PNAS

Supporting Information for
Molecular mechanisms controlling the biogenesis of the TGF-β signal Vg1

P. C. Dave P. Dingal[a,b,1], Adam N. Carte[a,c,d,2], Tessa G. Montague[a,2], Medel B. Lim Suan, Jr.[b], and Alexander F. Schier[a,e,f,1]

[a]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, 02138
[b]Department of Bioengineering, The University of Texas at Dallas, Richardson, TX 75080
[c]Systems, Synthetic, and Quantitative Biology Program, Harvard University, Cambridge, MA 02138
[d]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115
[e]Biozentrum, University of Basel, 4056 Basel, Switzerland
[f]Allen Discovery Center for Cell Lineage Tracing, University of Washington, Seattle, WA 98109.

[1]To whom correspondence may be addressed. Email: davedingal@utdallas.edu or alex.schier@unibas.ch.
[2]A.N.C. and T.G.M. contributed equally to this work.

This PDF file includes:

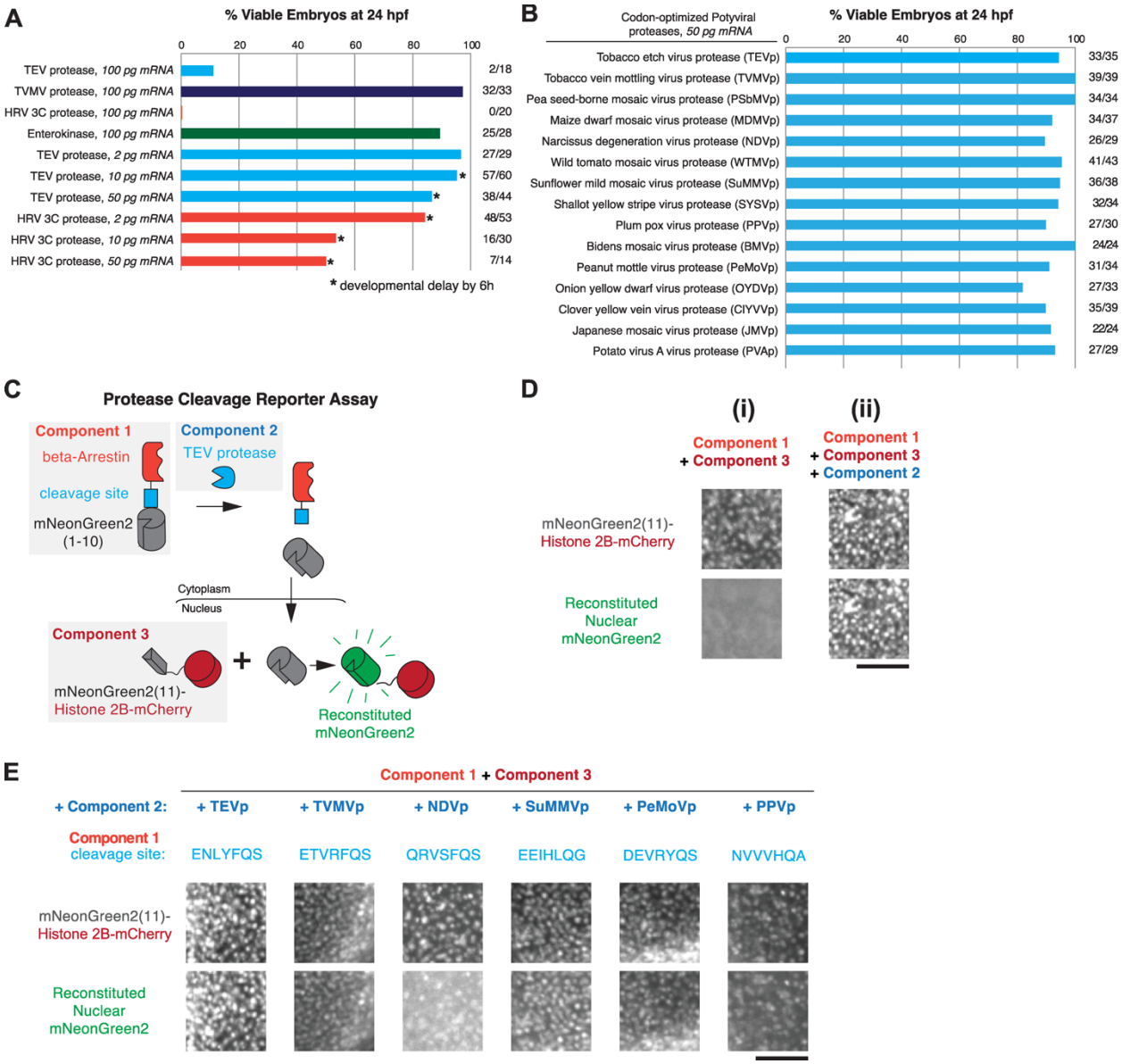

Fig. S1. Developing the SynPro system. (**A**) Viability at 24 hpf of wild-type embryos injected with mRNA of commercially available proteases: tobacco etch virus (TEV) protease (Addgene plasmid #8835), tobacco vein mottling virus (TVMV) protease (Addgene plasmid #8832), enterokinase (Addgene plasmid #49048), and human rhinovirus 3C (HRV 3C) protease (Addgene plasmid #78571). Developmental delays and loss of viability were observed in embryos injected with TEV protease and HRV 3C protease. (**B**) Viability at 24 hpf of wild-type embryos injected with 50 pg of zebrafish codon-optimized mRNA sequences of 15 potyviral proteases. In all conditions, more than 80% of embryos were viable at 24 hpf. (**C**) Schematic of fluorescence reporter assay for protease cleavage, which is composed of three components: Component 1 contains the cleavage site flanked by beta-Arrestin and mNeonGreen2(1-10); Component 2 is the codon-optimized Potyviral protease (e.g. TEV protease recognizes the sequence ENLYFQS); and Component 3 is nuclear histone 2b tagged with mCherry and fused with mNeonGreen2(11). See Methods for construct details. (**D**) Testing the cleavage efficiency of TEV protease. (**i**) Co-expression of Components 1 and 3 does not exhibit mNeonGreen2 fluorescence, whereas (**ii**) co-expression of all 3 components leads to proteolytic cleavage, nuclear translocation, and binding of mNeonGreen2(1-10) to mNeonGreen2(11) – leading to reconstitution of mNeonGreen2 fluorescence in the nucleus. (**E**) Using the reporter assay in zebrafish embryos, six out of fifteen Potyviral proteases exhibited proteolytic activity, based on reconstitution of nuclear mNeonGreen2 fluorescence (bottom row images). Scale bars, 100 µm.

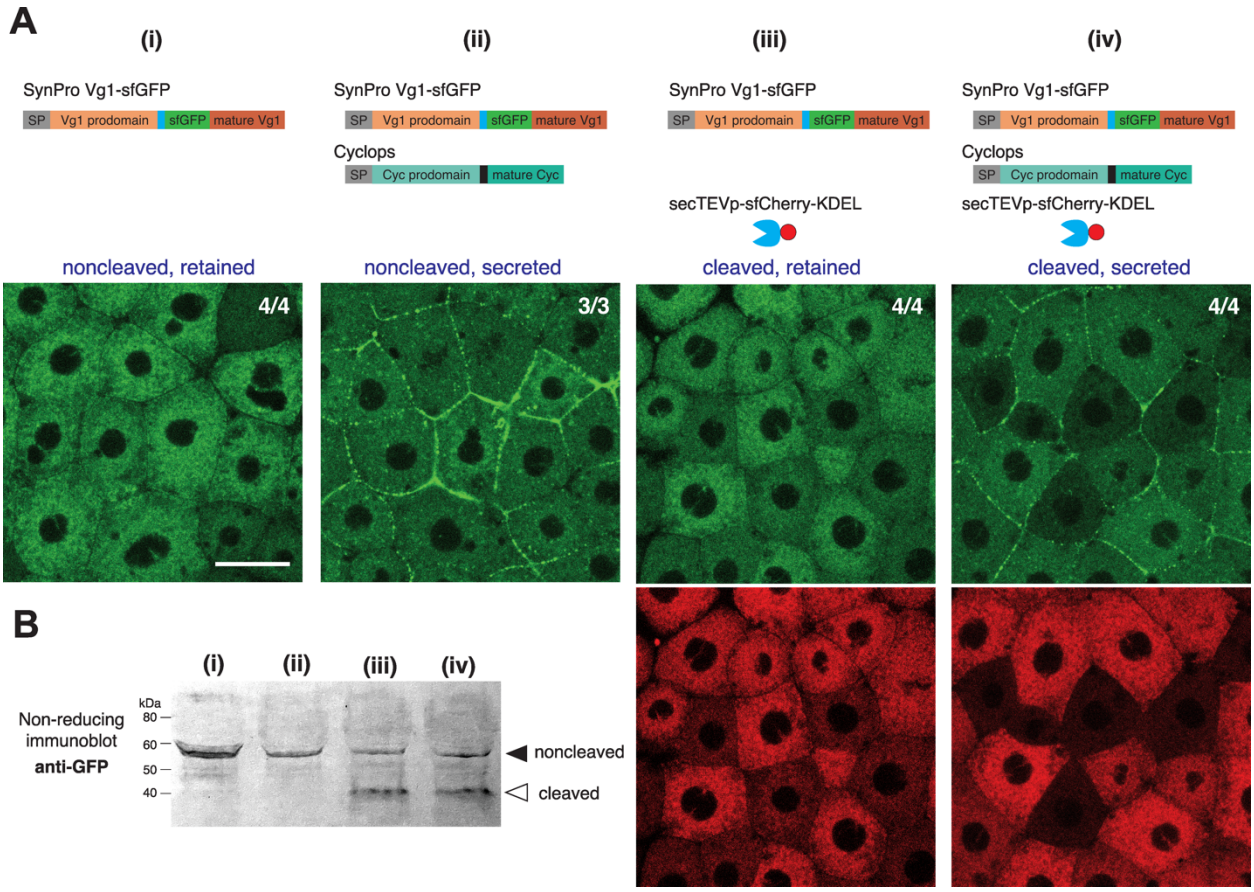

Fig. S2. Vg1 prodomain processing does not promote secretion. (**A**) Fluorescence images of fixed M*vg1* embryos injected with 25 pg mRNA of sfGFP-tagged *SynPro vg1*, *secTEVp-sfCherry-KDEL*, and *cyc* at the indicated combinations. *sfGFP* was inserted upstream of the *vg1* mature domain. Scale bar, 20 µm. (**B**) Anti-GFP non-reducing immunoblot of M*vg1* embryos injected with the indicated constructs as in (**A**). Black arrowhead indicates the position of uncleaved SynPro Vg1, open arrowhead indicates secTEVp-cleaved Vg1. 15 embryos at 3 hpf were loaded per well.

