EpCAM Multiple Sequence Alignment Xenopus & vertebrates

```
CLUSTAL 2.0.12 multiple sequence alignment

xEmCAMa     --------------------MCLVLVAQVQSQGCKCRTHYMGKCDNSGASSDCQCTLTIG  40
xEmCAMb     MKFVSVLRLG-------AALMCLVLVTRAQNPGCKCKTHYLGKCDNSGASSDCQCALSIG  53
xtEpCAM     MHLSTVLRLG-------AALLCFALVAQAQSPGCTCSTLYMGKCDNSGAG-GCQCTLAIG  52
hEpCAM1     MAPPQVLAF-------GLLLAAATATFAAAQEECVCENYKLAVNCFVNNNRQCQCT-SVG  52
mEpCAM1     MAGPQALAF-------GLLLAVVTATLAAAQRDCVCDNYKLATSCSLNEYGECQCT-SYG  52
cEpCAM      MELLRGAAL-------LLLLCAA----ACAQDSCTCTKNKRVTNCRLIDN-VCHCN-SIG  47
hEpCAM2     MARGPGLAPPPLRLPLLLLVLAAVTGHTAAQDNCTCPTNKMTVCSPDGPGGRCQCR-ALG  59
mEpCAM2     MARGLDLAP------LLLLLLAMATRFCTAQSNCTCPTNKMTVCDTNGPGGVCQCR-AMG  53
zEpCAM1     MK-------------VLVALFVVALVD-VTSQCTCKTMKWANCDD-----SCSCSLTLT  40
zEpCAM2     MK-------------VLVALFVVALVDVVTSQCACKTMKWANCDD-----SCSCSLTLT  41
                           *  *  .               * *  :

xEmCAMa     P-DSQPVNCSKLIPKCWLMKRESL---GTKAGR-RVKP-AQALIDNDGLYNPECDTNGVF  94
xEmCAMb     P-ASQAVDCTKLIPKCWLMKRESL---GTKAGR-RVKP-VQALIDNDGLYDPECETNGVF 107
xtEpCAM     T-ATQSINCSALIPKCWLMKRESL---GTKAGR-RVKP-VQALVDNDGLYDPECDVNGVF 106
hEpCAM1     --AQNTVICSKLAAKCLVMKAEMN---GSKLGR-RAKP-EGALQNNDGLYDPDCDESGLF 105
mEpCAM1     --TQNTVICSKLASKCLAMKAEMT---HSKSGR-RIKP-EGAIQNNDGLYDPDCDEQGLF 105
cEpCAM      --SSVPVNCEILTSKCLLMKAEMA---NTKSGR-REKP-KDALQDTDGLYDPECENNGLF 100
hEpCAM2     --SGMAVDCSTLTSKCLLLKARMS---APKNARTLVRPSEHALVDNDGLYDPDCDPEGRF 114
mEpCAM2     --SQVLVDCSTLTSKCLLLKARMS---ARKSGRSLVMPSEHAILDNDGLYDPECDDKGRF 108
zEpCAM1     ESSKQTLDCSKLVPKCFLMKAEMYRARHNLGTRKTGKPDENAFVDNDGIYDPECQSDGKF 100
zEpCAM2     ESSTQTLNCSKLVPKCFLMQAEMYRACNHQDTRSGGKPVETAFVDNDGIYDPVCESDGKF 101
                :  *   * .**  ::  .        *      *    *: :.**:*:* *: .* *

xEpCAMa     KARQCNNTDTCWCVNTAGVRRTDKGDKNWKCPELVRTNWVYVEMKRNNTDS-VNDDDLKK 153
xEpCAMb     KARQCNNTDTCWCVNTAGVRRTDKGDKNWKCPELVRTNWVIVEMKRNNSDS-VNDDVLIQ 166
xtEpCAM     KARQCNNTDTCWCVNSAGVRRTDKGDKNWKCPELVKTNWVIVEMKRNGTES-VSDADLIQ 165
hEpCAM1     KAKQCNGTSTCWCVNTAGVRRTDK-DTEITCSERVRTYWIIIELKHKAREKPYDSKSLRT 164
mEpCAM1     KAKQCNGTATCWCVNTAGVRRTDK-DTEITCSERVRTYWIIIELKHKERESPYDHQSLQT 164
