Supplemental Materials and Methods

Identification and prediction of full-length DICP transcripts

BLAST analyses were used to identify ESTs encoding zebrafish DICPs. RACE strategies were used to clone partial and full-length DICP cDNAs. Tissues from adult zebrafish were dissected directly into Trizol and total RNA purified as recommended by the manufacturer (Invitrogen). Five µg of RNA (1.7 µg from kidney, 1.7 µg from spleen and 1.7 µg from intestine) were reverse transcribed using an oligo dT primer and Superscript™ III Reverse Transcriptase supplied with the GeneRacer™ kit and RACE strategies were conducted as recommended by the manufacturer (Invitrogen). PCR products were cloned into pGEM-T Easy (Promega) and sequenced. cDNA sequences have been submitted to GenBank with accession numbers JN416849-JN416863 and predicted proteins listed in Supplemental Fig. S3.

dicp3a, dicp3b, dicp3i, dicp3k, dicp3p and dicp3s

A full-length *dicp3a* EST (IMAGE clone 7037237, GenBank CK019181) was identified by BLAST, acquired and sequenced (*dicp3a*^{2326}). A second *dicp3a* transcript, *dicp3a*^{2614}, was recovered by RT-PCR with primers designed to include the translational start and stop codons (For2 CAGAATGGCTGATAGGAGTCCTCTG and DICP3abikps-Rev1 CATTAAGCTCTTCTGCTGACG translational start and stop codons are underlined). A *dicp3i* cDNA was obtained by 3′ RACE (*dicp3i*^{2114}) with nested forward primers designed to be complementary to the exon encoding the predicted leader sequence (For1 GTCAGACACACAGAATGGCTGATAG and For2 CAGAATGGCTGATAGGAGTCCTCTG). The exons encoding the transmembrane domains and cytosplasmic regions of *dicp3a* and *dicp3i* and the corresponding predicted exons of *dicp3b*, *dicp3k*, *dicp3p* and *dicp3s* share complete identity or differ by a single nucleotide but display variation in the D1 domain. *dicp3b*, *dicp3k*, *dicp3p* and *dicp3s* are predicted to encode a D2 domain. Dicp3a lacks a D2 domain, but includes four novel exons that encode a region of low sequence complexity. Dicp3i lacks a D2 domain but includes a short exon that is similar to the 5′ end of the D2 domain of chromosome 3 DICPS (e.g. polyserine sequence).

dicp3g and dicp3h

BLAST analyses identified EST IMAGE clone 7994924 (GenBank DT066294) as encoding Dicp3g. This EST was acquired and sequenced (*dicp3g*^{2323}). *dicp3h* was deduced by its similarity to *dicp3g* in genomic scaffold 262. The coding sequence of *dicp3g* and *dicp3h* from genomic scaffold 262 are 99% identical, and are predicted to encode receptors that differ by two residues.

dicp3cd and dicpef

Since the majority of DICP genes possess a single D1 and D2 domain, *dicp3c* and *dicp3d* were thought originally to represent two genes. Likewise, *dicp3e* and *dicp3f* also were thought to represent two genes. However, it was noted that *dicp3c* and *dicp3d* were in the same orientation with *dicp3c* lacking a leader exon and being downstream of *dicp3d*. In addition, *dicp3e* and *dicp3f* are in the same orientation with *dicp3e* lacking a leader exon and being downstream of *dicp3f*. Therefore, it was reasoned that *dicp3c* and *dicp3d* could encode a single four Ig domain receptor and that *dicp3e* and *dicp3f* could encode a second four Ig domain receptor. RT-PCR successfully identified a *dicp3ef* transcript with a D1-D2-D1

configuration, *dicp3ef*^{1512} demonstrating that D1-D2-D1 and possibly D1-D2-D1-D2 transcripts can be generated from *dicp3cd* and *dicp3ef* (using primers dicp3ef-for1 AGAGAAACTCAGAGAGACTGAATCTGGG and dicp3ef-rev1 AACCAATCAGCTGAACTGCTGCTGTCTG). Exons 3' of the D1 and D2 domains were predicted by comparison to an EST (IMAGE:8109230, GenBank DT866261.1) and a full-length D1-D2 transcript, *dicp3cd*^{2748}, recovered by 5' RACE (primers dicp3c-5'RACE-rev1 CAGGGCAGCTCAGTTCTCTATGAG and dicp3c-5'RACE-rev2 GTGTAATCGCAGTTCATGCTGTGC). Full length D1-D2-D1-D2 transcripts are predicted, but not recovered.

dicp3q

Two variants of *dicp3q* have been identified. The *dicp3q*^{1910} partial cDNA was recovered by 3' RACE using primers in the D1 domain (For1 CAGATCTGGAGAAACTGTTCATCTG and For2 GACAGCAGCAGTTCAGCTGATTGG) and encodes a secreted protein with a single Ig (D1) domain. The *dicp3q*^{2145} cDNA was recovered by 3' RACE using primers complementing the leader domain (For1 GTCAGACACACAGAATGGCTGATAG and For2 CAGAATGGCTGATAGGAGTCCTCTG) and encodes a membrane bound receptor with a single Ig (D1) domain.

