst) within RosettaScripts. Local Monte-Carlo-based protein-protein docking using RosettaScripts[20] was performed in 2500 replicas. Here, first, the "Docking" mover was used with low resolution (backbone plus centroid) flags before a high resolution (full atom) docking (using ref2015 scoring function). Other parameters as distance perturbation and angle perturbation were left on default. The Interface RMSD (iRMSD) was plotted against the score of the interface residues (i_SC) of the top 500 structures (determined by i_SC) and mean binding energies were calculated. Lower interface score correlates with favourable interface energies. Lower iRMSD correlates to docking structures closer to the starting AlphaFold2-multimer conformation. For structural representation, the top 10 hits showing the lowest i_SC were structurally investigated for the final homology model. All used scripts for structure predication and interface energy calculations are available under https://github.com/agknaus/Jatzlau-Trumpp-et-al.-structural-computational-analysis/.

Molecular Clock and Evolutionary Rate Analysis of LBDs

Amino-acid sequences of LBD regions were aligned in MEGA X [70]. Alignments were generated using the MUSCLE algorithm with default parameters. Pairwise evolutionary distances between LBD sequences were calculated in MEGA X using the Analyze → Compute Pairwise Distances function. Distances were estimated under the Jones–Taylor–Thornton (JTT) substitution model with gamma-distributed rate heterogeneity ($\Gamma = 1.0$) and a homogeneous substitution pattern among lineages. The substitution type was set to Amino acid, and gaps or missing data were handled using pairwise deletion. Variance estimation was performed using the bootstrap method with 1000 replicates. For each focal species (chicken, horse, salmon, zebrafish), distances were calculated between its inner LBD and the orthologous inner LBDs of other species within the same clade. Distances between the two LBD copies within the focal species were computed separately to quantify intra-species divergence associated with the domain duplication event.

Species divergence times for all pairwise comparisons were obtained directly from TimeTree.org [71]. These values were used as the independent variable in all molecular-clock analyses.For each clade, the amino-acid substitution rate (r, substitutions per site per million years) was estimated by performing a weighted linear regression of pairwise amino-acid

distances against their corresponding TimeTree-derived species divergence times. The slope of the regression provided the clade-specific molecular-clock rate for the inner LBD domain. The observed amino-acid distance between the two LBD copies within each focal species was divided by the clade-specific molecular-clock rate to obtain the duplication-specific evolutionary rate ($r_{(}dup_{)}$). This value reflects the effective rate of amino-acid substitution between the two paralogous domains since their duplication.

Statistical analysis

All statistical tests were performed using GraphPad Prism version 10.5.0 software and are listed in the figure legends. Normal distribution of data sets was tested with the Shapiro-Wilk normality test. In cases of failure to reject the null hypothesis, the ANOVA and Tukey's or Dunnett post hoc test were used to check for statistical significance under the normality assumption. For all experiments statistical significance was assigned, with an alpha-level of $p < 0.05$.

Graphical schemes and figures

Figures were generated and assembled using PyMOL, GraphPad Prism, Biorender.com and Adobe® Photoshop (Adobe Systems, San José, USA).

Acknowledgments

J.J. and P.K. acknowledge the support from Deutsche Forschungsgemeinschaft DFG (SFB1444) and the Einstein Center ECRT). M.T. and L.O. was supported by the Max Planck Research School IMPRS-BAC. Funding has been in part provided by the German Research Foundation (DFG) STO493/8-1 to MS. CW was funded by grants from the Singapore Ministry of Education (grant numbers MOE-2016-T3-1-005 and MOE-T2EP30221-0008). Further, the authors would like to thank Francesca Bottanelli (Freie Universität Berlin) for providing access to her commercial expert line Abberior STED microscope. W.B. was supported by Berlin–Brandenburg School for Regenerative Therapies. The authors would like to thank BioRender for providing a platform to generate elements of Figure 1 – 7, Supplemental Figure 3 – 5, Supplemental Figure 8 – 10, Supplemental Figure 17 (Created in BioRender. Trumpp, M. (2025) https://BioRender.com/8p64dhh).

Author Contributions

J.J., M.T. designed the study, evaluated the data and generated all figures; M.T., J.J., J.K., Y.L., H.B. and performed the experiments; J.J. and M.T. performed the phylogenetic analysis; L.O., J.K. performed receptor homology modeling & docking calculations; M.T. produced hTGFβ1-SiR-d12; Y.L. performed zebrafish experiments; M.S. collected the tissue samples of *G. petersii*, *P. buchholzi*, *E. calabaricus*, *A. anguilla* and *X. laevis*, and prepared RNA and cDNA and discussed ancient/recent genome duplications; C.W. provided *D. rerio* cDNA samples. M.T., J.J., J.K, and H.B. cloned all Halo-receptor constructs. H.K. processed the transcriptome data and analyzed the results. M.P. performed RNA splicing analysis, P.M. analyzed scRNA-seq data, W.B. analyzed LSBA data; P.K., M.S., C.W., and S.M. discussed the project and gave critical input. J.J., M.T. and P.K. wrote the manuscript; all authors commented on the manuscript.

Conflict of Interest

The authors declare no competing interest.

Data availability

The data supporting the findings of this study are available from the corresponding authors upon request. RNA raw sequence datasets for *Gnathonemus petersii, Pantodon buchholzi, Erpetoichthys calabaricus and Anguilla anguilla* have been submitted to NCBI/SRA under the BioProject accession: PRJNA1314022.

References

1 Pires-daSilva, A. & Sommer, R. J. The evolution of signalling pathways in animal development. *Nat Rev Genet* **4**, 39-49 (2003). https://doi.org/10.1038/nrg977

2 Barberan, S., Martin-Duran, J. M. & Cebria, F. Evolution of the EGFR pathway in Metazoa and its diversification in the planarian Schmidtea mediterranea. *Sci Rep* **6**, 28071 (2016). https://doi.org/10.1038/srep28071

3 Bogdan, S. & Klambt, C. Epidermal growth factor receptor signaling. *Curr Biol* **11**, R292-295 (2001). https://doi.org/10.1016/s0960-9822(01)00167-1

4 Itoh, N. & Ornitz, D. M. Evolution of the Fgf and Fgfr gene families. *Trends Genet* **20**, 563-569 (2004). https://doi.org/10.1016/j.tig.2004.08.007

5 Kipryushina, Y. O., Yakovlev, K. V. & Odintsova, N. A. Vascular endothelial growth factors: A comparison between invertebrates and vertebrates. *Cytokine Growth Factor Rev* **26**, 687-695 (2015). https://doi.org/10.1016/j.cytogfr.2015.04.001

6 Trumpp, M. *et al.* Characterization of Fibrodysplasia Ossificans Progessiva relevant Acvr1/Acvr2 Activin receptors in medaka (*Oryzias latipes*). *Plos One* **18** (2023). https://doi.org/10.1371/journal.pone.0291379

7 Zinski, J., Tajer, B. & Mullins, M. C. TGF-β Family Signaling in Early Vertebrate Development. *Cold Spring Harb Perspect Biol* **10** (2018). https://doi.org/10.1101/cshperspect.a033274

8 Hiepen, C., Yadin, D., Rikeit, P., Dorpholz, G. & Knaus, P. Actions from head to toe: An update on Bone/Body Morphogenetic Proteins in health and disease. *Cytokine Growth Factor Rev* **27**, 1-11 (2016). https://doi.org/10.1016/j.cytogfr.2015.12.006

9 Nickel, J. & Mueller, T. D. Specification of BMP Signaling. *Cells* **8** (2019). https://doi.org/10.3390/cells8121579

10 Yadin, D., Knaus, P. & Mueller, T. D. Structural insights into BMP receptors: Specificity, activation and inhibition. *Cytokine Growth Factor Rev* **27**, 13-34 (2016). https://doi.org/10.1016/j.cytogfr.2015.11.005

11 Bragdon, B. *et al.* Bone morphogenetic proteins: a critical review. *Cell Signal* **23**, 609-620 (2011). https://doi.org/10.1016/j.cellsig.2010.10.003

12 Upadhyay, A., Moss-Taylor, L., Kim, M. J., Ghosh, A. C. & O'Connor, M. B. TGF-β Family Signaling in Drosophila. *Cold Spring Harb Perspect Biol* **9** (2017). https://doi.org/10.1101/cshperspect.a022152

13 Huminiecki, L. *et al.* Emergence, development and diversification of the TGF-beta signalling pathway within the animal kingdom. *BMC Evol Biol* **9**, 28 (2009). https://doi.org/10.1186/1471-2148-9-28

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

15 Hoegg, S., Brinkmann, H., Taylor, J. S. & Meyer, A. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J Mol Evol* **59**, 190-203 (2004). https://doi.org/10.1007/s00239-004-2613-z

16 Ning, B. *et al.* Genomic organization, intragenic tandem duplication, and expression analysis of chicken TGFBR2 gene. *Poultry Science* **101**, 102169 (2022). https://doi.org/https://doi.org/10.1016/j.psj.2022.102169

17 Heinecke, K. *et al.* Receptor oligomerization and beyond: a case study in bone morphogenetic proteins. *Bmc Biol* **7**, 59 (2009). https://doi.org/10.1186/1741-7007-7-59

18 Sebald, W., Nickel, J., Zhang, J.-L. & Mueller, T. D. Molecular recognition in bone morphogenetic protein (BMP)/receptor interaction. *Biological Chemistry* **385**, 697-710 (2004). https://doi.org/doi:10.1515/BC.2004.086

19 Chu, K. Y., Malik, A., Thamilselvan, V. & Martinez-Hackert, E. Type II BMP and activin receptors BMPR2 and ACVR2A share a conserved mode of growth factor recognition. *J Biol Chem* **298**, 102076 (2022). https://doi.org/10.1016/j.jbc.2022.102076

20 Fleishman, S. J. *et al.* RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *Plos One* **6**, e20161 (2011). https://doi.org/10.1371/journal.pone.0020161

21 Jatzlau, J. *et al.* A versatile Halo- and SNAP-tagged BMP/TGFβ receptor library for quantification of cell surface ligand binding. *Commun Biol* **6** (2023). https://doi.org/10.1038/s42003-022-04388-4

22 Radaev, S. *et al.* Ternary Complex of Transforming Growth Factor-β1 Reveals Isoform-specific Ligand Recognition and Receptor Recruitment in the Superfamily*. *Journal of Biological Chemistry* **285**, 14806-14814 (2010). https://doi.org/https://doi.org/10.1074/jbc.M109.079921

23 Sun, J. *et al.* Origination of New Immunological Functions in the Costimulatory Molecule B7-H3: The Role of Exon Duplication in Evolution of the Immune System. *Plos One* **6**, e24751 (2011). https://doi.org/10.1371/journal.pone.0024751

24 Gunning, P. W., Schevzov, G., Kee, A. J. & Hardeman, E. C. Tropomyosin isoforms: divining rods for actin cytoskeleton function. *Trends Cell Biol* **15**, 333-341 (2005). https://doi.org/10.1016/j.tcb.2005.04.007

25 Piera-Velazquez, S. & Jimenez, S. A. Endothelial to Mesenchymal Transition: Role in Physiology and in the Pathogenesis of Human Diseases. *Physiol Rev* **99**, 1281-1324 (2019). https://doi.org/10.1152/physrev.00021.2018

26 Goumans, M.-J., van Zonneveld, A. J. & ten Dijke, P. Transforming Growth Factor β–Induced Endothelial-to-Mesenchymal Transition: A Switch to Cardiac Fibrosis? *Trends in Cardiovascular Medicine* **18**, 293-298 (2008). https://doi.org/https://doi.org/10.1016/j.tcm.2009.01.001

27 Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial–mesenchymal transition. *Nature Reviews Molecular Cell Biology* **15**, 178-196 (2014). https://doi.org/10.1038/nrm3758

28 Dooley, S. & ten Dijke, P. TGF-β in progression of liver disease. *Cell Tissue Res* **347**, 245-256 (2012). https://doi.org/10.1007/s00441-011-1246-y

29 Preiß, H. *et al.* Regulation of Nodal signaling propagation by receptor interactions and positive feedback. *Elife* **11** (2022). https://doi.org/10.7554/eLife.66397

30 Pomreinke, A. P. *et al.* Dynamics of BMP signaling and distribution during zebrafish dorsal-ventral patterning. *eLife* **6**, e25861 (2017). https://doi.org/10.7554/eLife.25861

31 Verstraeten, A., Alaerts, M., Van Laer, L. & Loeys, B. Marfan Syndrome and Related Disorders: 25 Years of Gene Discovery. *Hum Mutat* **37**, 524-531 (2016). https://doi.org/10.1002/humu.22977

32 Ye, D. *et al.* Insights into bone morphogenetic proteins in cardiovascular diseases. *Front Pharmacol* **14**, 1125642 (2023). https://doi.org/10.3389/fphar.2023.1125642

33 Kaplan, F. S., Al Mukaddam, M., Stanley, A., Towler, O. W. & Shore, E. M. Fibrodysplasia ossificans progressiva (FOP): A disorder of osteochondrogenesis. *Bone* **140**, 115539 (2020). https://doi.org/10.1016/j.bone.2020.115539

34 Hinck, A. P., Mueller, T. D. & Springer, T. A. Structural Biology and Evolution of the TGF-β Family. *Cold Spring Harb Perspect Biol* **8** (2016). https://doi.org/10.1101/cshperspect.a022103

35 Jaillon, O. *et al.* Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431**, 946-957 (2004). https://doi.org/10.1038/nature03025

36 Nakatani, Y., Takeda, H., Kohara, Y. & Morishita, S. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res* **17**, 1254-1265 (2007). https://doi.org/10.1101/gr.6316407

37 Session, A. M. *et al.* Genome evolution in the allotetraploid frog Xenopus laevis. *Nature* **538**, 336-343 (2016). https://doi.org/10.1038/nature19840

38 Wang, J.-T., Li, J.-T., Zhang, X.-F. & Sun, X.-W. Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (Cyprinus carpio). *BMC Genomics* **13**, 96 (2012). https://doi.org/10.1186/1471-2164-13-96

39 Gundappa, M. K. *et al.* Genome-Wide Reconstruction of Rediploidization Following Autopolyploidization across One Hundred Million Years of Salmonid Evolution. *Molecular Biology and Evolution* **39** (2021). https://doi.org/10.1093/molbev/msab310

40 Deng, Z. *et al.* TGF-β signaling in health, disease and therapeutics. *Signal Transduction and Targeted Therapy* **9**, 61 (2024). https://doi.org/10.1038/s41392-024-01764-w

41 Herpin, A., Lelong, C. & Favrel, P. Transforming growth factor-β-related proteins: an ancestral and widespread superfamily of cytokines in metazoans. *Developmental & Comparative Immunology* **28**, 461-485 (2004). https://doi.org/https://doi.org/10.1016/j.dci.2003.09.007

42 Rivera, A. M. & Swanson, W. J. The Importance of Gene Duplication and Domain Repeat Expansion for the Function and Evolution of Fertilization Proteins. *Front Cell Dev Biol* **10**, 827454 (2022). https://doi.org/10.3389/fcell.2022.827454

43 Kuhl, H. *et al.* Multi-genome comparisons reveal gain-and-loss evolution of anti-Mullerian hormone receptor type 2 as a candidate master sex-determining gene in Percidae. *Bmc Biol* **22**, 141 (2024). https://doi.org/10.1186/s12915-024-01935-9

44 Wen, M. *et al.* An ancient truncated duplication of the anti-Müllerian hormone receptor type 2 gene is a potential conserved master sex determinant in the Pangasiidae catfish family. *Molecular Ecology Resources* **22**, 2411-2428 (2022). https://doi.org/https://doi.org/10.1111/1755-0998.13620

45 Qu, M. *et al.* Seadragon genome analysis provides insights into its phenotype and sex determination locus. *Sci Adv* **7** (2021). https://doi.org/10.1126/sciadv.abg5196

46 Feron, R. *et al.* Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, Perca flavescens. *Mol Ecol Resour* **20**, 531-543 (2020). https://doi.org/10.1111/1755-0998.13133

47 Suga, H., Ono, K. & Miyata, T. Multiple TGF-β receptor related genes in sponge and ancient gene duplications before the parazoan–eumetazoan split 1. *FEBS Letters* **453**, 346-350 (1999). https://doi.org/https://doi.org/10.1016/S0014-5793(99)00749-8

48 Herpin, A. *et al.* Structural and functional evidence for a singular repertoire of BMP receptor signal transducing proteins in the lophotrochozoan Crassostrea gigas suggests a shared ancestral BMP/activin pathway. *Febs j* **272**, 3424-3440 (2005). https://doi.org/10.1111/j.1742-4658.2005.04761.x

49 Goebel, E. J., Hart, K. N., McCoy, J. C. & Thompson, T. B. Structural biology of the TGFβ family. *Exp Biol Med (Maywood)* **244**, 1530-1546 (2019). https://doi.org/10.1177/1535370219880894

50 Saito, T. *et al.* Structural Basis of the Human Endoglin-BMP9 Interaction: Insights into BMP Signaling and HHT1. *Cell Reports* **19**, 1917-1928 (2017). https://doi.org/https://doi.org/10.1016/j.celrep.2017.05.011

51 Kim, S. K., Henen, M. A. & Hinck, A. P. Structural biology of betaglycan and endoglin, membrane-bound co-receptors of the TGF-beta family. *Exp Biol Med (Maywood)* **244**, 1547-1558 (2019). https://doi.org/10.1177/1535370219881160

52 Peng, Y. *et al.* IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. *Bioinformatics* **29**, i326-i334 (2013). https://doi.org/10.1093/bioinformatics/btt219

53 He, S. *et al.* Mandarin fish (Sinipercidae) genomes provide insights into innate predatory feeding. *Commun Biol* **3**, 361 (2020). https://doi.org/10.1038/s42003-020-1094-y

54 Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Research* **12**, 656-664 (2002). https://doi.org/10.1101/gr.229202