cEpCAM      KAKQCNGT-TCWCVNTAGVRRTDKHDTDLKCNQLVRTTWIIIEMRHAERKTPLNAESLTR 159
hEpCAM2     KARQCNQTSVCWCVNSVGVRRTDKGDLSLRCDELVRTHHILIDLRHRPTAGAFNHSDLDA 174
mEpCAM2     KARQCNQTSVCWCVNSVGVRRTDKGDQSLRCDEVVRTHHILIELRHRPTDRAFNHSDLDS 168
zEpCAM1     KAVQCNNTEVCWCVNSAGVRRSDKKDKNIKC-EPAETYWVRAEMTHKSVDVPIDVANLRM 159
zEpCAM2     KAVQCNNTEVCWCVNSAGVRRSDKKDKNIKC-EPAETYWVRVEMKHKSVDVPIDATKLRT 160
            ** *** *  .*****:.****:** *  .  * :  ..*   :  :: :       .  *

xEmCAMa     ALKTTIVNRYGLPEKCVSVELEGPS--LIYVDLKQNGSQKLPGEVDITDVAYYMEKDIKG 211
xEmCAMb     ALKTTILNRYGLPEKYVSVELEGSS--FIYIDLKQNGTQKLPGEVDITDVGYYMEKDIKG 224
xtEpCAM     ALKTTITNRYGLPDKYISVELETP---LIYIDLKQNTSQKLPGEVDITDVAYYMEKDVKG 222
hEpCAM1     ALQKEITTRYQLDPKFITSILYENN--VITIDLVQNSSQKTQNDVDIADVAYYFEKDVKG 222
mEpCAM1     ALQEAFTSRYKLNQKFIKNIMYENN--VITIDLMQNSSQKTQDDVDIADVAYYFEKDVKG 222
cEpCAM      YLKDTITSRYMLDGRYISGVVYENP--TITIDLKQNSSDKTPGDVDITDVAYYFEKDVKD 217
hEpCAM2     ELRRLFRERYRLHPKFVAAVHYEQP--TIQIELRQNTSQKAAGDVDIDDAAYYFERDIKG 232
mEpCAM2     ELRRLFQERYKLHPSFLSAVHYEEP--TIQIELRQNASQKGLRDVDIADAAYYFERDIKG 226
zEpCAM1     GIENALQQRYFLDKNFVSEVQYDKDARLIVVDVKKDKNDR---TTDLSLMTYYLEKDIKV 216
zEpCAM2     GIENVLQQRYGLDKKLVSEVQYDKDGRLIVVDVKKDKDDR---TTDLSLMTYYMEKDIKV 217
            :.  :  ** *    :        * ::: :: ::   .*:   **:*:*:*

xEmCAMa     DSLFHPDEKFEILVNGNNFAVKEP--IIYYIDEKPHEISMKHLTPGVIAVIVVVVLAIVA 269
xEmCAMb     DPLFHPDEKFEILVNGKNFGVKEP--VIYYVDEKPHEITMKHLTPGVIAVIVVVVLAVVA 282
xtEpCAM     DSLFPANNQFQILANGNKISVKEP--MIYYIDEKPHEISMRHLTPGVIAVIVVVVLAVVA 280
hEpCAM1     ESLFHS-KKMDLTVNGEQLDLDPGQTLIYYVDEKAPEFSMQGLKAGVIAVIVVVVIAVVA 281
mEpCAM1     ESLFHSSKSMDLRVNGEPLDLDPGQTLIYYVDEKAPEFSMQGLTAGIIAVIVVVSLAVIA 282
cEpCAM      DSIFLN-NKLNMNIDNEELKFDN--MMVYYVDEVPPEFSMKSLTAGVIAVIVIVVLAIVA 274
hEpCAM2     ESLFQGRGGLDLRVRGEPLQVER--TLIYYLDEIPPKFSMKRLTAGLIAVIVVVVVALVA 290
mEpCAM2     ESLFMGRRGLDVQVRGEPLHVER--TLIYYLDEKPPQFSMKRLTAGVIAVIAVVSVAVVA 284
zEpCAM1     KPLFSDEKPFVLSVQGKNVTMEN--VLIYYVDDKAPTFTMQKLTGGIIAVIVVVSLIVIG 274
zEpCAM2     LPLFWNGQPFEVDVPGTKVSMEN--VLIYYVDDRAPTFTMQKLTGGIIAVIVVVSLIVIG 275
            .:*    : :   .  .  .    ::**:*: .  ::*: *. *:*****.:* : ::.