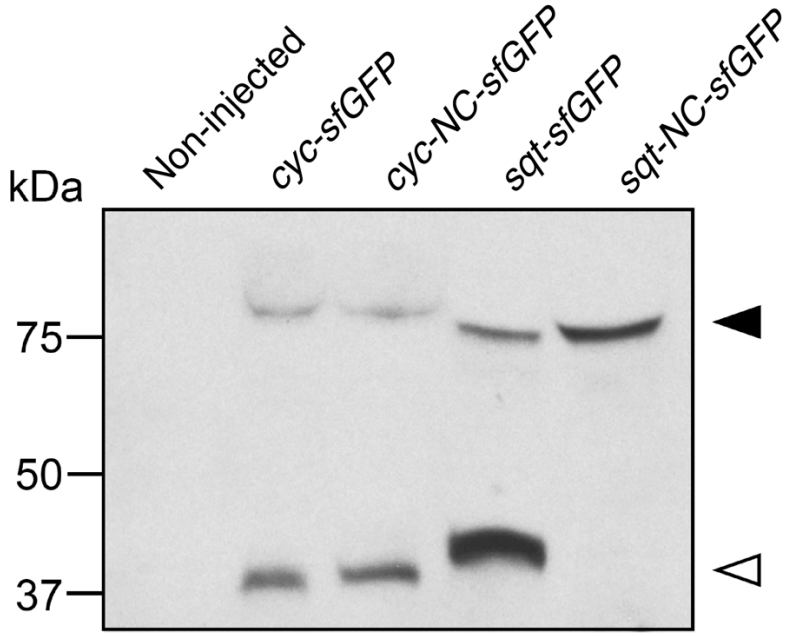

Fig. S3. Non-cleavable Squint. Anti-GFP reducing immunoblot of M*vg1* embryos injected with 50 pg mRNA of *cyc-sfGFP*, *cyc-NC-sfGFP* (RRGRR → SQNTS), *sqt-sfGFP* or *sqt-NC-sfGFP* mRNA. Black arrowhead indicates the position of full-length protein, open arrowhead indicates processed mature domain.