55 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010). https://doi.org/10.1093/bioinformatics/btq033
56 Dyer, S. C. *et al.* Ensembl 2025. *Nucleic Acids Research* **53**, D948-D957 (2024). https://doi.org/10.1093/nar/gkae1071
57 Keller, O., Kollmar, M., Stanke, M. & Waack, S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**, 757-763 (2011). https://doi.org/10.1093/bioinformatics/btr010
58 Wang, J. *et al.* The conserved domain database in 2023. *Nucleic Acids Res* **51**, D384-d388 (2023). https://doi.org/10.1093/nar/gkac1096
59 Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)* **2020** (2020). https://doi.org/10.1093/database/baaa062
60 Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* **52**, W78-w82 (2024). https://doi.org/10.1093/nar/gkae268
61 Lange, M. *et al.* Zebrahub – Multimodal Zebrafish Developmental Atlas Reveals the State-Transition Dynamics of Late-Vertebrate Pluripotent Axial Progenitors. *bioRxiv*, 2023.2003.2006.531398 (2023). https://doi.org/10.1101/2023.03.06.531398
62 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529 (2021). https://doi.org/https://doi.org/10.1016/j.cell.2021.04.048
63 Hulsen, T. DeepVenn--a web application for the creation of area-proportional Venn diagrams using the deep learning framework Tensorflow. js. *arXiv preprint arXiv:2210.04597* (2022).
64 Kolberg, L. *et al.* g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research* **51**, W207-W212 (2023). https://doi.org/10.1093/nar/gkad347
65 Balasubramanian, H., Sankaran, J., Pandey, S., Goh, C. J. H. & Wohland, T. The dependence of EGFR oligomerization on environment and structure: A camera-based N&B study. *Biophys J* **121**, 4452-4466 (2022). https://doi.org/10.1016/j.bpj.2022.11.003
66 Le, Y. *et al.* Midkine-a interacts with Ptprz1b to regulate neural plate convergence and midline formation in the developing zebrafish hindbrain. *Dev Biol* **521**, 52-74 (2025). https://doi.org/10.1016/j.ydbio.2025.02.004
67 Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev Dyn* **203**, 253-310 (1995). https://doi.org/10.1002/aja.1002030302
68 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org/10.1038/s41586-021-03819-2
69 Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034 (2021). https://doi.org/10.1101/2021.10.04.463034
70 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**, 1547-1549 (2018). https://doi.org/10.1093/molbev/msy096
71 Kumar, S. *et al.* TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol Biol Evol* **39** (2022). https://doi.org/10.1093/molbev/msac174

Main Figures

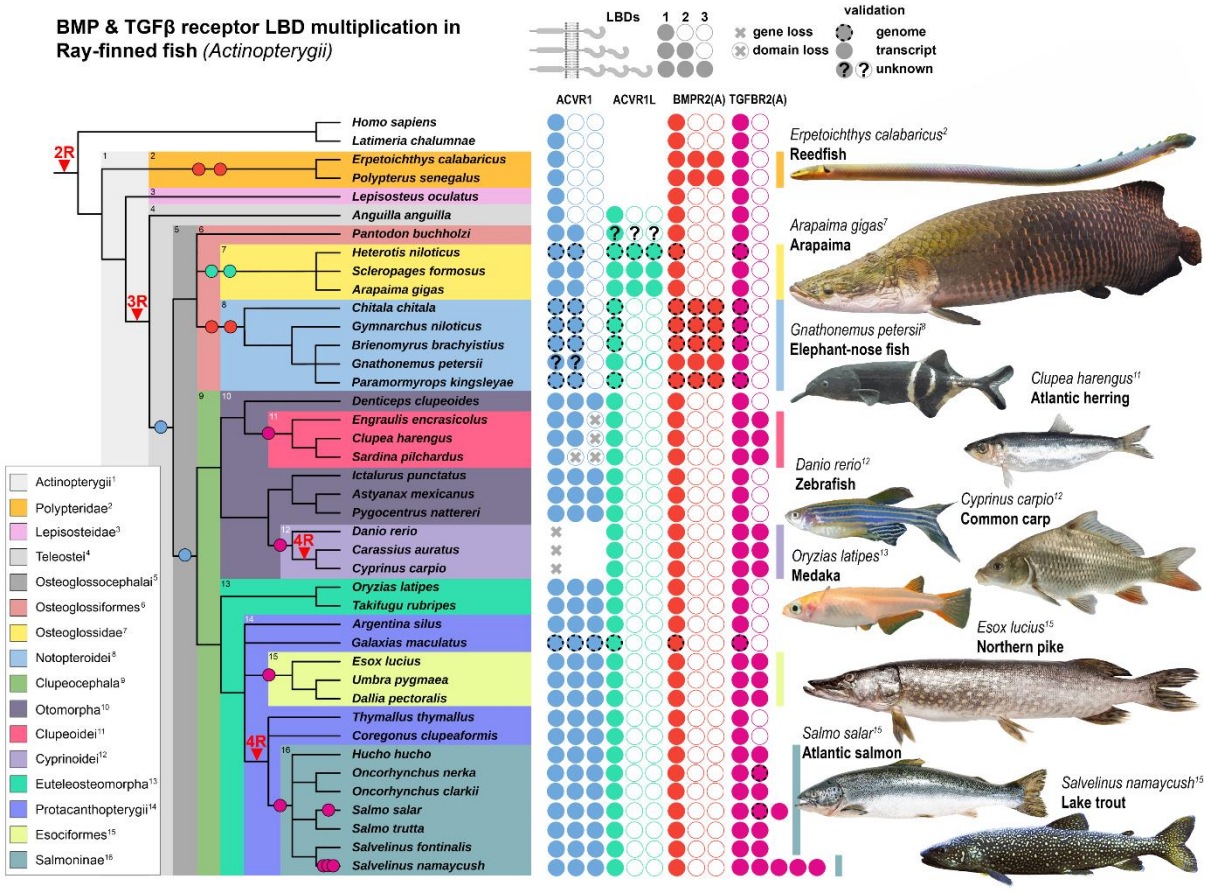


Figure 1: Convergent evolution of TGFβ & BMP-Receptor LBD multiplication in ray-finned fish. Evolutionary tree displaying the relationship between different ray-finned fish clades. Ligand Binding Domain (LBD) duplication events (circles) occurred at least 5 times in four BMP and TGFβ-Receptor genes homologs (ACVR1, ACVR1L, BMPR2 and TGFBR2). Additionally, LDB triplication events have occurred 4 times subsequently or in parallel (ACVR1, ACVR1L and BMPR2). Vertebrate-specific (2R), teleost-specific (3R) and carp-specific (4R) genome duplication events are indicated. Number of filled circles represent 1 to 3 LBDs. Transcripts are either validated by RNA-seq or predicted genes from available genomes assemblies (dashed-line).

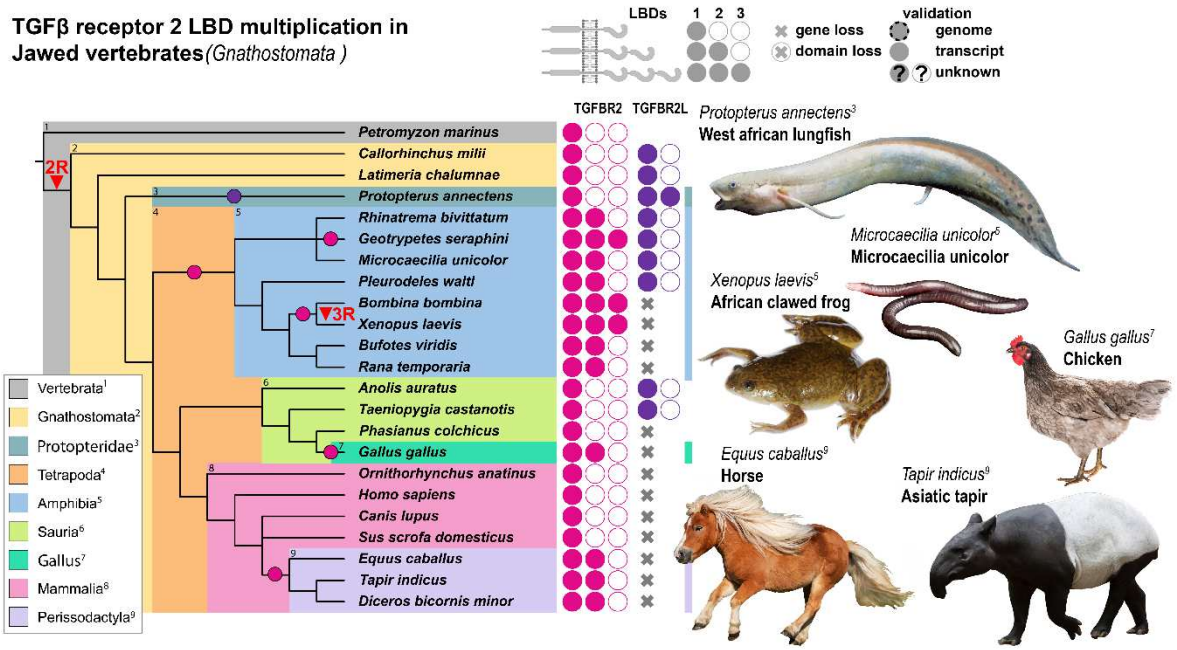


Figure 2: Convergent LBD multiplication in TGFBR2 orthologs of jawed vertebrates. Evolutionary tree displaying the relationship of jawed vertebrates. Ligand Binding Domain (LBD) duplication events (circles) occurred at least 4 times in the TGFβ-Receptor type 2 genes paralogs *TGFBR2* and *TGFBR2L*. Additionally, TGFBR2 LDB triplication events have occurred 2 times independently in different amphibian clades. Vertebrate-specific ancestral whole-genome duplication (2R) is indicated and gave rise to the paralogs *TGFBR2* and *TGFBR2L*. *TGFBR2L* has been lost in all mammals, some bird and amphibian species. Number of filled circles represent 1 to 3 LBDs. Transcripts are all validated by RNA-seq.

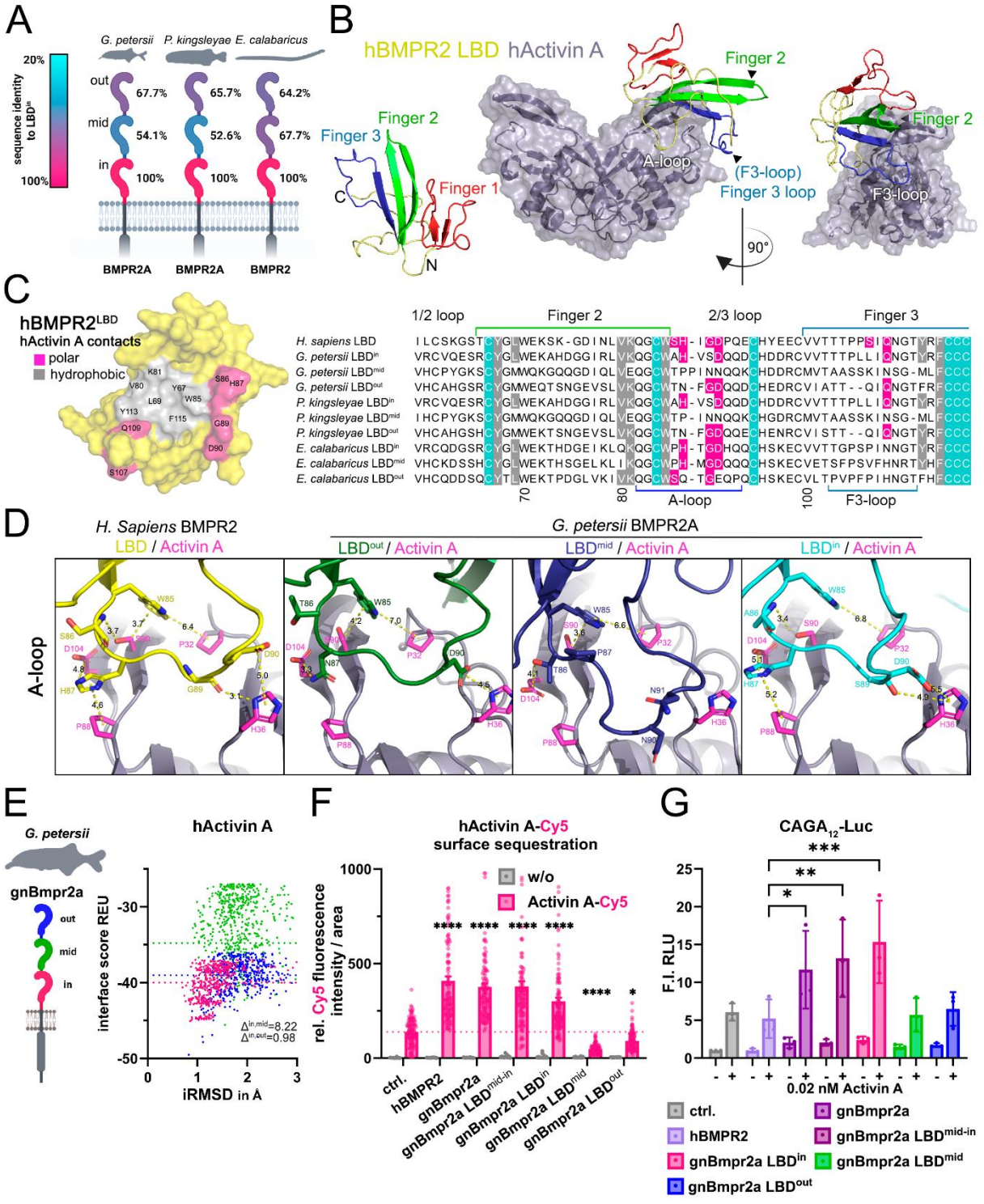


Figure 3: Inner LBD conveys signaling competence of LBD multimerized BMPR2 ortholog. (**A**) Scheme of BMPR2 LBD sequence identity comparison relative to respective inner LBD within *G. petersii*, *G. niloticus* and *E. calabaricus*. (**B**) Structure of hActivin A with the LBD interface of hBMPR2 obtained via Rosetta docking starting from an AlphaFold2-multimer prediction. Interaction is facilitated by finger motif 3 (F3-loop) and A-loop of receptor LBD. (**C**) Molecular surface of hBMPR2 LBD showing the hActivin A interface, residues forming polar or hydrophobic contacts are colored in pink and gray, respectively (**left**). Sequence alignment of BMPR2 LBDs of *H. sapiens*, *G. petersii*, *G. niloticus* and *E. calabaricus* LBDs with hActivin A contact residues and critical backbone cysteins (teal) highlighted; interaction sites F3-loop and A-loop are indicated with colored lines (**right**). (**D**) Cartoon and stick representation of respective animal receptor LBD interaction with hActivin A (A-loop) based on

AlphaFold2-multimer models PyMOL was used for image representation. (**E, left**) Illustration of *G. petersii* Bmpr2a 3 LBD receptor domain structure as reference, highlighting the inner (pink), middle (green) and outer (blue) ligand binding domain. (**E, right**) *In silico* interface analysis via Rosetta docking of hActivin A interactions with single gnBmpr2a domains, depicted as interface score (REU) in relation to interface root mean square deviation (iRMSD) in angstroms. Differences of the mean REU of out and mid LBD are calculated against inner LBD. (**F**) Activin A-Cy5 surface binding to Halo-hBMPR2 and Halo-gnBmpr2a variants is represented as relative fluorescence intensity per area. (**G**) pSMAD2/3 sensitive $CAGA_{12}$ Luciferase reporter activity of Halo-gnBmpr2a receptor variants as FI ± SD of relative Luminescence Units (RLU). Statistical significance was calculated using two-way ANOVA and Dunnett post-hoc test (**F**) relative to control (**G**) relative to hBMPR2 (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

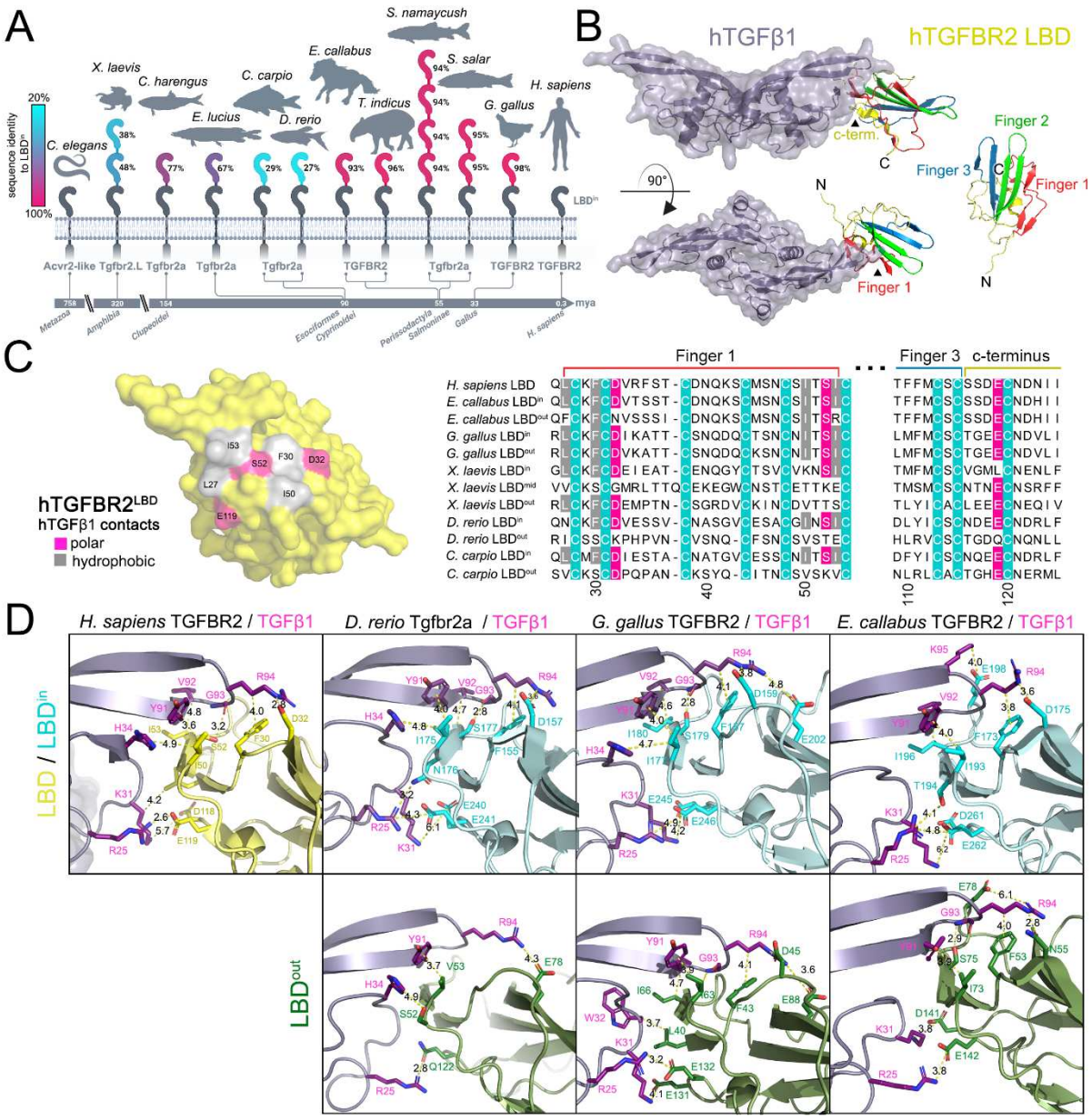


Figure 4: Multimerized LBDs of TGFBR2 orthologs display varying degree of TGFβ1-interface conservation. (**A**) Scheme of TGFBR2 LBD sequence identity comparison relative to respective inner LBD within each depicted animal. (**B**) Structure overview of hTGFβ1 with the LBD interface of hTGFBR2 obtained via Rosetta docking starting from an AlphaFold2-multimer prediction. Interaction is facilitated by c-terminus, finger motif 1 and 3 of receptor LBD. (**C**) Molecular surface of hTGFBR2 LBD showing the hTGFβ1 interface residues forming polar or hydrophobic contacts are colored in pink and gray, respectively (**left**). Sequence alignment of TGFBR2 LBDs of *Homo sapiens*, *Equus caballus*, *Gallus gallus*, *Xenopus laevis*, *Danio rerio*, *Cyprinus carpio* with TGFβ1 contact residues and critical backbone cysteins (teal) highlighted; interaction sites Finger 1, 3 and c–terminus are indicated with colored lines above (**right**). (**D**) Cartoon and stick representation of respective animal receptor LBD interaction with hTGFβ1 (c-terminus, finger motif 1 and 3) based on AlphaFold2-multimer models docked to hTGFβ1 using the lowest energy structure after performing a Rosetta docking protocol. PyMOL was used for image representation.