dicp3l, dicp3n, dicp3o, dicp3r, dicp3t

Transcripts corresponding to *dicp3l*, *dicp3n*, *dicp3o*, *dicp3r* and *dicp3t* have not been identified.

dicp3jP, dicp3mP, dicp3u

The exons predicted to encode the D1 domains of *dicp3j* and *dicp3m* possess frame shifts that result in stop codons within the D1 domains. EST searches as well as RACE and PCR strategies have not been successful in identifying *dicp3j* or *dicp3m* transcripts. Based on these observations, we propose that they represent psuedogenes and designate them *dicp3jP* and *dicp3mP*. The D1 domain of *dicp3u* was identified at the end of scaffold 262 and it is unclear if this exon represents a pseudogene or a functional gene.

dicp14a

Nested forward primers were designed for 3' RACE based on the predicted D1 of *dicp14a* (For1 GAGTCTGGAGTAAATCTCCAAACAG and For2 CCTCTGTAGATCACTGTAGCATCAG) and employed to identify the 3' UTR of *dicp14a*. Subsequently, nested reverse primers were designed for 5' RACE based on the 3' UTR of *dicp14a* (Rev1 CAGTCTTAAGGTACAGGTTGAGAG and Rev2 CAGAGGTAMATTCAGCAAATGTGTC) and employed to clone a cDNA encoding the complete ORF of *dicp14a*^{1938}.

dicp14b

A putative leader as well as D1 and D2 domains were identified adjacent to the genomic sequence of *dicp14a* that are predicted to encode a putative *dicp14b*. Nested forward primers were designed against the leader sequence of *dicp14b* (For1 GTGACAGAGTGACAAATAGACATGCTG and For2 GACATGCTGGGACTGATCATTTTCTGC, translational start codon is underlined) and applied to 3' RACE. A *dicp14b* cDNA possessing a complete ORF that did not include the D1 exon (*dicp14b* transcript variant 2, *dicp14b-tv2*^{2011}) was identified. Reverse-transcriptase PCR (RT-PCR) was used with cDNA (from zebrafish kidney and intestine) and primers designed to amplify the entire ORF of *dicp14b*, For2 and Rev3

(TTACTTCTGTAGGGGTCTTGACGTGGC), which generated the expected band for *dicp14b-tv2* along with longer and shorter cDNAs that were cloned and sequenced. The longer cDNA encodes a *dicp14b* transcript variant (*dicp14b-tv1*^{2271}) that includes both D1 and D2 exons and a TM domain. The shorter cDNA encodes a *dicp14b* variant (*dicp14b-tv3*^{2247}) that lacks the D1 and TM exons.

dicp16a

A zebrafish kidney EST (IMAGE:6794317; GenBank CA473423) was identified as being highly similar to the predicted *dicp16a* D1 domain, and was sequenced. This EST encodes the complete ORF of *dicp16a-tv1*^{2065}. An alternate form of dicp16a, *dicp16a-tv2*^{1912}, was identified by 3' RACE employing primers complementary to the D1 domain (dicp16a-for1 (CAGATCCGGTGAAAATGTCAGTCTG and dicp16a-for2 CTGTAATAATGCTCTTTCTGGCTGC).

dicp16c

A zebrafish brain EST (Genome Institute of Singapore clone FDR103-P00066-DEPE-F_J19; GenBank EH455442) was identified as being highly similar to the predicted *dicp16c* D1 (generously provided by W.C. Lin, Genome Institute of Singapore) and sequenced. This EST was found to encode the complete ORF of *dicp16c*^{2081}.

dicp16b, dicp16f

Transcripts corresponding to dicp16b and dicp16f have not been identified.

dicp16dP and dicp16eP

The chromosome 16 reference sequence (scaffold 1952) possesses two additional sets of D1 and D2 domains named *dicp16d* and *dicp16e;* however, frameshifts are present in the D2 domain of *dicp16d* and in the D1 and D2 domains of *dicp16e*. All frameshifts would result in premature stop codons within transcripts possessing these exons. Interestingly, *dicp16d*, which is encoded in a different genomic clone (BAC CH73-34H11), does not have a frame shift in D2 and may encode a functional protein. Although dicp16e in BAC CH73-34H11 does not have a frame shift in D1, a frame shift is present in D2. Although alternative mRNA splicing might produce transcripts that lack these exons, multiple RACE strategies were not successful in identifying *dicp16d* or *dicp16e* transcripts. In addition, no additional sequences flanking *dicp16d* and *dicp16e* that share identity with other DICP exons were identified. Based on these observations we propose that *dicp16d* and *dicp16e* are pseudogenes and refer to them as *dicp16dP* and *dicp16eP*.