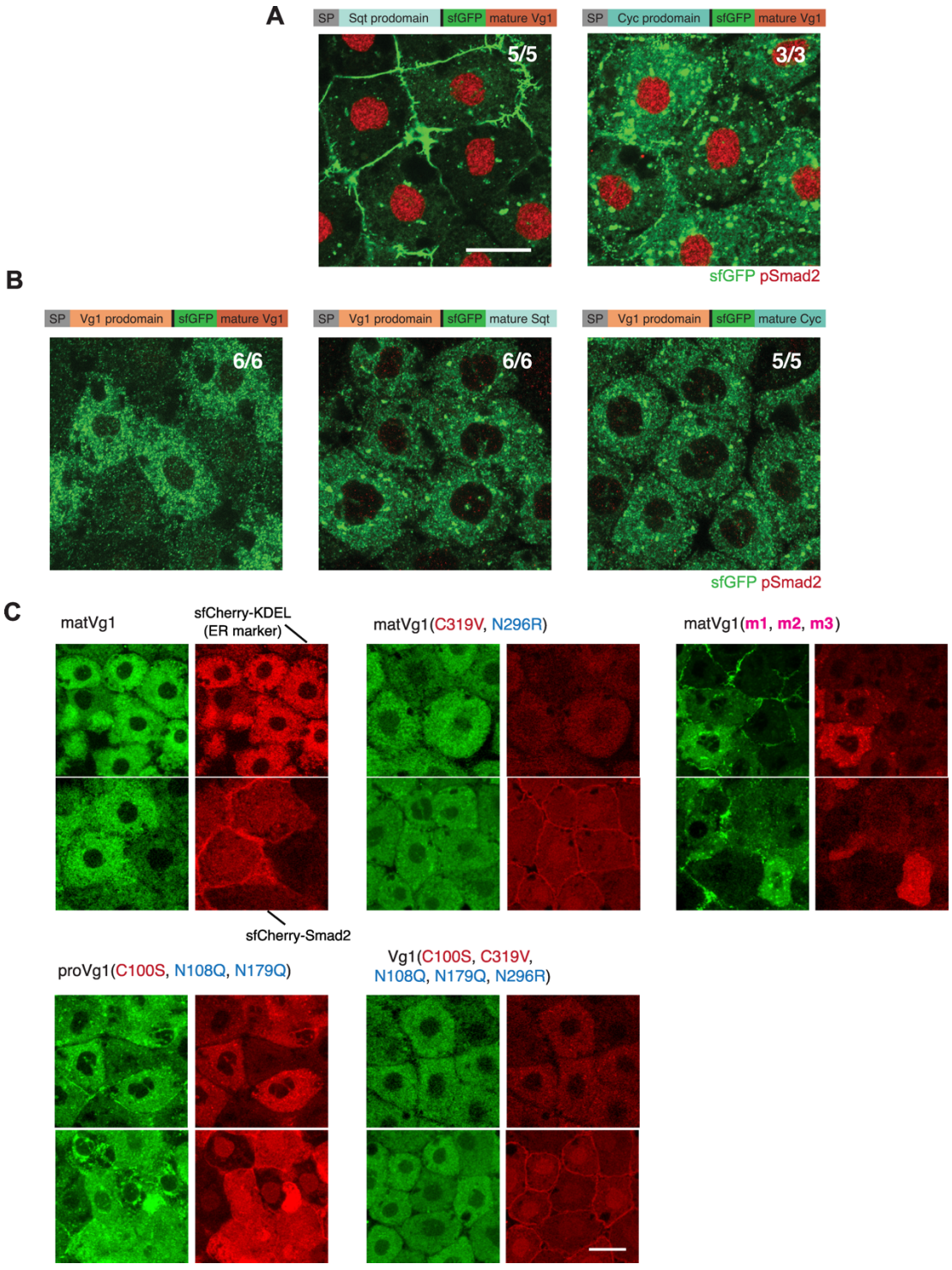

Fig. S4. Localization of Vg1 and Nodal variants. **(A)** Immunofluorescence images of fixed M*vg1* embryos injected with 50 pg mRNA of *sqt* prodomain (left) or *cyc* prodomain (right) fused to *sfGFP*-tagged mature *vg1* domain; or injected with 50 pg mRNA of (**B**) *vg1-sfGFP* (left), *vg1* prodomain fused to *sfGFP*-tagged mature *sqt* domain (middle) or fused to *sfGFP*-tagged mature *cyc* domain (right). Embryos were immunostained with antibodies against sfGFP and pSmad2. (**C**) Fluorescence images of fixed M*vg1* embryos to verify secretion or localization in the ER of several Vg1 constructs. sfCherry-KDEL was used to mark the ER compartment; whereas sfCherry-Smad2 was used to mark the inner leaflet of the plasma membrane, cytoplasm, and nucleus (see Materials and Methods for construct details). Embryos were injected with 50 pg mRNA per construct. Scale bars, 20 µm.

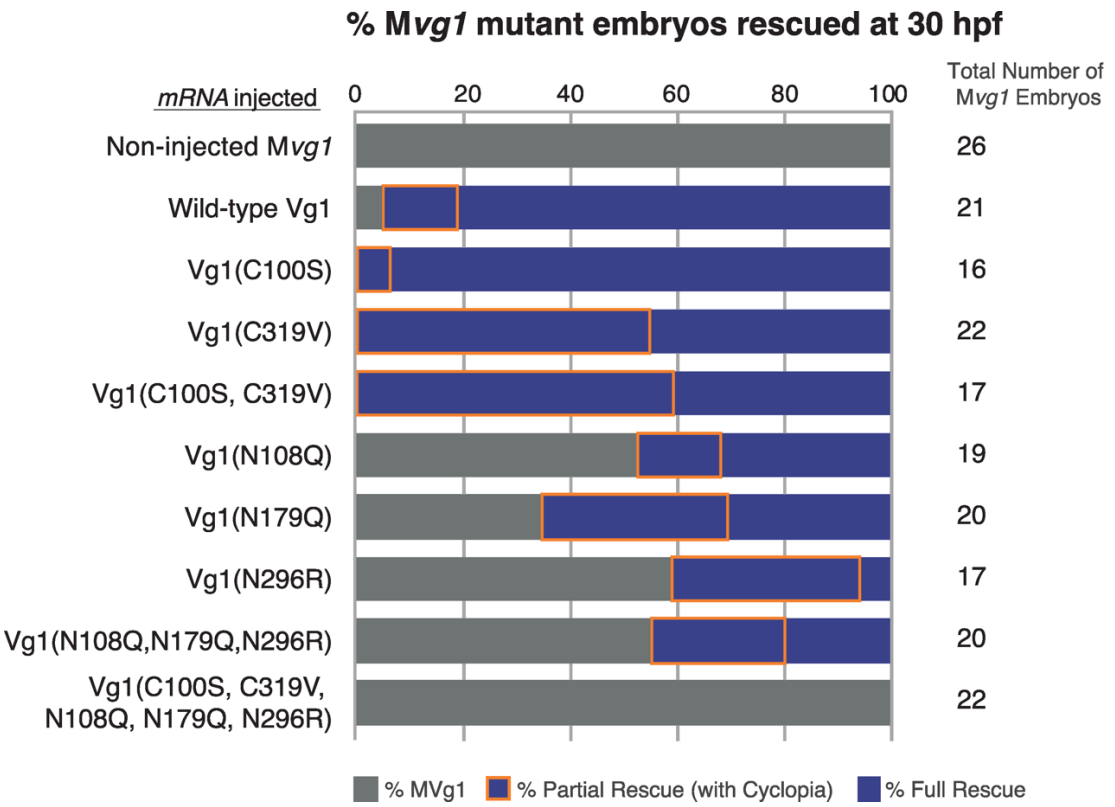

Fig. S5. Vg1 mRNA rescue of M*vg1* embryos. Rescue percentage after 30 hpf of M*vg1* embryos injected with 50 pg of *vg1* mRNAs with the indicated mutations. Full or partial (with cyclopia) rescue percentages are respectively indicated as bars that are blue or blue with orange edges.