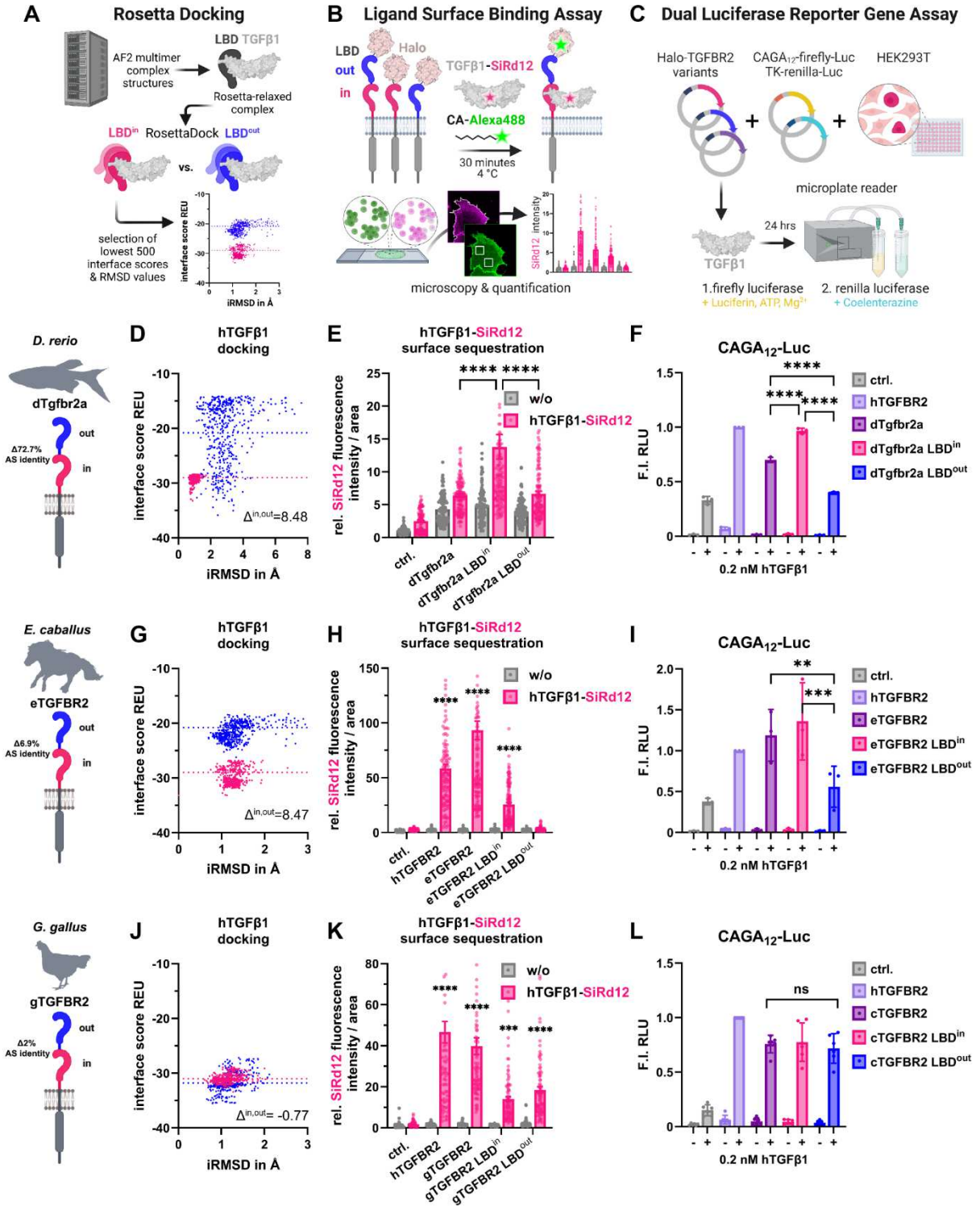


Figure 5: TGFBR2 LBD duplications differently finetune signaling competence. (**A**) Schematic representation of *in silico* binding analysis workflow using Rosetta docking. Alphafold (AF2 multimer) LBD-TGFβ1 complex structures are generated, followed by initial coordinate constrained relaxation to obtain optimized H-bonds. Docking of LBDs is then performed through RosettaScripts (n=2500). Interface score (REU) versus iRMSD plots identify the top 500 models with the lowest scores, indicating different ligand binding capabilities. (**B**) Illustration of LSBA for visualization and quantification of fluorescent hTGFβ1 binding on COS-7 cells expressing respective Halo-tagged TGFBR2 receptor constructs. (**C**) Schematic representation of dual luciferase reporter gene assay used to assess TGFβ1

signaling activity through Halo-tagged TGFBR2 receptor orthologs. HEK293T cells are co-transfected with respective receptor variant, $CAGA_{12}$-firefly-Luc, and TK-renilla-Luc plasmids. Following TGFβ1 treatment for 24 hours, Firefly luciferase activity (indicative of TGFβ1 pathway activation) and Renilla luciferase activity (normalization control) are measured. (**D. G, J**) *in silico* binding analysis via Rosetta docking of hTGFβ1 to inner (pink) and outer LBD (blue) of *D. rerio*, *E. caballus* and *G. gallus* depicted as interface score (REU) in relation to interface root mean square deviation (iRMSD) in angstroms. REU differences of LBD^{in} and LBD^{out} are calculated by respective mean REU values of each LBD. (**E, H, K**) hTGFβ1-SiR-d12 surface binding of *D.rerio*, *E. caballus* and *G. gallus* receptor variants represented as relative fluorescence intensity per area. (**F, I, L**) pSMAD2/3 sensitive $CAGA_{12}$ Luciferase reporter activity of *D. rerio*, *E. caballus* and *G. gallus* receptor variants as FI ± SD of RLU. Statistical significance was calculated using two-way ANOVA and (**F, H, I, K, L**) Tukey's post-hoc test within groups or (**E, H**) Dunnett post-hoc test relative to control (**p < 0.01, ***p < 0.001, ****p < 0.0001).

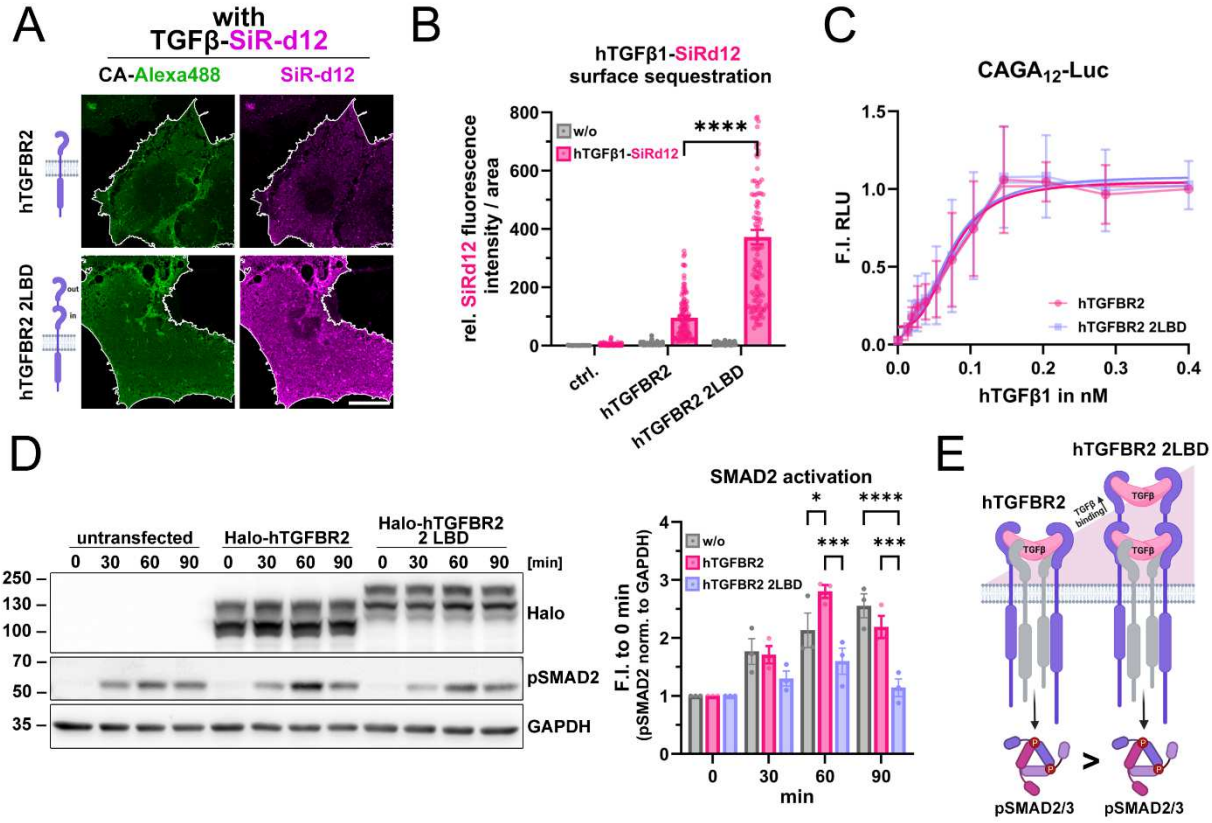


Figure 6: Human TGFBR2 LBD duplication enhances ligand binding without affecting downstream signaling. (**A**) Representative confocal microscopy images of LSBA for hTGFβ1-SiR-d12 stimulated COS-7 cells expressing hTGFBR2 or hTGFBR2-2LBD. Scale bar ≙ 20 µm. (**B**) hTGFβ1-SiR-d12 surface binding of hTGFBR2 and hTGFBR2-2LBD represented as relative fluorescence intensity per area. (**C**) pSMAD2/3 sensitive $CAGA_{12}$ Luciferase reporter activity of hTGFBR2 and hTGFBR2-2LBD transfected HEKT293 cells stimulated with increasing hTGFβ1 concentrations depicted as FI ± SD of relative Luminescence Units (RLU). (**D**, left) Immunoblot analysis of HEK293T expressing hTGFBR2 or hTGFBR2-2LBD, stimulated with 0.1 nM hTGFβ1 for 0 – 90 minutes and stained for Halo, pSMAD2 and GAPDH as loading reference, (right) densitometric quantification of pSMAD2 levels relative to GAPDH represented as F.I., n=3. (**B, D**) Statistical significance was calculated using two-way ANOVA in between (**B**) or within (**D**) groups and Tukey's post-hoc test (*p < 0.05, ***p < 0.001, ****p < 0.0001). (**E**) Summary scheme comparing hTGFBR2 and hTGFBR2-2LBD ligand binding and downstream pSMAD2/3 signaling. Increased ligand binding is not reflected on SMAD2/3 phosphorylation.

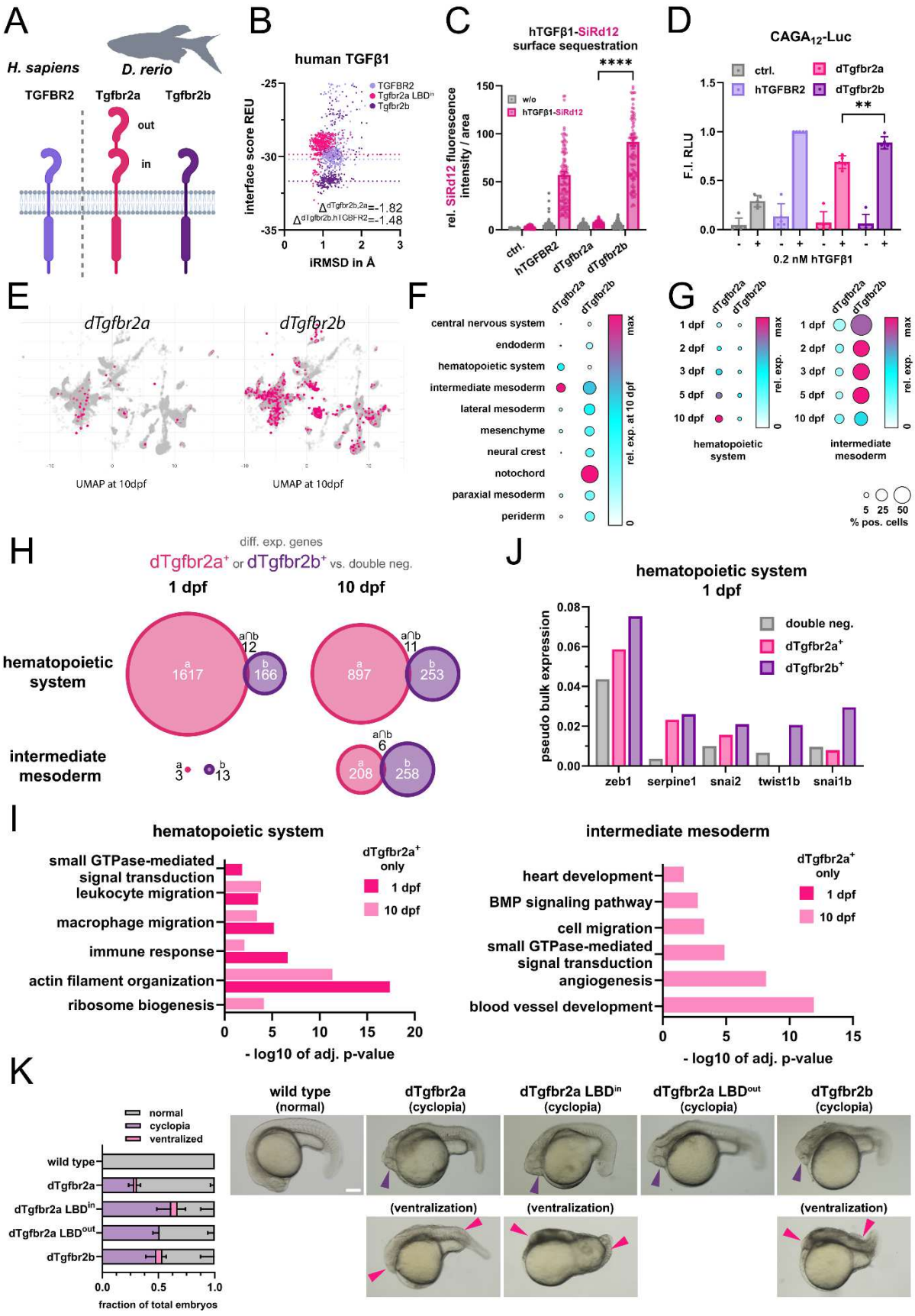
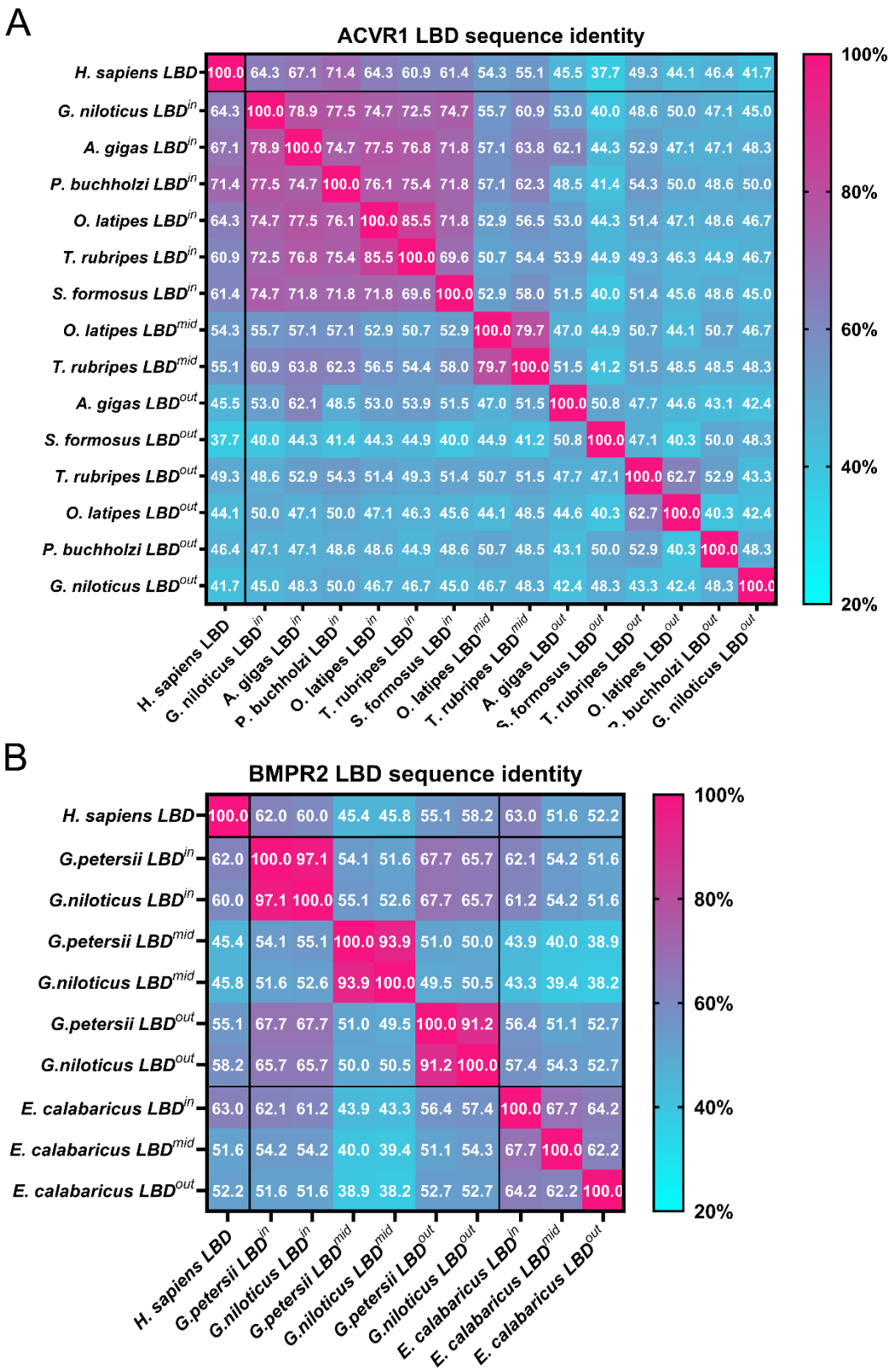


Figure 7: Duplicated LBD tgfbr2 paralog exhibits limited signaling competence and cell type specific expression pattern in zebrafish. (**A**) Schematic comparison of human hTGFBR2, *Danio rerio* dTgfbr2a and dTgfbr2b receptor. (**B**) *In silico* binding analysis via Rosetta docking of hTGFβ1 to inner LBD of dTgfbr2a (pink), dTgfbr2b LBD (purple) and