xEmCAMa     LIAVLIFTRRKR-GKYQKAEMKELNEMQQEAST 301
xEmCAMb     LIAVLIFTRRKK-GKYQKAEMKELNEMQKEVST 314
xtEpCAM     LIAVLIFTRRKK-ARYQKAEMKELNEMQKEVST 312
hEpCAM1     GIVVLVISRKKRMAKYEKAEIKEMGEMHRELNA 314
mEpCAM1     GIVVLVISTRKKSAKYEKAEIKEMGEIHRELNA 315
cEpCAM      GIIGLVLSRRRK-GKYVKAEMKEMNEMHRGLNA 306
hEpCAM2     GMAVLVITNRRKSGKYKKVEIKELGELRKEPSL 323
mEpCAM2     GVVVLVVTKRRKSGKYKKVELKELGEMRSEPSL 317
zEpCAM1     GFLVLFFLARRQKAHYSKAQAREMETIS----- 302
zEpCAM2     GFLVLFFLARRQKAQYSKAQAREMETIS----- 303
            .  *..  ::: .:* *.: :*:  :
```

Accession Numbers
- xEmCAMa: *Xenopus laevis* NP_001086975
- xEmCAMb: *Xenopus laevis* AAN86618
- xtEpCAM: *Xenopus tropicalis* AAH84898
- hEpCAM1: *Homo sapiens* AAH14785
- hEpCAM2: *Homo sapiens* CAG47056
- mEpCAM1: *Mus musculus* NP_032558
- mEpCAM2: *Mus musculus* NP_064431
- cEpCAM: *Gallus gallus* NP_001012582
- zEpCAM1: *Danio rerio* NP_998340
- zEpCAM2: *Danio rerio* NP_001017593