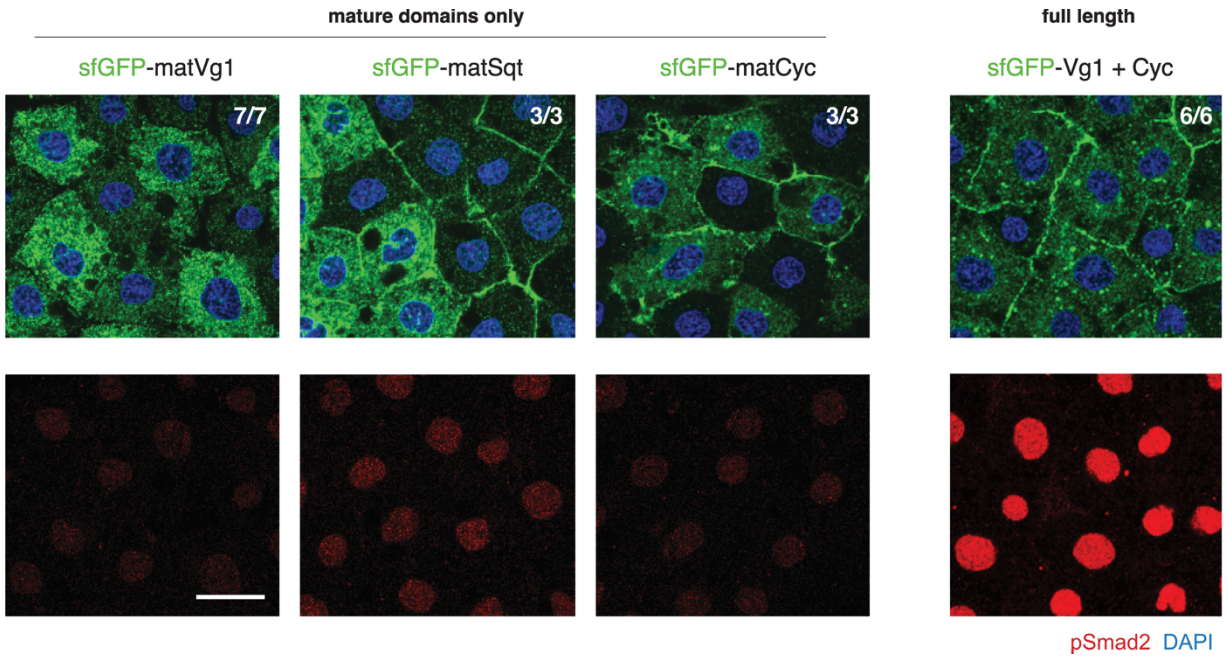

Fig. S6. Localization and activity of Vg1 and Nodal mature domains. **(A)** (Left to right) Immunofluorescence images of fixed M*vg1* embryos injected with 50 pg mRNA of *sfGFP-matVg1*, *sfGFP-matSqt*, *sfGFP-matCyc*, or *sfGFP-matVg1* and *cyc*. The observed fluorescence in the ER for matCyc and matSqt conditions may be a result of delayed or inefficient release of the overexpressed proteins. Embryos were immunostained with antibodies against sfGFP and pSmad2. DAPI, nuclei. Scale bar, 20 µm.