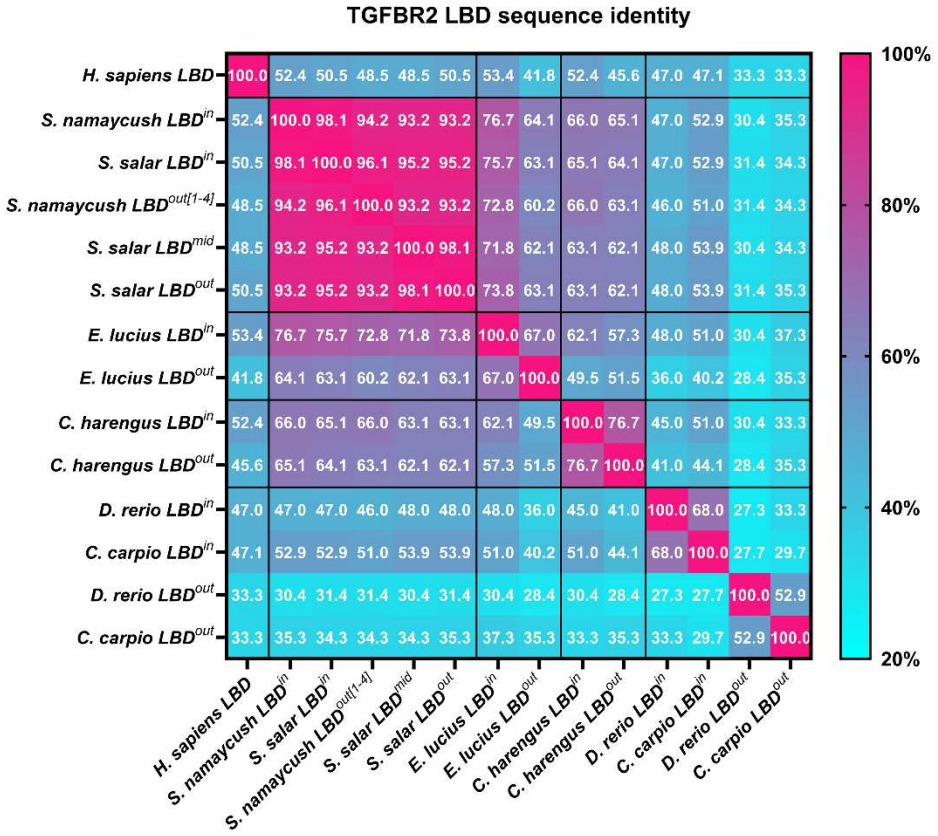
hTGFBR2 (lavender) depicted as REU in relation to iRMSD in angstroms. REU differences are calculated by respective mean REU values of dTgfbr2b to dTgfbr2a or hTGFBR2. (**C**) hTGFβ1-SiR-d12 surface binding of hTGFBR2, dTgfbr2a and dTgfbr2b represented as relative fluorescence intensity per area. (**D**) pSmad2/3 sensitive $CAGA_{12}$ Luciferase reporter activity of dTgfbr2b to dTgfbr2a or hTGFBR2 as FI ± SD of relative Luminescence Units (RLU). Significance was calculated using two-way ANOVA and Tukey's post-hoc test within groups (**p < 0.01, ****p < 0.0001). (**E**) UMAP representation of *D. rerio* transcriptome at 10 dpf, highlighting dTgfbr2a and dTgfbr2b positive cells (pink). (**F**) relative expression of dTgfbr2a and dTgfbr2b at 10 dpf in respective tissue; size of circle represents percentage of positive cells (**G**) relative expression of dTgfbr2a and dTgfbr2b in hematopoietic system and intermediate mesoderm at different developmental stages; size of circle represents percentage of positive cells (**H**) Venn diagrams of differentially expressed genes in dTgfbr2a (pink) and dTgfbr2b (purple) positive cells compared to double-negative controls at 1 and 10 dpf in hematopoietic system and intermediate mesoderm (**J**) Pseudo bulk expression of known SMAD2/3 target genes in dTgfbr2a, dTgfbr2b positive and double negative cells of the hematopoietic system at 1 dpf (**I**) Go term analysis of dTgfbr2a positive cells at 1 and 10 dpf of the hematopoietic system and intermediate mesoderm. (**K**) Overexpression of dTgfbr2a, dTgfbr2a-LBD^{in}, dTgfbr2a-LBD^{out} and dTgfbr2b in zebrafish embryos at 22 hpf. (left) Fraction of embryos showing normal, cyclopia or ventralized phenotypes after overexpression of indicated receptor constructs. (right) Representative brightfield images of embryos exhibiting indicated phenotypes. Purple arrowheads indicate fused eye structure in embryos with cyclopia phenotype. Pink arrowheads point to the reduction of head and neural tube structures in ventralized embryos. Scale bar = 100 µm.

Supplemental Figures:

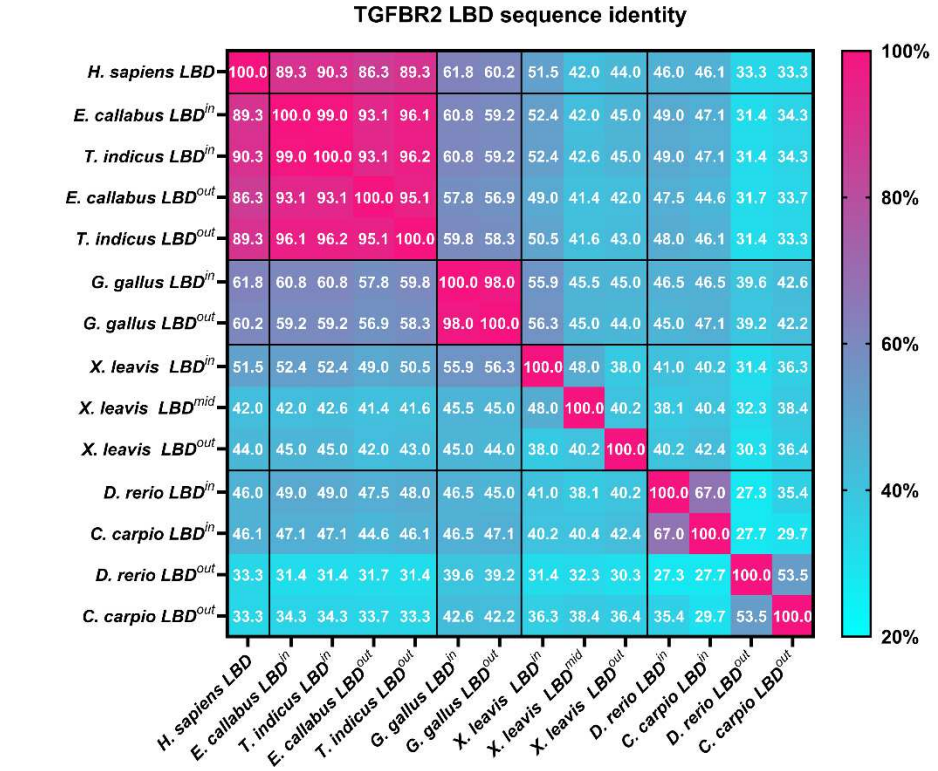


Supplemental Figure 1: LBD sequence identity comparison of ACVR1 and BMPR2 orthologs. (**A**) ACVR1 LBD sequence identity matrix comparing all LBD ACVR1 variants of indicated species. (**B**) BMPR2 LBD sequence identity matrix comparing all LBD BMPR2 variants of indicated species with each other.

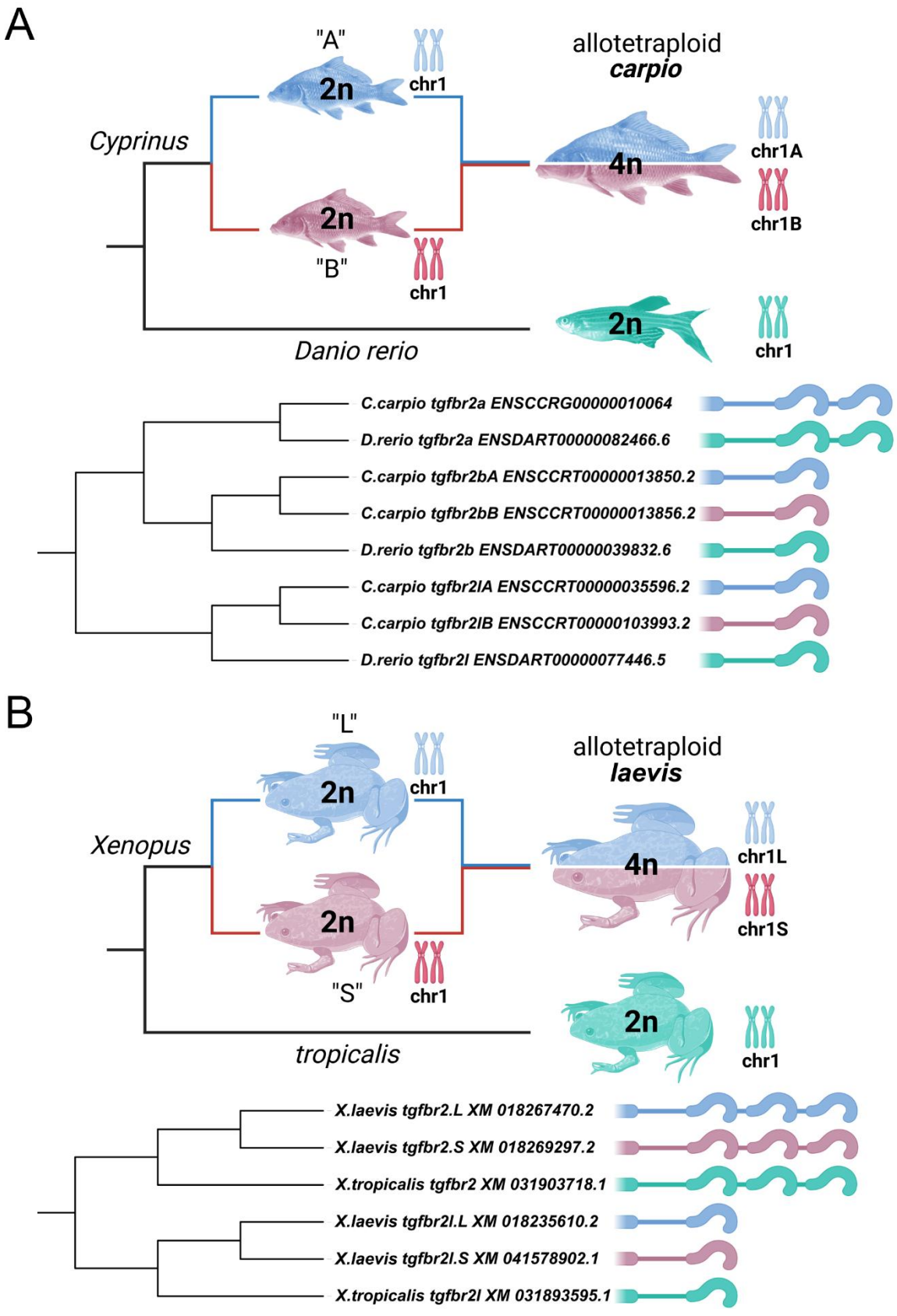
A



B

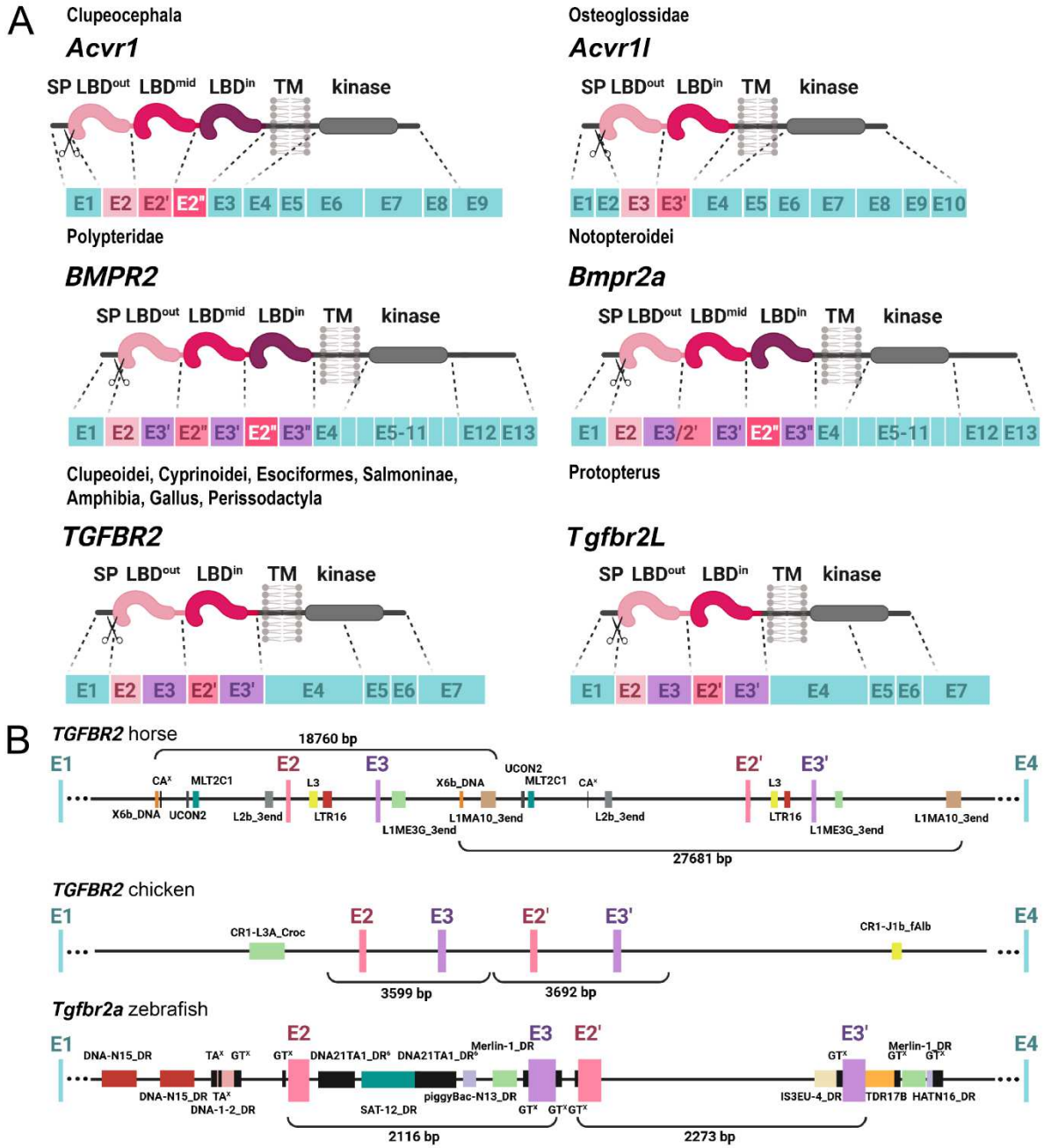


Supplemental Figure 2: LBD sequence identity comparison of TGFBR2 orthologs. (**A**) TGFBR2 LBD sequence identity matrix comparing all LBD TGFBR2 variants of human and indicated ray finned fish species. (**B**) TGFBR2 LBD sequence identity matrix comparing all LBD TGFBR2 variants of indicated jaw vertebrate as well as Cyprinoidei species.



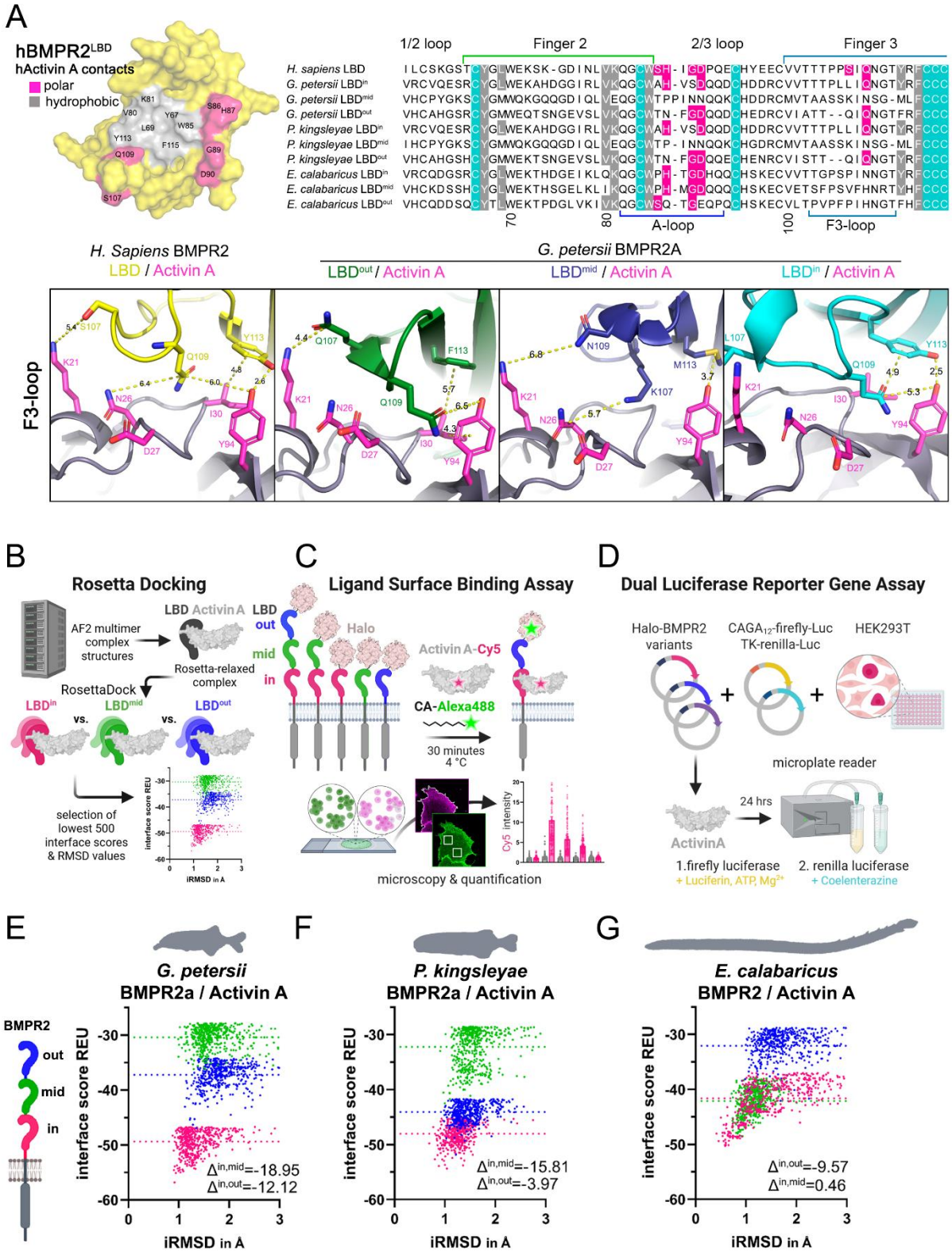
Supplemental Figure 3: TGFBR2 LBD duplication independent of allotetraploid chromosomal evolution in *Cyprinus carpio* and *Xenopus laevis*. (**A, top**) Phylogenetic depiction of chromosomal evolution in the Cyprinus lineage, illustrating the allopolyploid origin

of *Cyprinus carpio* (4n) from two diploid ancestors ("A" and "B"). Each ancestor contributed a set of chromosomes (chr1A in blue, chr1B in red) leading to a tetraploid *C. carpio* with both ancestral chromosome sets. *Danio rerio* is depicted with its diploid structure (2n), maintaining a single chromosome set (chr1). (**Bottom)** Phylogenetic tree showing the relationship of tgfbr2a gene variants between *C. carpio* and *D. rerio*. Structural variations in ligand-binding domains are illustrated beside each gene. (**B, top**) Phylogenetic depiction of chromosomal evolution in the Xenopus lineage, illustrating the allopolyploid origin of *Xenopus laevis* (4n) from two diploid ancestors ("L" and "S"). Each ancestor contributed a set of chromosomes (chr1L in blue, chr1S in red) leading to a tetraploid *X. laevis* with both ancestral chromosome sets. *Xenopus tropicalis* is depicted with its diploid structure (2n), maintaining a single chromosome set (chr1). (**Bottom)** Phylogenetic tree showing the relationship of tgfbr2 gene variants between X. *laevis and X. tropicalis.* Structural variations in ligand-binding domains are illustrated beside each gene.