Figure S1. **Amino acid sequences of *Xenopus laevis* EpCAMa and b pseudoalleles and alignement with EpCAMs of representative vertebrate species.**
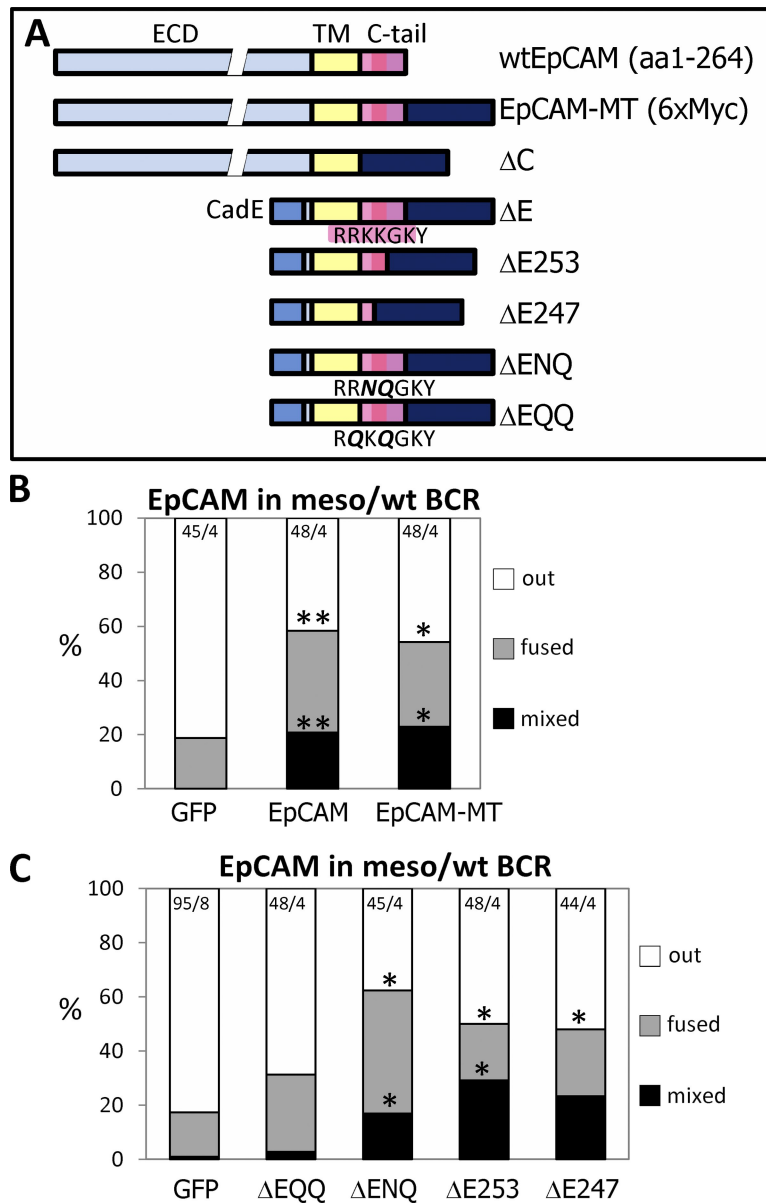
Figure S2. **Structure–function analysis of EpCAM.** (A) Diagram of the EpCAM constructs used in this study. The transmembrane (TM) and cytoplasmic (Ctail) domains and the 6xmyc tag, but not the extracellular domain (ECD), are presented on scale. (B) C-terminally myc-tagged EpCAM has the same tissue-mixing activity as wild-type EpCAM. (C) The mixing activity resides in specific residues of the cytoplasmic tail proximal to the transmembrane domain.
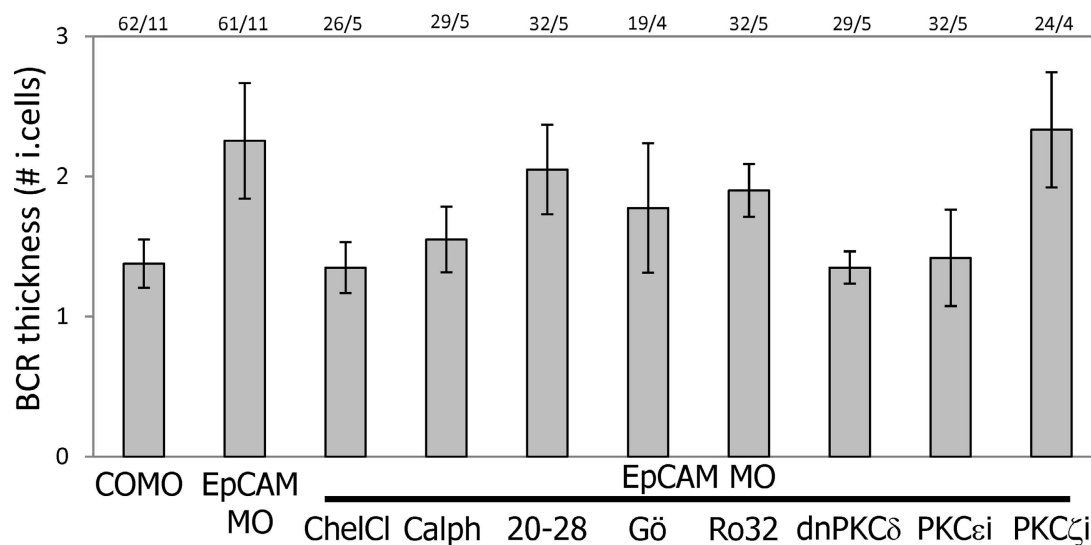
Figure S3. **Epiboly rescue experiments using the following PKC inhibitors.** ChelCl, chelerythrine chloride; Calph, calphostin C; 20–28, PKC-20-28; Gö, Gö6976; Ro32, Ro32-0432; dnPKC-δ, dominant-negative PKC-δ; PKC-εi, PKC-ε inhibitor; PKC-ζi, inhibitor of atypical PKCs.