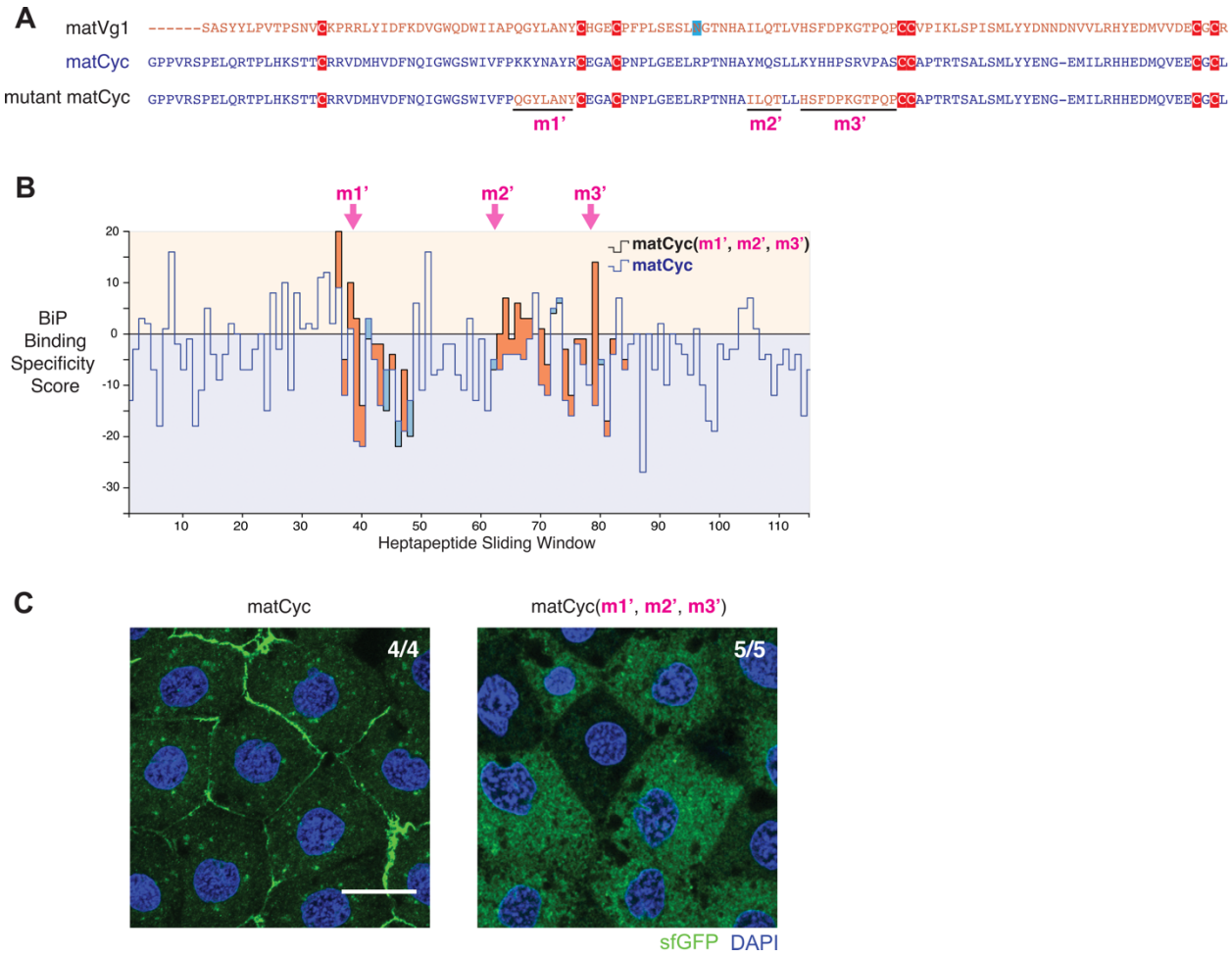

Fig. S7. ER retention of matCyc using BiP-binding features of matVg1. (**A**) Amino acid sequence alignment of the mature domains of Vg1 (matVg1) and Cyc (matCyc) and mutant matCyc. Cysteines (red) and asparagines (blue) are highlighted. (**B**) Difference charts of the BiP binding specificity scores (34) between wild-type and mutant matCyc sequences along a sliding window of seven amino acids. Orange fills indicate the mutant matCyc (black line) score > matCyc (blue line) score. Conversely, blue fills indicate matCyc score > mutant matCyc score. Note the gain of orange fills in *m1'*, *m2'*, and *m3'* regions (arrows). (**C**) Fluorescence images of fixed M*vg1* embryos injected with 50 pg mRNA of *sfGFP*-tagged *matCyc* and *sfGFP*-tagged mutant *matCyc(m1', m2', m3')*. *sfGFP* was inserted upstream of the mature *cyc* domain in all constructs. DAPI, nuclei. Scale bar, 20 μm.

Table S1. BiP Binding Specificity Scores of matVg1, matCyc, matVg1(*m1*, *m2*, *m3*), and matCyc(m1', m2', m3') sequences. These scores are visualized in Figure 4B-C and in SI Appendix, Fig. S7.

	matVg1		matCyc		matVg1 (*m1*, *m2*, *m3*)		matCyc (*m1'*, *m2'*, *m3'*)	
Window #	**Sequence**	**Score**	**Sequence**	**Score**	**Sequence**	**Score**	**Sequence**	**Score**
1			GPPVRSP	**-13**			GPPVRSP	**-13**
2			PPVRSPE	**-3**			PPVRSPE	**-3**
3			PVRSPEL	**3**			PVRSPEL	**3**
4			VRSPELQ	**2**			VRSPELQ	**2**
5			RSPELQR	**-7**			RSPELQR	**-7**
6			SPELQRT	**-18**			SPELQRT	**-18**
7	SASYYLP	**-6**	PELQRTP	**1**	SASYYLP	**-6**	PELQRTP	**1**
8	ASYYLPV	**6**	ELQRTPL	**16**	ASYYLPV	**6**	ELQRTPL	**16**
9	SYYLPVT	**-15**	LQRTPLH	**-2**	SYYLPVT	**-15**	LQRTPLH	**-2**
10	YYLPVTP	**-5**	QRTPLHK	**-7**	YYLPVTP	**-5**	QRTPLHK	**-7**
11	YLPVTPS	**1**	RTPLHKS	**-1**	YLPVTPS	**1**	RTPLHKS	**-1**
12	LPVTPSN	**-24**	TPLHKST	**-18**	LPVTPSN	**-24**	TPLHKST	**-18**
13	PVTPSNV	**-7**	PLHKSTT	**-11**	PVTPSNV	**-7**	PLHKSTT	**-11**
14	VTPSNVC	**1**	LHKSTTC	**5**	VTPSNVC	**1**	LHKSTTC	**5**
15	TPSNVCK	**-12**	HKSTTCR	**-4**	TPSNVCK	**-12**	HKSTTCR	**-4**