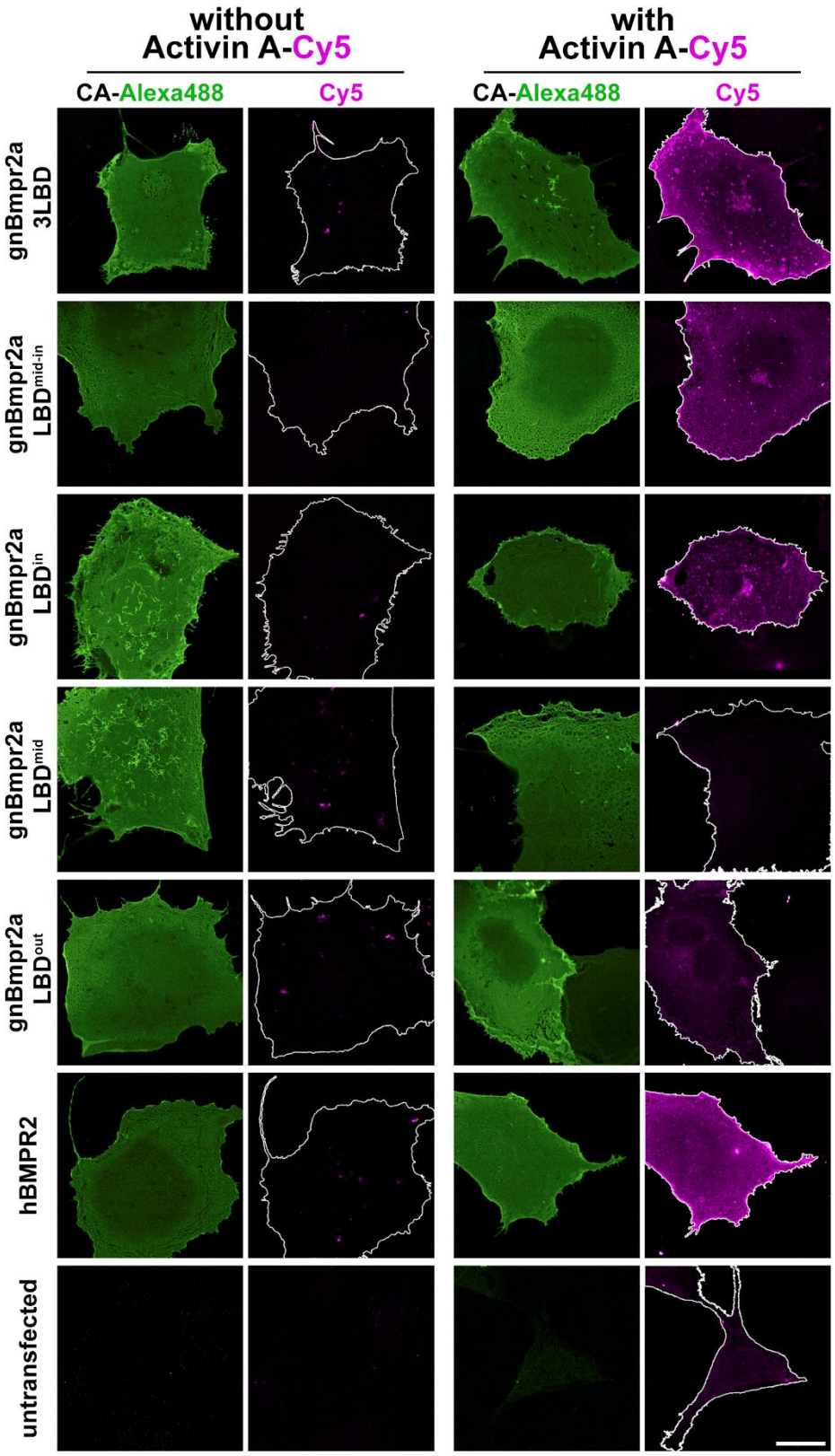
Supplemental Figure 4: Genetic architecture of duplicated LBDs of TGFβ receptor family members. (**A**) Schematic representation of receptor gene exon composition of Acvr1, Acvr1L,

BMPR2, Bmpr2a, TGFBR2 and TGFBR2L. (**B**) Genomic organization of TGFBR2 intronic sites flanking exons encoding in horse, chicken, and Tgfbr2a in zebrafish, including annotations size and common repetitive elements.

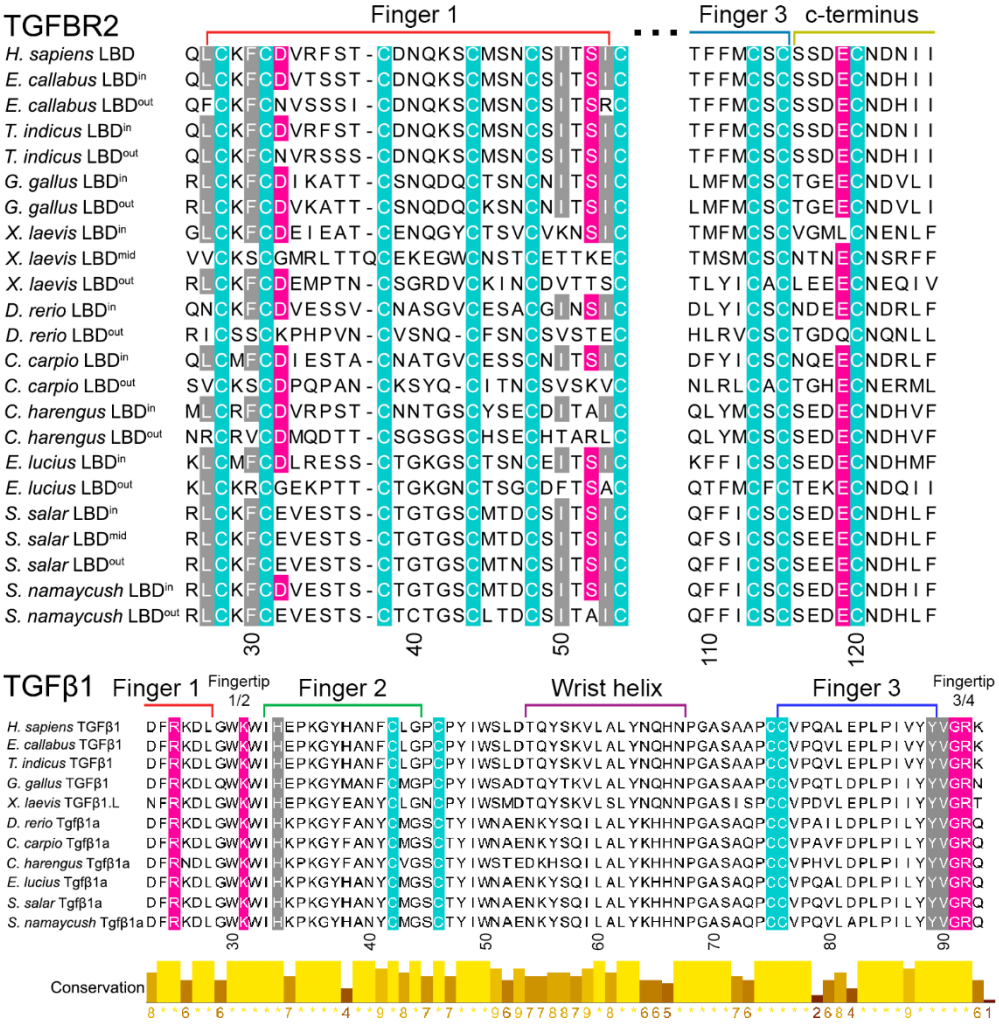


Supplemental Figure 5: Inner LBD conveys Activin A binding competence of LBD multimerized BMPR2 orthologs. (**A**) Molecular surface of hBMPR2 LBD showing the hActivin A interface of a rosetta-docked AlphaFold2-multimer structure prediction. Residues forming polar or hydrophobic contacts are colored in pink and gray, respectively (**left**). Sequence

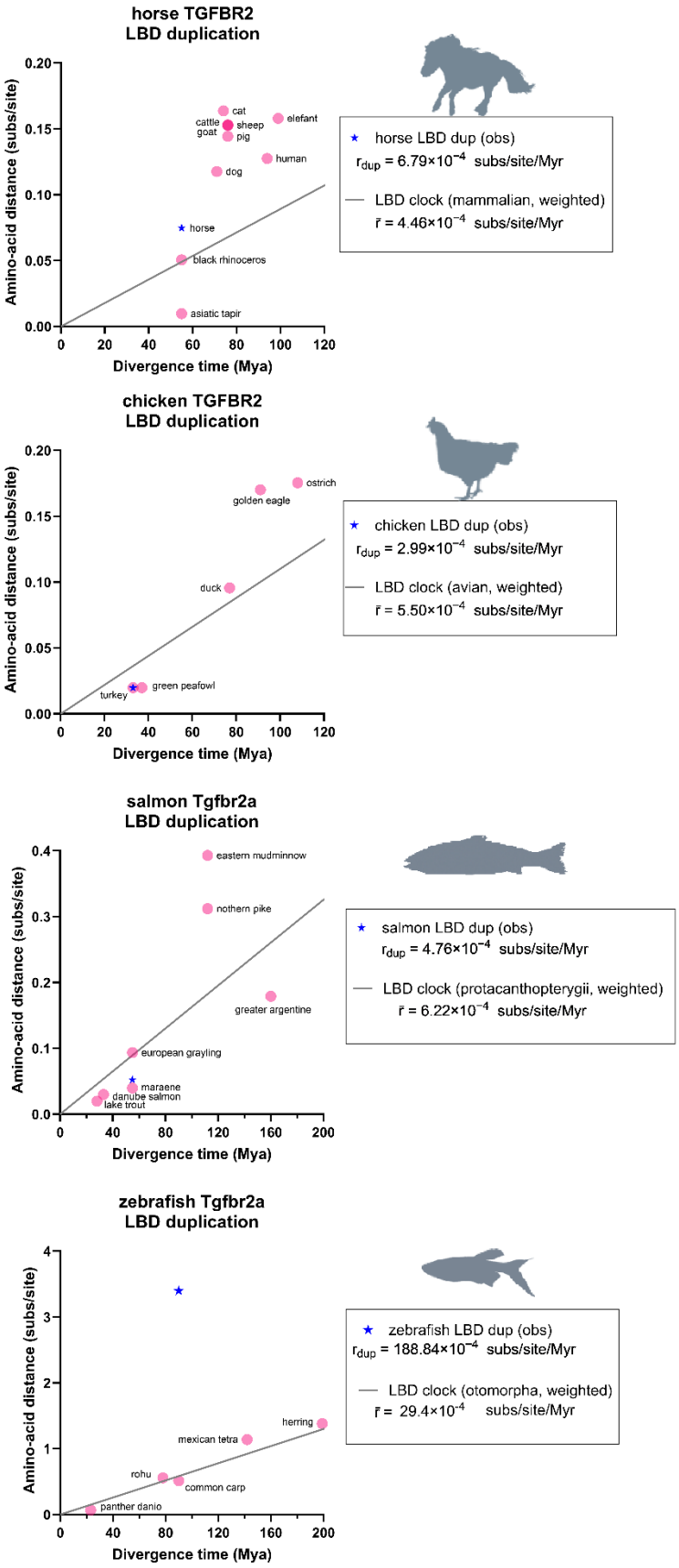
alignment of BMPR2 LBDs of *Homo Sapiens*, *G. petersii*, *G.niloticus* and *E. calabaricus* LBDs with hActivin A contact residues and critical backbone cysteins (teal) highlighted; interaction sites F3-loop and A-loop are indicated with colored lines (**right**). (**A, lower**) Cartoon and stick representation of respective animal receptor LBD interaction with hActivin A (A-loop) based on AlphaFold2-multimer models docked to hActivin A using Rosetta docking protocols. PyMOL was used for image representation. (**B**) Schematic representation of *in silico* binding analysis workflow using Rosetta docking. Alphafold (AF2 multimer) LBD-Activin A complex structures are generated, followed by initial coordinate constrained relaxation to obtain optimized H-bonds. Docking of LBDs is then performed through RosettaScripts (n=2500). Interface score (REU) versus iRMSD plots identify the top 500 models with the lowest scores, indicating different ligand binding capabilites. (**C**) Illustration of LSBA for visualization and quantification of fluorescent hActivin A binding on COS-7 cells expressing respective Halo-tagged BMPR2 receptor constructs. (**D**) Schematic representation of dual luciferase reporter gene assay used to assess Activin A signaling activity through Halo-tagged BMRP2 receptor orthologs. HEK293T cells are co-transfected with respective receptor variant, $CAGA_{12}$-firefly-Luc, and TK-renilla-Luc plasmids. Following Activin A treatment for 24 hours, Firefly luciferase activity (indicative of Activin A pathway activation) and Renilla luciferase activity (normalization control) are measured. *In silico* binding analysis of (**E**) *G. petersii* (**F**) *P. kingsleyae* (**G**) *E. calabaricus* Activin A interactions with its respective single Bmpr2a domains obtained via Rosetta Docking, depicted as interface score (REU) in relation to interface root mean square deviation (iRMSD) in angstroms. (**right**) Illustration of respective animal BMPR2a 3 LBD receptor domain structure as reference, highlighting the inner (pink), middle (green) and outer (blue) ligand binding domain. REU differences of out and mid LBD mean are calculated against inner LBD mean value.



Supplemental Figure 6: Representative LSBA images of Activin A bound to gnBmpr2a variants. Representative confocal microscopy images of LSBA for unstimulated and hActivinA-Cy5 stimulated COS-7 cells expressing *Gnathonemus petersii* receptor variants (gnBmpr2a (3LBD), gnBmpr2a-LBD^{midin}, gnBmpr2a-LBD^{in}, gnBmpr2a-LBD^{mid}, gnBmpr2a-LBD^{out}) or hBMPR2 or untransfected control. Scale bar ≙ 20 µm.

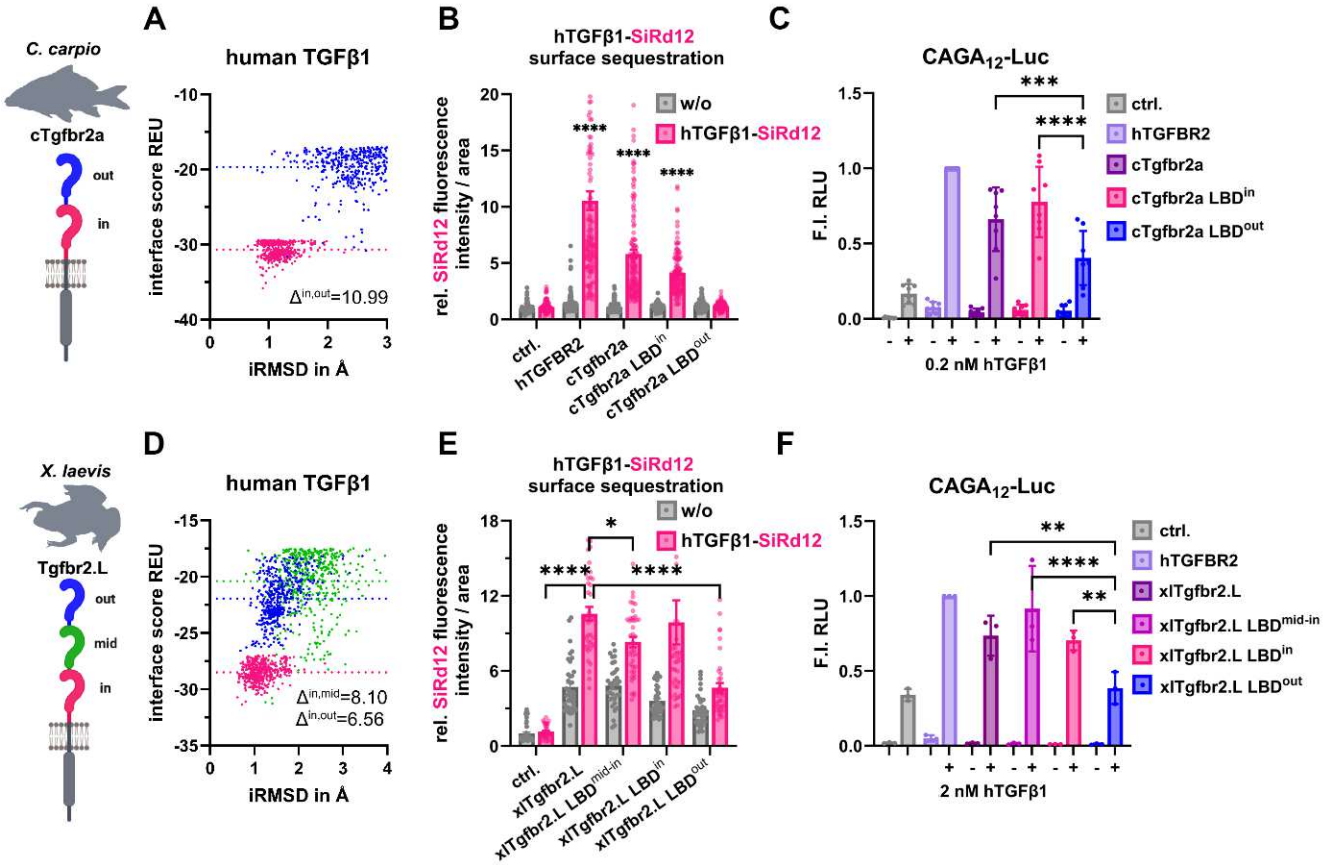


Supplemental Figure 7: Sequence alignment of multimerized TGFBR2 LBDs and TGFβ1 orthologs. TGFBR2 LBD and TGFβ1 ligand sequence alignment of *H. sapiens*, *E. callabus*, *T. indicus*, *G. Gallus*, *X. laevis*, *D. rerio*, *C. carpio*, *C. harengus*, *E. lucius*, *S. salar*, *S. namaycush*; polar receptor:ligand interaction sites are shown in light magenta and hydrophobic interaction sites in grey and highly conserved backbone cysteins in teal. Interaction domains are indicated above.