membrane GFP

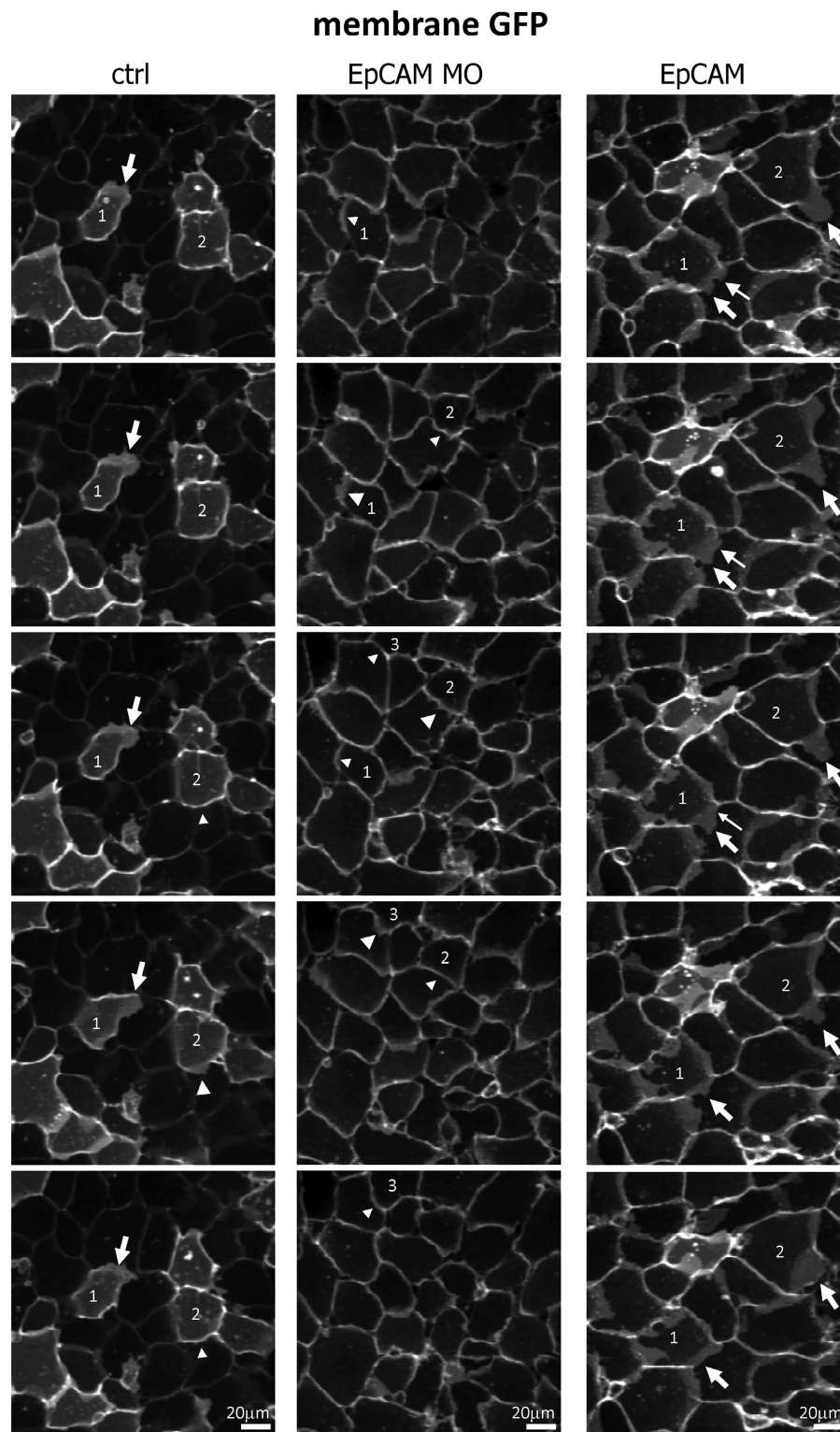ctrl   EpCAM MO   EpCAM

Figure S4.   **Selected frames from live time-lapse movies of BCRs expressing membrane GFP.** Every second frame is shown (12-min time interval). Numbers: examples of cells with long-lived protrusions (arrows) or short-lived protrusions (small arrowheads indicate the beginning and the end, large arrowheads the peak of protrusion extension). Note that large extensions often appear to be constituted of two or more parts that behaved independently (thin and large arrows in the last column [EpCAM mRNA]). In those cases, each part was scored as separate protrusions.