16	PSNVCKP	**-14**	KSTTCRR	**-9**	PSNVCKP	**-14**	KSTTCRR	**-9**
17	SNVCKPR	**-3**	STTCRRV	**-4**	SNVCKPR	**-3**	STTCRRV	**-4**
18	NVCKPRR	**-13**	TTCRRVD	**2**	NVCKPRR	**-13**	TTCRRVD	**2**
19	VCKPRRL	**5**	TCRRVDM	**0**	VCKPRRL	**5**	TCRRVDM	**0**
20	CKPRRLY	**6**	CRRVDMH	**-7**	CKPRRLY	**6**	CRRVDMH	**-7**
21	KPRRLYI	**-2**	RRVDMHV	**-7**	KPRRLYI	**-2**	RRVDMHV	**-7**
22	PRRLYID	**-5**	RVDMHVD	**-3**	PRRLYID	**-5**	RVDMHVD	**-3**
23	RRLYIDF	**2**	VDMHVDF	**0**	RRLYIDF	**2**	VDMHVDF	**0**
24	RLYIDFK	**-10**	DMHVDFN	**-15**	RLYIDFK	**-10**	DMHVDFN	**-15**
25	LYIDFKD	**3**	MHVDFNQ	**8**	LYIDFKD	**3**	MHVDFNQ	**8**
26	YIDFKDV	**-19**	HVDFNQI	**-3**	YIDFKDV	**-19**	HVDFNQI	**-3**
27	IDFKDVG	**-3**	VDFNQIG	**10**	IDFKDVG	**-3**	VDFNQIG	**10**
28	DFKDVGW	**-7**	DFNQIGW	**-11**	DFKDVGW	**-7**	DFNQIGW	**-11**
29	FKDVGWQ	**8**	FNQIGWG	**8**	FKDVGWQ	**8**	FNQIGWG	**8**
30	KDVGWQD	**1**	NQIGWGS	**1**	KDVGWQD	**1**	NQIGWGS	**1**
31	DVGWQDW	**-1**	QIGWGSW	**2**	DVGWQDW	**-1**	QIGWGSW	**2**
32	VGWQDWI	**9**	IGWGSWI	**1**	VGWQDWI	**9**	IGWGSWI	**1**
33	GWQDWII	**11**	GWGSWIV	**11**	GWQDWII	**11**	GWGSWIV	**11**
34	WQDWIIA	**0**	WGSWIVF	**12**	WQDWIIA	**0**	WGSWIVF	**12**
35	QDWIIAP	**5**	GSWIVFP	**2**	QDWIIAP	**5**	GSWIVFP	**2**
36	DWIIAPQ	**12**	SWIVFPK	**9**	DWIIAPK	**1**	SWIVFPQ	**20**
37	WIIAPQG	**-2**	WIVFPKK	**-12**	WIIAPKK	**-9**	WIVFPQG	**-5**
38	IIAPQGY	**-3**	IVFPKKY	**1**	IIAPKKY	**-12**	IVFPQGY	**10**
39	IAPQGYL	**1**	VFPKKYN	**-21**	IAPKKYN	**-23**	VFPQGYL	**3**
40	APQGYLA	**-26**	FPKKYNA	**-22**	APKKYNA	**-34**	FPQGYLA	**-14**
41	PQGYLAN	**-1**	PKKYNAY	**3**	PKKYNAY	**3**	PQGYLAN	**-1**
42	QGYLANY	**-2**	KKYNAYR	**-5**	KKYNAYR	**-5**	QGYLANY	**-2**
43	GYLANYC	**-2**	KYNAYRC	**-14**	KYNAYRC	**-14**	GYLANYC	**-2**
44	YLANYCH	**-16**	YNAYRCE	**-7**	YNAYRCH	**-8**	YLANYCE	**-15**
45	LANYCHG	**-1**	NAYRCEG	**-7**	NAYRCHG	**-4**	LANYCEG	**-4**
46	ANYCHGE	**-11**	AYRCEGA	**-17**	AYRCHGE	**-6**	ANYCEGA	**-22**
47	NYCHGEC	**-2**	YRCEGAC	**-19**	YRCHGEC	**-14**	NYCEGAC	**-7**
48	YCHGECP	**-10**	RCEGACP	**-13**	RCHGECP	**-3**	YCEGACP	**-20**