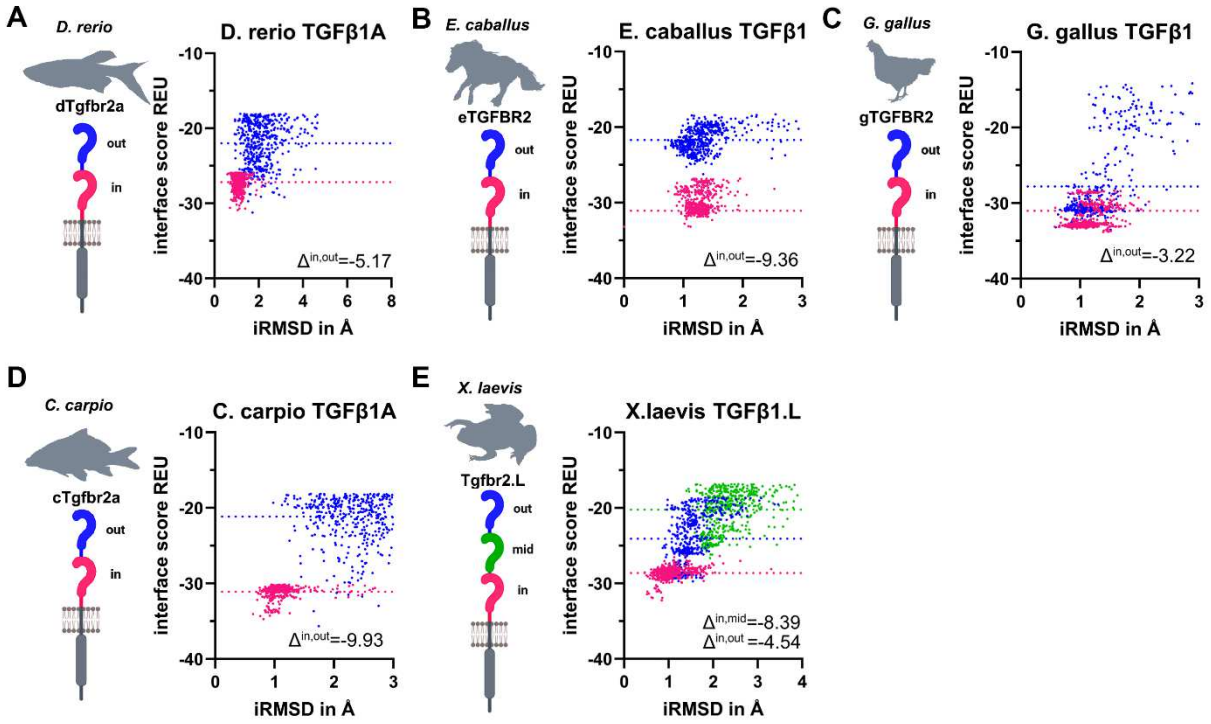


Supplemental Figure 8: Evolutionary divergence of duplicated LBD domains relative to species divergence times. Each panel shows the relationship between amino-acid distance and species divergence time for duplicated LBDs of TGFBR2/Tgfbr2a in four focal species:

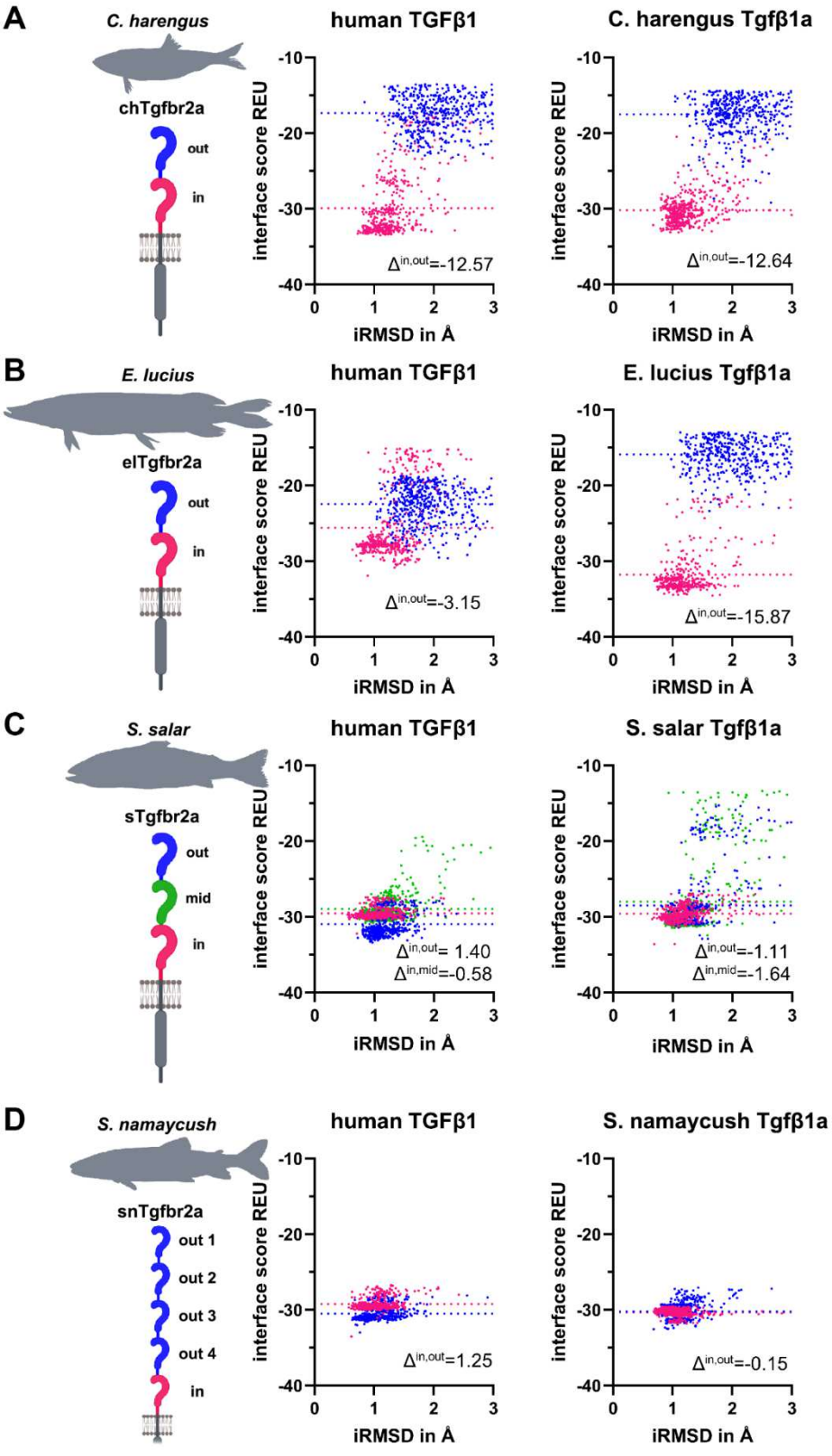
horse, chicken, salmon, and zebrafish. Pink circles represent pairwise comparisons between the inner LBD of the focal species and the orthologous inner LBD of other species in the same clade. The gray line depicts the clade-specific molecular clock, obtained by weighted linear regression of amino-acid distance against species divergence time; the resulting substitution rate ($\bar{r}$), expressed as substitutions per site per million years, is shown in each panel. The blue star marks the observed amino-acid divergence between the two duplicated LBD copies within the focal species, and the duplication-specific evolutionary rate ($r_{(}dup_{)}$) is reported in each panel.



Supplemental Figure 9: Functional implications of LBD multiplication in TGFBR2 orthologs of *Cyprinus carpio* and *Xenopus laevis*. (**A, D)** *in silico* binding analysis via Rosetta docking of hTGFβ1 to inner (pink) and outer LBD (blue) of *C. carpio*, *X. laevis* depicted as interface score (REU) in relation to interface root mean square deviation (iRMSD) in angstroms. REU differences of in and out are calculated by respective mean REU values of each LBD. (**B, E**) hTGFβ1-SiR-d12 surface binding of *C. carpio*, *X. laevis* receptor variants represented as relative fluorescence intensity per area. (**C**, **F,**) pSmad2/3 sensitive $CAGA_{12}$ Luciferase reporter activity of *C. carpio*, *X. laevis* receptor variants as FI ± SD of relative Luminescence Units (RLU). Statistical significance was calculated using two-way ANOVA and (**B,E**) Dunnett post-hoc test relative to control (**B**) or xlTgfbr2.L (**E**) or using two-way ANOVA and (**C,F**) Tukey's post-hoc test within groups (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

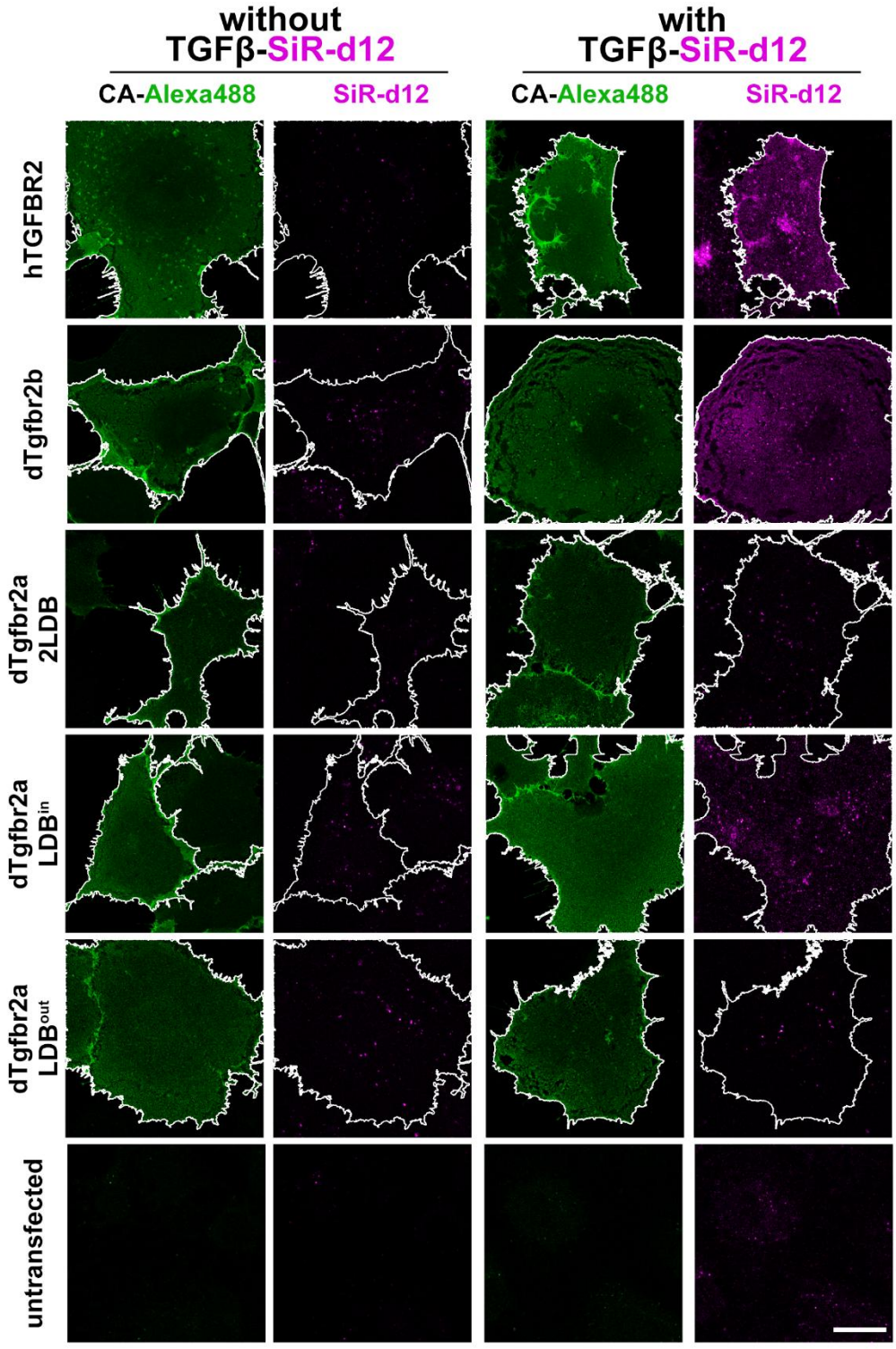


Supplemental Figure 10: *In silico* binding analysis of TGFBR2 orthologs with respective animal TGFβ1. *In silico* binding analysis via Rosetta docking and illustration of respective animal TGFBR2 LBD receptor domain structure as color reference of (**A**) *Danio rerio* (**B**) *Equus callabus* (**C**) *Gallus gallus* (**D**) *Cyprinus carpio* (**E**) *Xenopus laevis* TGFβ1 interactions with its respective single TGFBR2/Tgfbr2a/Tgfbr2.L domains, depicted as interface score (REU) in relation to interface root mean square deviation (iRMSD) in angstroms. REU differences of out and mid LBD mean are calculated against inner LBD mean value.

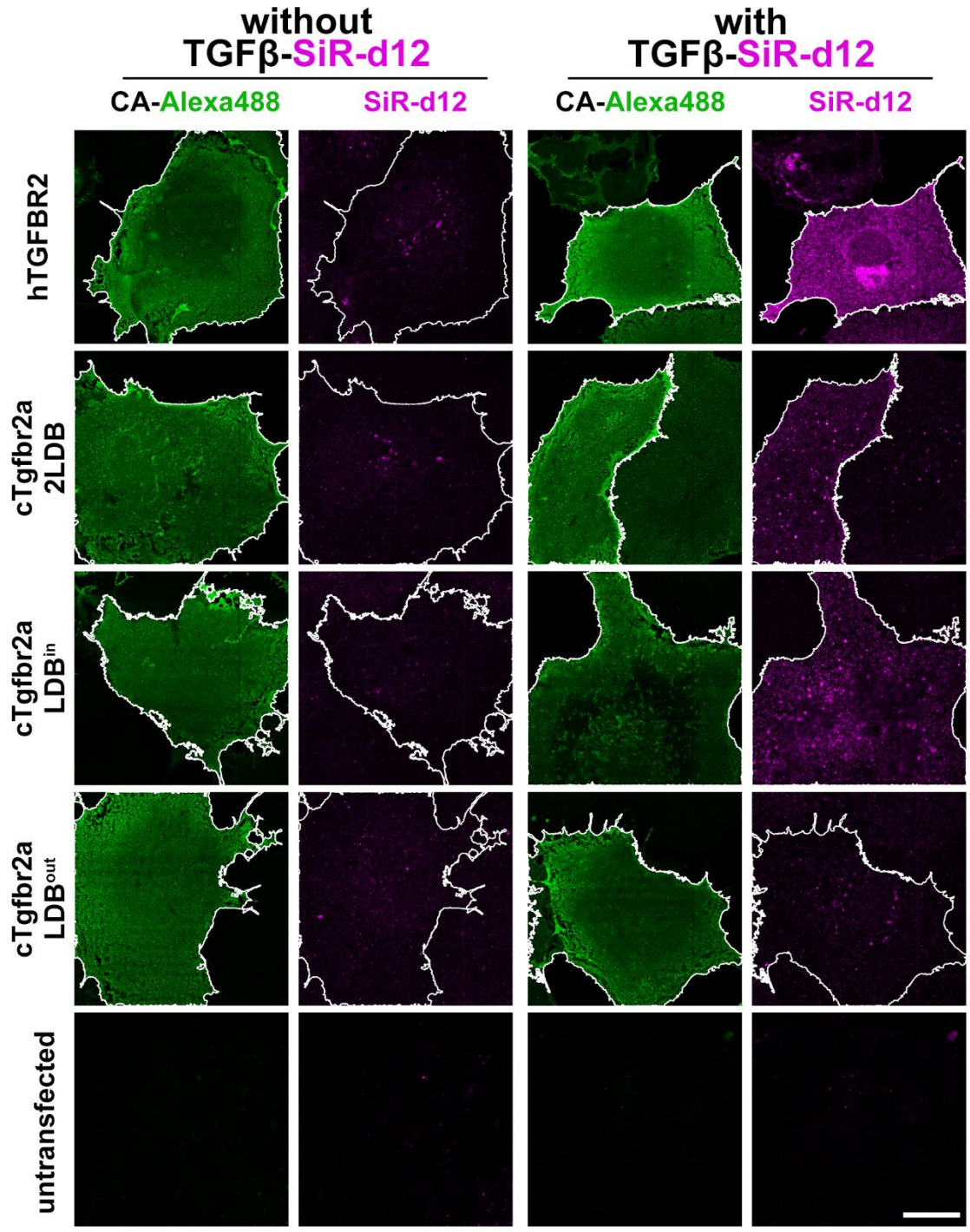


Supplemental Figure 11: *In silico* binding analysis of ray-finned fish TGFBR2 orthologs with human and respective animal TGFβ1. Illustration of respective animal TGFBR2 LBD receptor domain structure as color reference and Rosetta docking via *in silico* binding analysis of (**A**) *Clupea harengus* (**B**) *Esox lucius* (**C**) *Salmo salar* (**D**) *Salvelinus namaycush* TGFβ1 *and* hTGFβ1 interactions with respective single Tgfbr2a domains, depicted as interface score

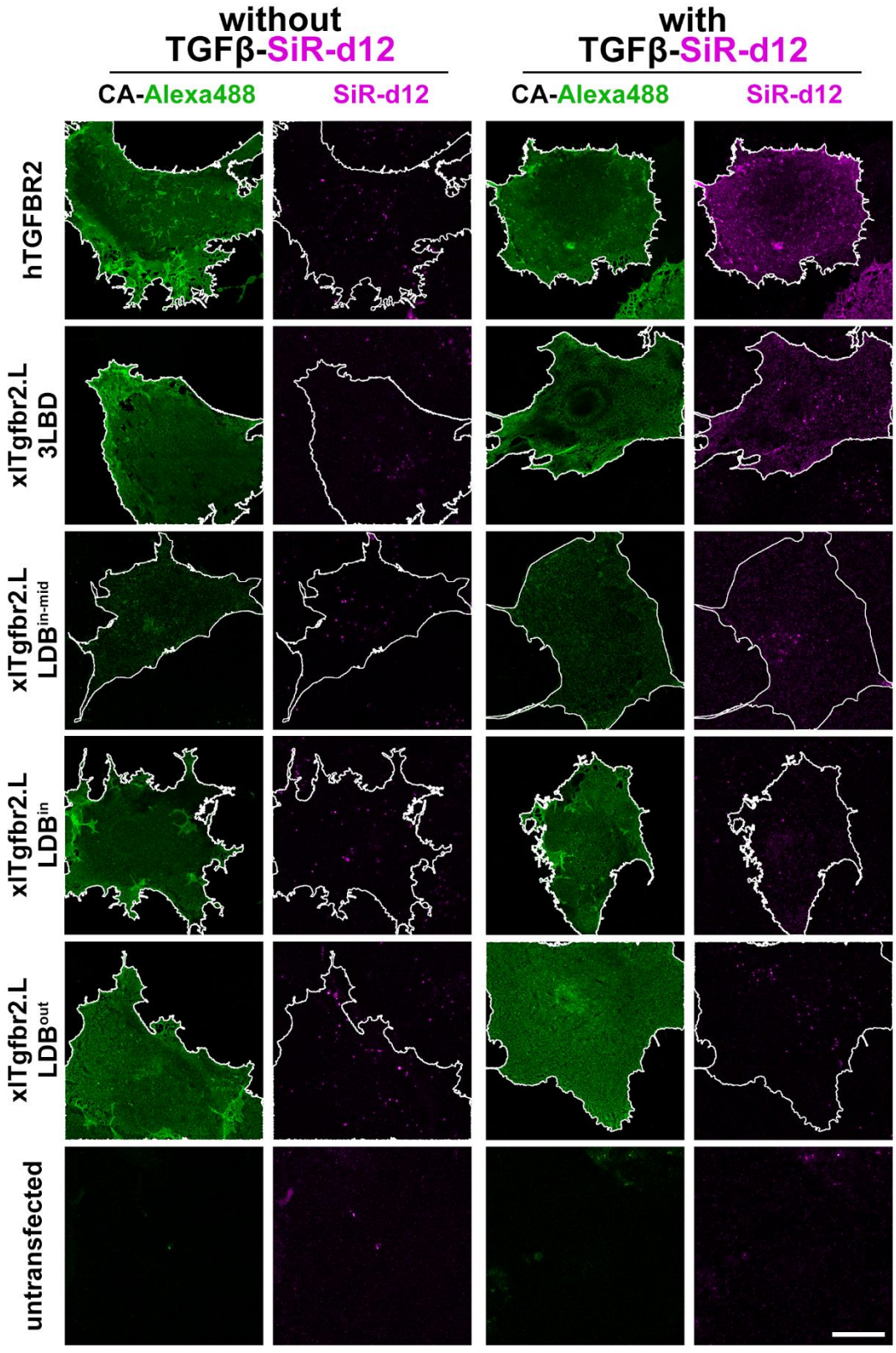
(REU) in relation to interface root mean square deviation (iRMSD) in angstroms. REU differences of out and mid LBD mean are calculated against inner LBD mean value.



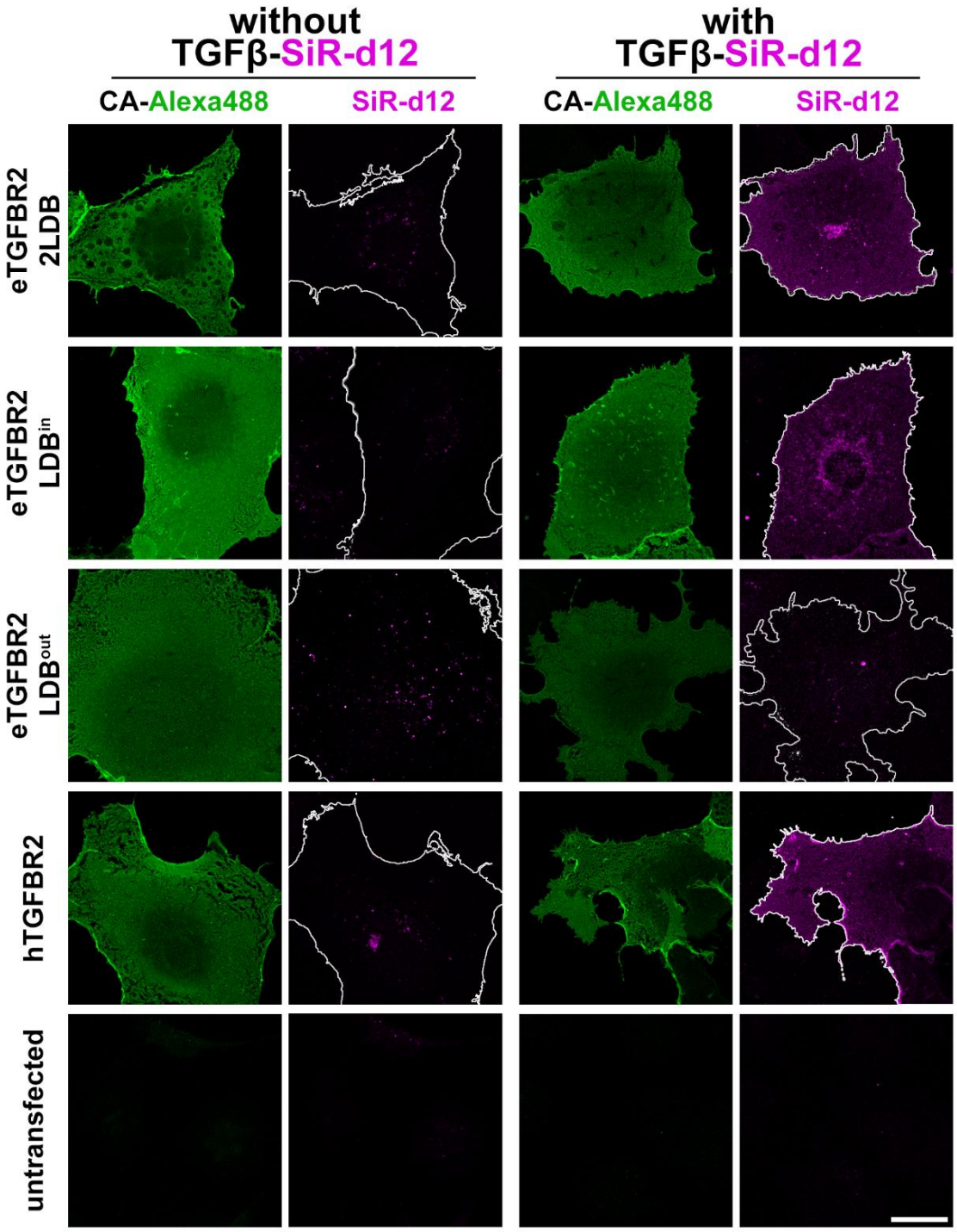
Supplemental Figure 12: Representative LSBA images of TGFβ1 bound to dTgfbr2a/b variants. Representative confocal microscopy images of LSBA for unstimulated and hTGFβ1-SiR-d12 stimulated COS-7 cells expressing human BMPR2 or *danio rerio* variants (dTgfbr2b, dTgfbr2a (2LBD), dTgfbr2a-LBD^{in}, dTgfbr2a-LBD^{out}) or untransfected control. Scale bar ≙ 20 µm.



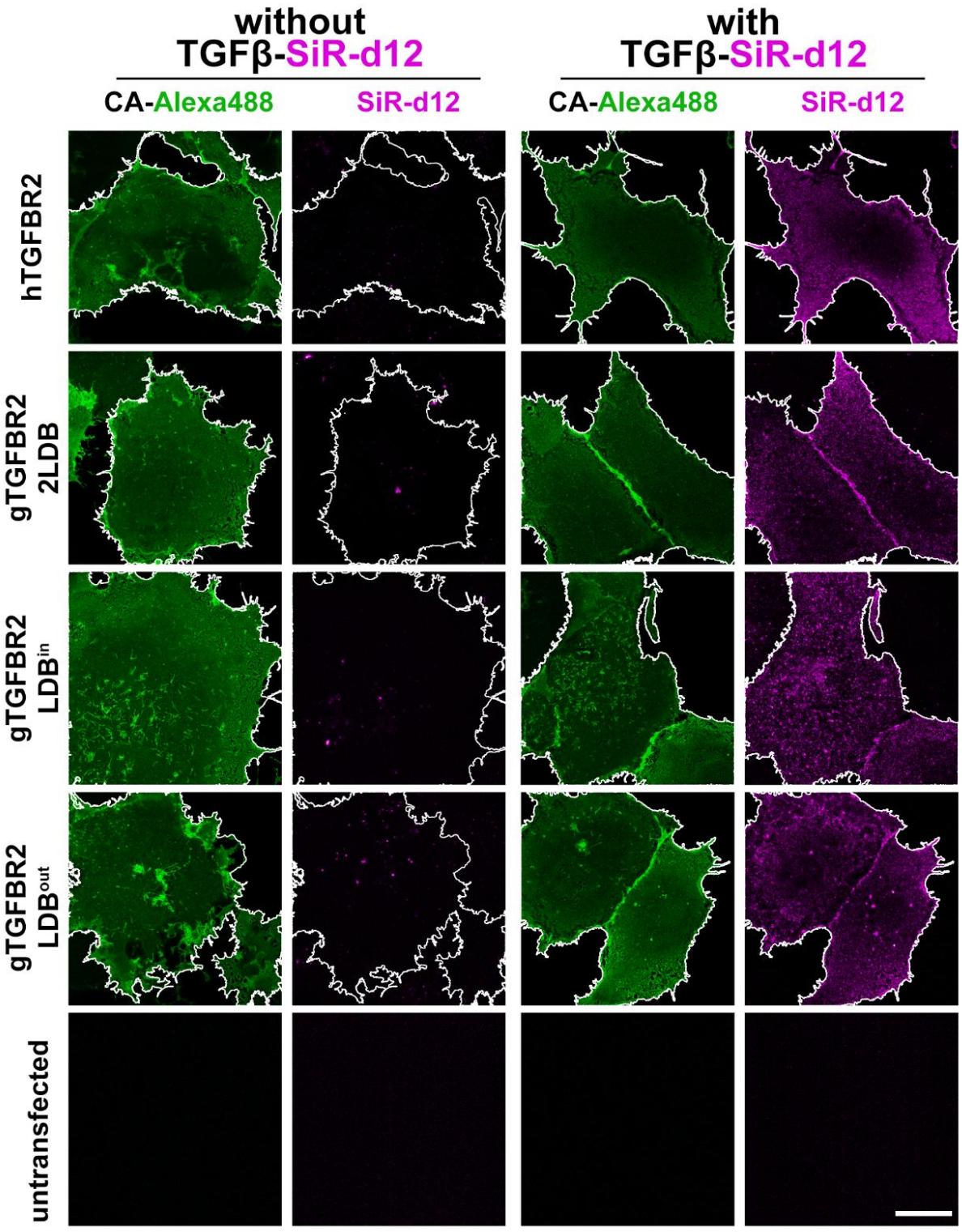
Supplemental Figure 13: Representative LSBA images of TGFβ1 bound to cTgfbr2a variants. Representative confocal microscopy images of LSBA for unstimulated and hTGFβ1-SiR-d12 stimulated COS-7 cells expressing human TGFBR2 or *Cyprinus carpio* variants (cTgfbr2a (2LBD), cTgfbr2a-LBD^{in}, cTgfbr2a-LBD^{out}) or untransfected control. Scale bar ≙ 20 µm.



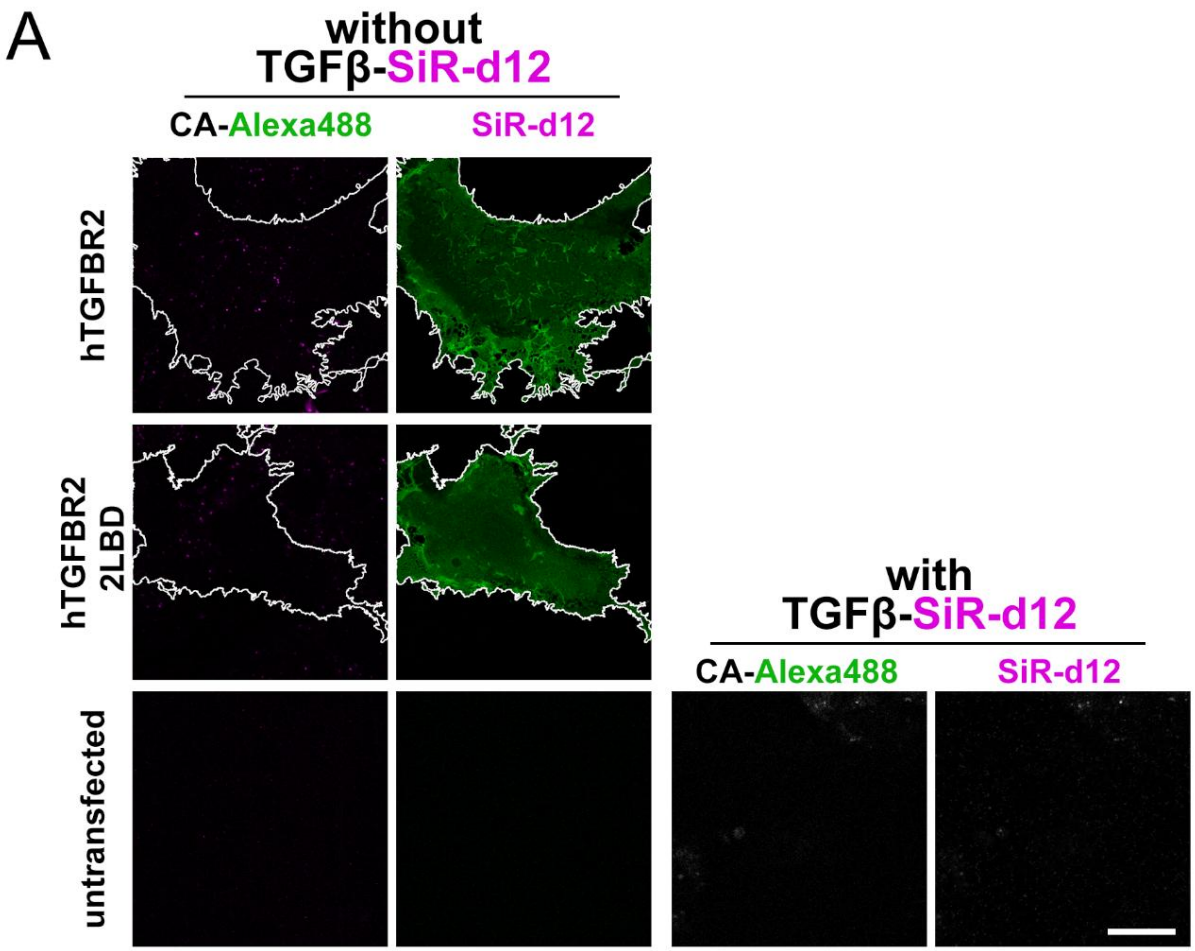
Supplemental Figure 14: Representative LSBA images of TGFβ1 bound to xlTgfbr2.L variants. Representative confocal microscopy images of LSBA for unstimulated and hTGFβ1-SiR-d12 stimulated COS-7 cells expressing human TGFBR2 or *Xenopus laevis* variants (xlTgfbr2.L (3LBD), xlTgfbr2.L-LBD^{mid-in}, xlTgfbr2.L-LBD^{in}, xlTgfbr2.L-LBD^{out}) or untransfected control. Scale bar ≙ 20 µm.



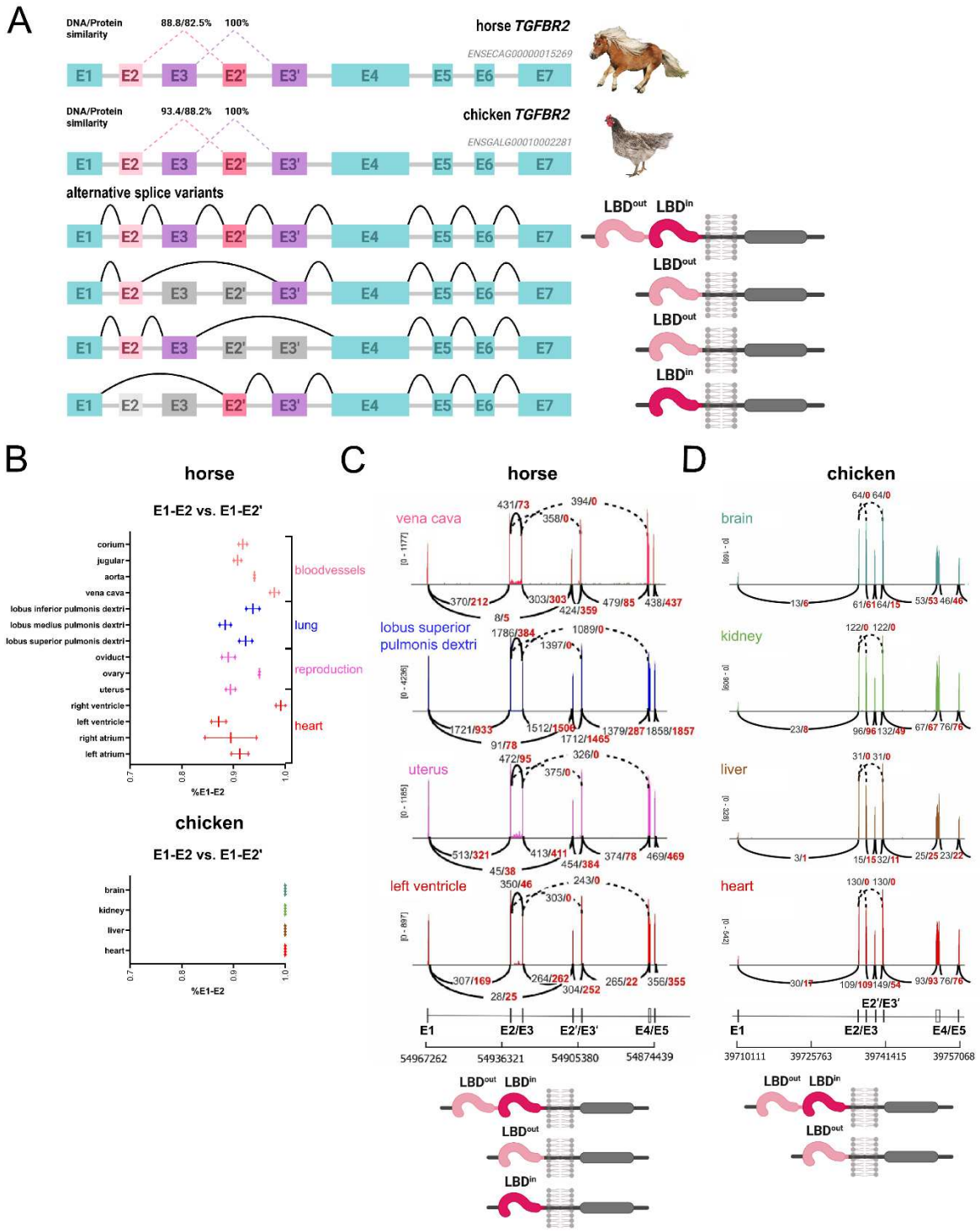
Supplemental Figure 15: Representative LSBA images of TGFβ1 bound to eTGFBR2 variants. Representative confocal microscopy images of LSBA for unstimulated and hTGFβ1-SiR-d12 stimulated COS-7 cells expressing *Equus caballus* variants *(*eTGFBR2 (2LBD), eTGFBR2-LBD^{in}, eTGFBR2-LBD^{out}) and human TGFBR2 or untransfected control. Scale bar ≙ 20 µm.