49	CHGECPF	**9**	CEGACPN	**6**	CHGECPF	**9**	CEGACPN	**6**
50	HGECPFP	**-13**	EGACPNP	**-11**	HGECPFP	**-13**	EGACPNP	**-11**
51	GECPFPL	**19**	GACPNPL	**16**	GECPFPL	**19**	GACPNPL	**16**
52	ECPFPLS	**-8**	ACPNPLG	**-8**	ECPFPLS	**-8**	ACPNPLG	**-8**
53	CPFPLSE	**5**	CPNPLGE	**-7**	CPFPLSE	**5**	CPNPLGE	**-7**
54	PFPLSES	**-7**	PNPLGEE	**-2**	PFPLSES	**-7**	PNPLGEE	**-2**
55	FPLSESL	**6**	NPLGEEL	**-2**	FPLSESL	**6**	NPLGEEL	**-2**
56	PLSESLN	**-19**	PLGEELR	**-8**	PLSESLR	**-14**	PLGEELR	**-8**
57	LSESLNG	**-5**	LGEELRP	**-11**	LSESLRP	**-7**	LGEELRP	**-11**
58	SESLNGT	**-3**	GEELRPT	**3**	SESLRPT	**11**	GEELRPT	**3**
59	ESLNGTN	**-8**	EELRPTN	**-13**	ESLRPTN	**-14**	EELRPTN	**-13**
60	SLNGTNH	**-13**	ELRPTNH	**-1**	SLRPTNH	**-2**	ELRPTNH	**-1**
61	LNGTNHA	**-9**	LRPTNHA	**-15**	LRPTNHA	**-15**	LRPTNHA	**-15**
62	NGTNHAI	**1**	RPTNHAY	**-5**	RPTNHAI	**-7**	RPTNHAI	**-7**
63	GTNHAIL	**-1**	PTNHAYM	**-7**	PTNHAIL	**0**	PTNHAIL	**0**
64	TNHAILQ	**7**	TNHAYMQ	**-4**	TNHAILQ	**7**	TNHAILQ	**7**
65	NHAILQT	**-1**	NHAYMQS	**-4**	NHAILQT	**-1**	NHAILQT	**-1**
66	HAILQTL	**6**	HAYMQSL	**-4**	HAILQTL	**6**	HAILQTL	**6**
67	AILQTLV	**-3**	AYMQSLL	**-5**	AILQTLV	**-3**	AILQTLL	**3**
68	ILQTLVH	**-2**	YMQSLLK	**-1**	ILQTLVK	**-7**	ILQTLLH	**3**
69	LQTLVHS	**5**	MQSLLKY	**8**	LQTLVKY	**2**	LQTLLHS	**8**
70	QTLVHSF	**-2**	QSLLKYH	**-10**	QTLVKYH	**-9**	QTLLHSF	**1**
71	TLVHSFD	**-9**	SLLKYHH	**-12**	TLVKYHH	**-14**	TLLHSFD	**-6**
72	LVHSFDP	**6**	LLKYHHP	**5**	LVKYHHP	**7**	LLHSFDP	**4**
73	VHSFDPK	**4**	LKYHHPS	**7**	VKYHHPS	**5**	LHSFDPK	**6**
74	HSFDPKG	**-3**	KYHHPSR	**-13**	KYHHPSR	**-13**	HSFDPKG	**-3**
75	SFDPKGT	**-12**	YHHPSRV	**-16**	YHHPSRV	**-16**	SFDPKGT	**-12**
76	FDPKGTP	**-1**	HHPSRVP	**-2**	HHPSRVP	**-2**	FDPKGTP	**-1**
77	DPKGTPQ	**-1**	HPSRVPA	**-6**	HPSRVPA	**-6**	DPKGTPQ	**-1**
78	PKGTPQP	**-10**	PSRVPAS	**-10**	PSRVPAS	**-10**	PKGTPQP	**-10**
79	KGTPQPC	**14**	SRVPASC	**-14**	SRVPASC	**-14**	KGTPQPC	**14**
80	GTPQPCC	**-6**	RVPASCC	**-5**	RVPASCC	**-5**	GTPQPCC	**-6**
81	TPQPCCV	**-6**	VPASCCA	**-20**	VPASCCV	**-9**	TPQPCCA	**-17**
82	PQPCCVP	**-4**	PASCCAP	**-4**	PASCCVP	**-7**	PQPCCAP	**-1**
83	QPCCVPI	**6**	ASCCAPT	**7**	ASCCVPI	**6**	QPCCAPT	**7**
84	PCCVPIK	**-11**	SCCAPTR	**-7**	SCCVPIK	**-13**	PCCAPTR	**-5**
85	CCVPIKL	**6**	CCAPTRT	**-2**	CCVPIKL	**6**	CCAPTRT	**-2**
86	CVPIKLS	**-4**	CAPTRTS	**0**	CVPIKLS	**-4**	CAPTRTS	**0**
87	VPIKLSP	**-13**	APTRTSA	**-27**	VPIKLSP	**-13**	APTRTSA	**-27**
88	PIKLSPI	**5**	PTRTSAL	**0**	PIKLSPI	**5**	PTRTSAL	**0**
89	IKLSPIS	**-2**	TRTSALS	**-7**	IKLSPIS	**-2**	TRTSALS	**-7**
90	KLSPISM	**-1**	RTSALSM	**2**	KLSPISM	**-1**	RTSALSM	**2**
91	LSPISML	**-3**	TSALSML	**-10**	LSPISML	**-3**	TSALSML	**-10**
92	SPISMLY	**-5**	SALSMLY	**0**	SPISMLY	**-5**	SALSMLY	**0**
93	PISMLYY	**3**	ALSMLYY	**-2**	PISMLYY	**3**	ALSMLYY	**-2**
94	ISMLYYD	**-4**	LSMLYYE	**-4**	ISMLYYD	**-4**	LSMLYYE	**-4**
95	SMLYYDN	**-8**	SMLYYEN	**-8**	SMLYYDN	**-8**	SMLYYEN	**-8**
96	MLYYDNN	**-5**	MLYYENG	**1**	MLYYDNN	**-5**	MLYYENG	**1**
97	LYYDNND	**-1**	LYYENGE	**-10**	LYYDNND	**-1**	LYYENGE	**-10**
98	YYDNNDN	**-15**	YYENGEM	**-17**	YYDNNDN	**-15**	YYENGEM	**-17**
99	YDNNDNV	**-13**	YENGEMI	**-19**	YDNNDNV	**-13**	YENGEMI	**-19**
100	DNNDNVV	**-20**	ENGEMIL	**-2**	DNNDNVV	**-20**	ENGEMIL	**-2**
101	NNDNVVL	**2**	NGEMILR	**-5**	NNDNVVL	**2**	NGEMILR	**-5**
102	NDNVVLR	**-9**	GEMILRH	**-3**	NDNVVLR	**-9**	GEMILRH	**-3**