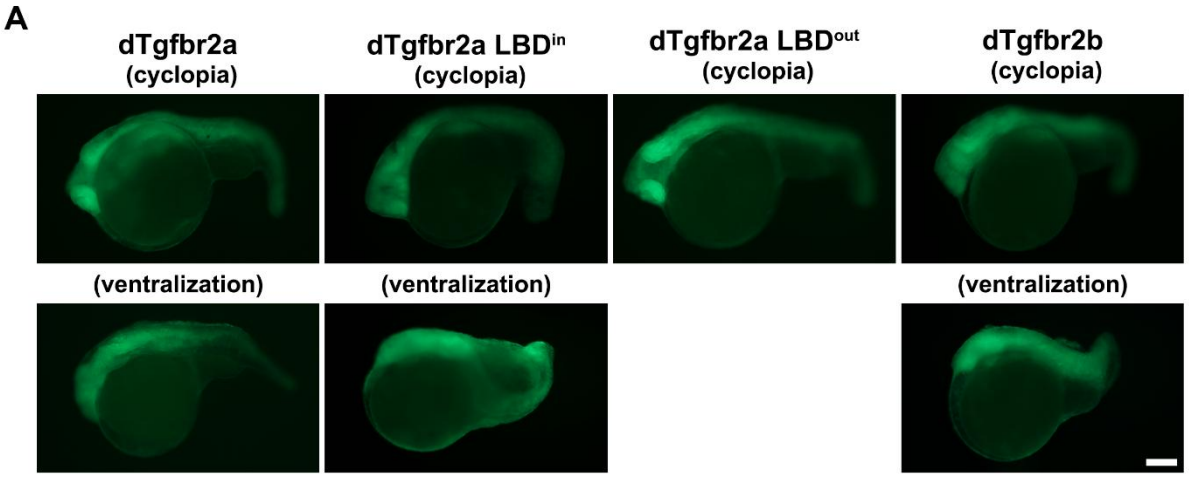
Supplemental Figure 16: Representative LSBA images of TGFβ1 bound to gTGFBR2 variants. Representative confocal microscopy images of LSBA for unstimulated and hTGFβ1-SiR-d12 stimulated COS-7 cells expressing human TGFBR2 and *Gallus gallus* variants (gTGFBR2 (2LBD), gTGFBR2-LBD^{in}, gTGFBR2-LBD^{out}) and human TGFBR2 or untransfected control. Scale bar ≙ 20 µm.



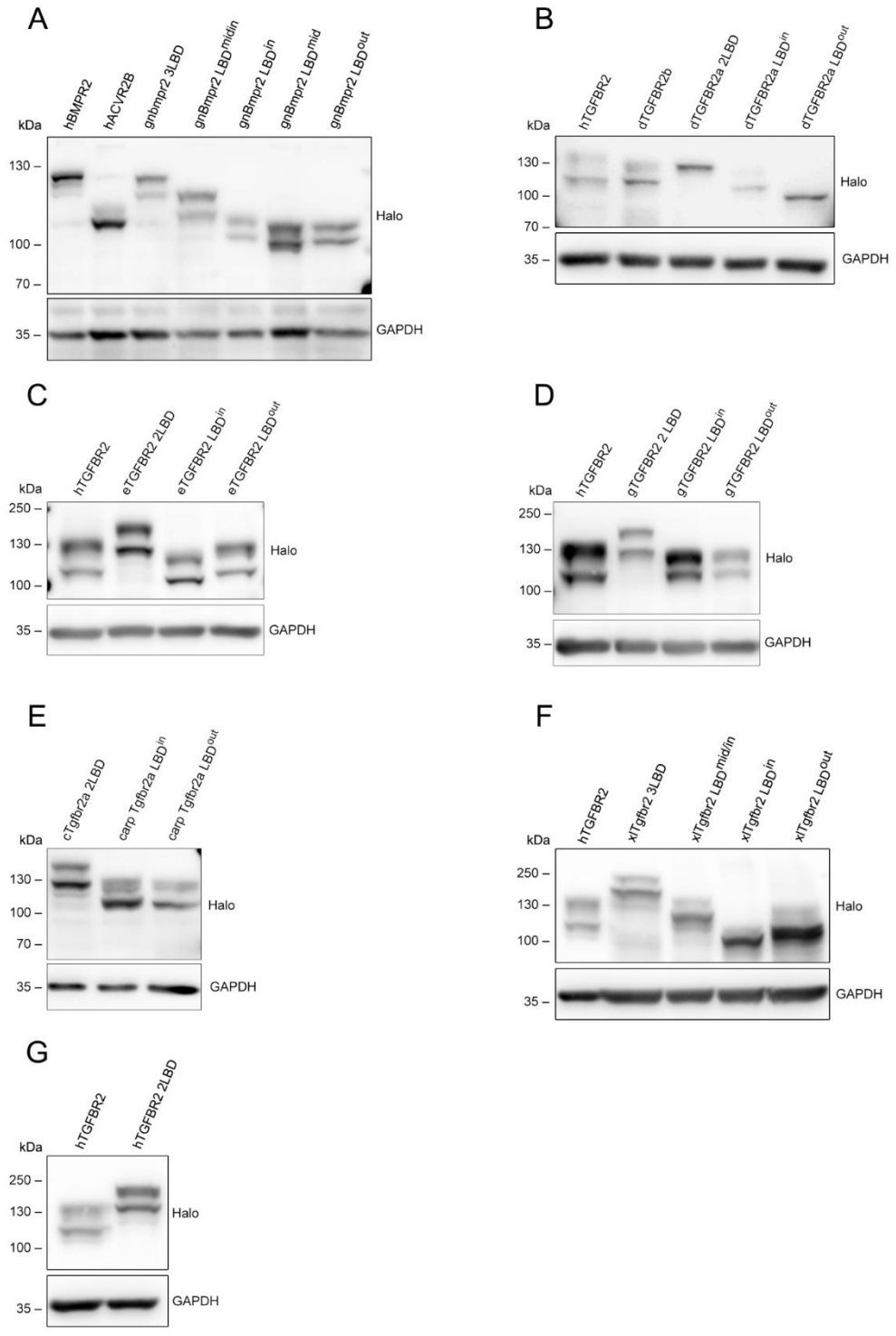
Supplemental Figure 17: Representative LSBA images of TGFβ1 bound to hTGFBR2 variants. Representative confocal microscopy images of LSBA for unstimulated COS-7 cells expressing hTGFBR2 or hTGFBR2-2LBD or untransfected control and hTGFβ1-SiR-d12 stimulated untransfected cells. Scale bar ≙ 20 µm.



Supplemental Figure 18: Alternative splice variants of TGFBR2 in horse and chicken. (**A, top**) Exon organization of horse eTGFBR2 and chicken gTGFBR2 with DNA/protein similarity percentages between corresponding exons. (**bottom**) Illustration of alternative splice variants, showing different combinations of exons leading to 2LBD, LBD_{out}, LBD_{in} receptor variants. (**B**) Summary of relative usage 2a/2b usage across 14 horse tissues. Tissues are indicated on the left (reproduction, blood vessels, lung, heart). Tissues used for Sashimi plots are highlighted. (**C,D**) Sashimi plot of the E1 to E5 region of horse (left) and chicken (right) TGFBR2. Exon coverage is shown on the left. Junctions are indicated by lines, junctions not supported by unique reads are shown as dotted lines. Next to each junction the black number shows all reads mapped to this junction, the red number reports unique reads only.



Supplemental Figure 19: Coinjected membrane bound PMT-mEGFP expression control in zebrafish embryos. (**A**) Expression of coinjected PMT-mEGFP allows selection of dTgfbr2 overexpression embryos. Scale bar = 100 µm.



Supplemental Figure 20: Western blot expression controls of Halo-tagged receptor constructs. Western blot analysis of HEK293T cell transiently expressing (**A**) *Gnathonemus petersii* receptor variants *(*gBmpr2a (3LBD), gBmpr2a-LBD^{midin}, gBmpr2a-LBD^{in}, gBmpr2a-LBD^{mid}, gBmpr2a-LBD^{out}) or hBMPR2 and hACVR2B as size reference (**B**) *Danio rerio* receptor variants *(*dTgfbr2b, dTgfbr2a (2LBD), dTgfbr2a-LBD^{in}, dTgfbr2a-LBD^{out}) or hTGFBR2 as size reference (**C**) *Equus caballus* variants *(*eTGFBR2 (2LBD), eTGFBR2-LBD^{in}, eTGFBR2-LBD^{out}) and hTGFBR2 as size reference (**D**) *Gallus gallus* variants *(*gTGFBR2 (2LBD), gTGFBR2-LBD^{in}, gTGFBR2-LBD^{out}) and hTGFBR2 as size reference (**E**) *Cyprinus carpio* variants *(*cTgfbr2a (2LBD), cTgfbr2a-LBD^{in}, cTgfbr2a-LBD^{out}) or hTGFBR2 as size reference (F)

Xenopus laevis variants *(*xlTgfbr2.L (3LBD), xlTgfbr2.L-LBD$^{mid\text{-}in}$, xlTgfbr2.L-LBD^{in}, xlTgfbr2.L-LBD^{out}) or hTGFBR2 as size reference (**G**) hTGFBR2 and hTGFBR2-2LBD. GAPDH was used as loading reference.

Table 1: N-terminally tagged HaloTag expression constructs

Plasmid Name	Description	Backbone	Resistance
eTGFBR2-Halo	full length eTGFBR2 with N-terminal HaloTag	pcDNA3.1	Ampicillin
eTGFBR2 LBD^{in}-Halo	eTGFBR2 LBD^{in} variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
eTGFBR2 LBD^{out}-Halo	eTGFBR2 LBD^{out} variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
dTgfbr2a-Halo	full length dTgfbr2a with N-terminal HaloTag	pcDNA3.1	Ampicillin
dTgfbr2a LBD^{in}-Halo	dTgfbr2a LBD^{in} variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
dTgfbr2a LBD^{out}-Halo	dTgfbr2a LBD^{out} variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
dTgfbr2b-Halo	dTgfbr2b with N-terminal HaloTag	pcDNA3.1	Ampicillin
cTgfbr2a-Halo	full length cTgfbr2a with N-terminal HaloTag	pcDNA3.1	Ampicillin
cTgfbr2a LBD^{in}-Halo	cTgfbr2a LBD^{in} variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
cTgfbr2a LBD^{out}-Halo	cTgfbr2a LBD^{out} variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
gnBmpr2a-Halo	full length gnBmpr2a with N-terminal HaloTag	pcDNA3.1	Ampicillin
gnBmpr2a LBD^{in}-Halo	gnBmpr2a LBD^{in} variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
gnBmpr2a LBD$^{mid\text{-}in}$-Halo	gnBmpr2a LBD$^{mid\text{-}in}$ variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
gnBmpr2a LBD^{out}-Halo	gnBmpr2a LBD^{out} variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
gnBmpr2a LBD^{mid}-Halo	gnBmpr2a LBD^{mid} variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
gTGFBR2-Halo	full length gTGFBR2 with N-terminal HaloTag	pcDNA3.1	Ampicillin
gTGFBR2 LBD^{in}-Halo	gTGFBR2 LBD^{in}-Halo variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
gTGFBR2 LBD^{out}-Halo	gTGFBR2 LBD^{in}-Halo variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
xlTgbfr2.L – blunt end	full length xlTgbfr2.L in blunt end vector	pCR™-Blunt-vector	Ampicillin
xlTgbfr2.L	full length xlTgbfr2.L with N-terminal HaloTag	pcDNA3.1	Ampicillin
xlTgbfr2.L LBD^{in}	xlTgbfr2.L LBD^{in} variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
xlTgbfr2.L LBD$^{mid\text{-}in}$	xlTgbfr2.L LBD$^{mid\text{-}in}$ variant with N-terminal HaloTag	pcDNA3.1	Ampicillin
pcs2+PMT-mEGFP	Membrane bound mEGFP: Gift from Dr. Thorsten Wohland (Balasubramanian et. al, 2022, RRID: Addgene_203777), subcloned into pCS2+ backbone for capped mRNA synthesis (Le et al., 2025)	pcs2+	Ampicillin
xlTgbfr2.L LBD^{out}	xlTgbfr2.L LBD^{out} variant with N-terminal	pcDNA3.1	Ampicillin

1261 **Table 2: Oligonucleotides used for molecular cloning of LBD-Halo constructs**

Name	Oligonucleotide-sequence (5'-3')	
	for	**rev**
eTGFBR2	CATAATTTTCAGTGAAAACCCCGACTTGC TTTTAGTTATCTTTCAAG	ATAACTAAAAGCAAGTCGGGGTTTTCACTGA AAATTATGTGGTCGTTGC
eTGFBR2 LBD^{in}	GCACTGCGGCTGCTGGATCCGCAGAAAT CGGTACTGGCTTTCCATTC	ACTAAAAGCAAGTCGGGGTTTTCACTGAAAA TTATGTGGTCG
eTGFBR2 LBD^{out}	AACCCCGACTTGCTTTTAGTTATCTTTCA AGTGACAGGGG	GGATCCAGCAGCCGCAGTGCTGACCAGCAG GATGGTCCATGG
dTgfbr2a	GGTGGAGAATTCCGCATCTGCAGCTCCT GTAAGCCTCATCCAGTCAACTG	GCTCGAGCGGCCGCCTACATCTCTTGTATCC CAGAATCC
dTgfbr2a LBD^{in}	GGTGGAGAATTCTTGTCGAAGACAAACT CCAGCAGC	GCTCGAGCGGCCGCCTACATCTCTTGTATCC CAGAATCC
dTgfbr2a LBD^{out}	GATTTGAAACAATATCCAGGGGTTCAG	CAATTTCTCCTCCTCAGTGAGTTTG
dTgfbr2b	GGTGGAGAATTCCGCGGAGTCGATGGA AATAATCC	GCTCGAGCGGCCGCTCATTTGTCGTCGCTCA CGCTCACG
cTgfbr2a	GGTGGAGAATTCAATCCACACAGTGTGT GTAAGTCATGTGACC	GCTCGAGCGGCCGCCTATTTCTCATCATTCC CAGAATCC
cTgfbr2a LBD^{in}	GATAACGACACCAAGTTAGTTATATTG	ATTGAATTCTCCACCTCCACCTCCGC
cTgfbr2a LBD^{out}	GTTTCGTCTCCTCCAGGAGACTATAAG	GAGTCTGAACCCTCCTTTCGTTTC
gnBmpr2a	GGTGGAGAATTCGCCCAAGCGGAAGAC CGCGTGTGC	GCTCGAGCGGCCGCCTAATTAACGGTAGGT GACACG
gnBmpr2a^{in}	GGTGGAGAATTCGCACCTAATGAAGAAA GGGAATGC	GCTCGAGCGGCCGCCTAATTAACGGTAGGT GACACG
gnBmpr2a $^{mid-in}$	GGTGGAGAATTCGCCGAAAGAGACGAAT TGC	GCTCGAGCGGCCGCCTAATTAACGGTAGGT GACACG
gnBmpr2a^{out} (Gibson)	**Backbone:** CCGCTGTTCGGTCCAACATCCACTCAAC CCCTCTATCCAC **LBD^{out}:** GAGGTGGAGGTGGAGAATTCGCCCAAG CGGAAGACCGCGTG	**Backbone:** GAATTCTCCACCTCCACCTCCGCCGGAAATC TCGAGCGTCG **LBD^{out}:** GATGTTGGACCGAACAGCGGAAAATCTTCGG TGAAATTAAGATTAC
gnBmpr2a^{mid} (Gibson)	**Backbone:** CCGCTGTTCGGTCCAACATCCACTCAAC CCCTCTATCCAC **LBD^{mid}:** GAGGTGGAGGTGGAGAATTCGCCGAAA GAGACGAATTGCAATGC	**Backbone:** GAATTCTCCACCTCCACCTCCGCCGGAAATC TCGAGCGTCG **LBD^{mid}:** GATGTTGGACCGAACAGCGGAGACAGATCG GTGAAGTTTAC
gTGFBR2 & gTGFBR2 LBD^{out}	GGTGGAGAATTCAGAAGCAAGGAAAATG GACTGC	GCTCGAGCGGCCGCCTACTTGGCGGTGGTC ACGGAGC
gTGFBR2, gTGFBR2-LBD^{out} (Gibson)	**Backbone:** GGCATCCAGATGGTGTGCGAGACCCTTA TCGAGTGCTGGG **Exon 6/7 Del:** TGCCCATTGAACTGGACATTGTTGTTGG CAAAGGAAGGTTTGCTGAAGTG	**Backbone:** GGCATCCAGATGGTGTGCGAGACCCTTATCG AGTGCTGGG **Exon 6/7 Del:** TCGCACACCATCTGGATGCCCTGATGGTTAA GCCAGGAGCTGG
gTGFBR2 LBD^{in}	CCCCATAAGCCAGAGGAGAAAGATGAAA TTTCC	CCTAAATCTGTTTGCAGAAAAAATAAG
xlTgbfr2.L – blunt end	GGTACCTCCACCAAGGGCGAATTCTGC	ACATTTCCCTTATACCTGGGTTCTAGTC
xlTgbfr2.L LBD^{in} (Gibson)	**Backbone:** GATGACCCATCAAAAGACATACTTGTGA AAGCTCTTCTCATC **LBD^{in}:** GAGGTGGAGGTGGAGAATTCAGTATAAA GGATTGTGC	**Backbone:** GAATTCTCCACCTCCACCTCCGCCGGAAATC TCGAGCGTCGACAGCC **LBD^{in}:** ATGTCTTTTGATGGGTCATCTTTAGAAAAAAA CAAGTTTTCATTACACAGC
xlTgbfr2.L LBD^{mid-in} (Gibson)	**Backbone:** GATGACCCATCAAAAGACATACTTGTGA AAGCTCTTCTCATC	**Backbone:** GAATTCTCCACCTCCACCTCCGCCGGAAATC TCGAGCGTCGACAGCC

	LBD^{mid-in}: CCGGCGGAGGTGGAGGTGGAGAATTCC CTGTTCATGTACCTGTTGTATGC	**LBD^{mid-in}:** ATGTCTTTTGATGGGTCATCTTTAGAAAAAAA CAAGTTTTCATTACACAGC
xlTgbfr2.L LBD^{out} (Gibson)	**Backbone:** GATGACCCATCAAAAGACATACTTGTGA AAGCTCTTCTCATC **LBD^{out}:** GAGGTGGAGGTGGAGAATTCTACCTGG GTTCTAGTCGTTTATGC	**Backbone:** GAATTCTCCACCTCCACCTCCGCCGGAAATC TCGAGCGTCGACAGCC **LBD^{out}:** ATGCTTTTGATGGGTCATCATCTGTGAAGACT ATTTGTTCATTGC

1262

1263