103	DNVVLRH	**-14**			DNVVLRH	**-14**		
104	NVVLRHY	**5**	EMILRHH	**5**	NVVLRHY	**5**	EMILRHH	**5**
105	VVLRHYE	**1**	MILRHHE	**7**	VVLRHYE	**1**	MILRHHE	**7**
106	VLRHYED	**-3**	ILRHHED	**1**	VLRHYED	**-3**	ILRHHED	**1**
107	LRHYEDM	**-1**	LRHHEDM	**-5**	LRHYEDM	**-1**	LRHHEDM	**-5**
108	RHYEDMV	**-14**	RHHEDMQ	**-4**	RHYEDMV	**-14**	RHHEDMQ	**-4**
109	HYEDMVV	**-15**	HHEDMQV	**-12**	HYEDMVV	**-15**	HHEDMQV	**-12**
110	YEDMVVD	**-11**	HEDMQVE	**-6**	YEDMVVD	**-11**	HEDMQVE	**-6**
111	EDMVVDE	**-5**	EDMQVEE	**-3**	EDMVVDE	**-5**	EDMQVEE	**-3**
112	DMVVDEC	**-11**	DMQVEEC	**-7**	DMVVDEC	**-11**	DMQVEEC	**-7**
113	MVVDECG	**2**	MQVEECG	**-4**	MVVDECG	**2**	MQVEECG	**-4**
114	VVDECGC	**-8**	QVEECGC	**-16**	VVDECGC	**-8**	QVEECGC	**-16**
115	VDECGCR	**-14**	VEECGCL	**-7**	VDECGCR	**-14**	VEECGCL	**-7**