The Company of **Biologists**

RESEARCH ARTICLE

Modelling the structure of Short Gastrulation and generation of a toolkit for studying its function in *Drosophila*

Sophie L. Frampton[1], Catherine Sutcliffe[1], Clair Baldock[2,*] and Hilary L. Ashe[1,*]

ABSTRACT

A BMP gradient is essential for patterning the dorsal-ventral axis of invertebrate and vertebrate embryos. The extracellular BMP binding protein Short Gastrulation (Sog) in *Drosophila* plays a key role in BMP gradient formation. In this study, we combine genome editing, structural and developmental approaches to study Sog function in *Drosophila*. We generate a *sog* knockout fly stock, which allows simple reintegration of altered versions of the *sog* coding sequence. As proof-of-principle, we test the requirement for two cysteine residues that were previously identified as targets for palmitoylation, which has been proposed to enhance Sog secretion. However, we show that the *sog*[C27,28S] mutant is viable with only very mild phenotypes, indicating that these residues and their potential modification are not critical for Sog secretion *in vivo*. Additionally, we use experimental negative stain EM imaging and hydrodynamic data to validate the AlphaFold structure prediction for Sog. The model suggests a more compact shape than the vertebrate ortholog Chordin and conformational flexibility between the C-terminal von Willebrand C domains. We discuss how this altered compactness may contribute to mechanistic differences in Sog and Chordin function during BMP gradient formation.

This article has an associated First Person interview with the first author of the paper.

KEY WORDS: ***Drosophila***, **Short gastrulation, CRISPR-Cas9, AlphaFold, Model, Palmitoylation**

INTRODUCTION

Bone morphogenetic proteins (BMPs) are a large class of highly conserved signalling molecules that belong to the TGF-beta superfamily. BMPs perform essential roles during animal development and adult tissue homeostasis, the significance of which is reflected in the variety of human diseases attributed to aberrant BMP activity (Bandyopadhyay et al., 2013; Wang et al., 2014). BMPs bind to their receptors resulting in the phosphorylation of a receptor regulated Smad, which then forms a complex with the common mediator Smad that accumulates in the nucleus

to regulate target gene transcription (Schmierer and Hill, 2007). A key developmental role for BMPs is the patterning of the dorsal-ventral (DV) axis in early vertebrate and invertebrate embryos. BMP gradient formation is mediated by a conserved network of regulators, including two BMP binding proteins, Sog/Chordin and Twisted Gastrulation (Tsg), as well as a protease, Tolloid (Tld) (Madamanchi et al., 2021; Matsuda et al., 2016; Shilo et al., 2013).

The most potent BMP signalling molecule in the early *Drosophila* embryo is a heterodimer of the BMP ligands Decapentaplegic (Dpp) and Screw (Scw) (Shimmi et al., 2005b), which have uniform expression in the dorsal ectoderm (St Johnston and Gelbart, 1987; Arora et al., 1994; Shimmi et al., 2005b). *tsg* and *tld* are also expressed in the dorsal ectoderm, while *sog* is expressed ventro-laterally in the neuroectoderm (Francois et al., 1994; Marqués et al., 1997; Mason et al., 1994). During embryogenesis, reciprocal gradients of Dpp/Scw and Sog are established across the dorsal ectoderm (Ferguson and Anderson, 1992; Ashe and Levine, 1999; Srinivasan et al., 2002; Shimmi et al., 2005b; Wang and Ferguson, 2005). A narrow stripe of peak BMP signalling occurs along the dorsal midline and is flanked by lower signalling levels (Dorfman and Shilo, 2001; Rushlow et al., 2001; Sutherland et al., 2003; Shimmi et al., 2005b) thereby subdividing the dorsal ectoderm into amnioserosa and dorsal epidermis, respectively (Raftery and Sutherland, 2003).

A favoured model of BMP gradient formation requires the shuttling of BMP ligands dorsally in a multi-protein complex (Holley et al., 1996; Marqués et al., 1997; Eldar et al., 2002; Shimmi et al., 2005b; Umulis et al., 2006, 2009; Sawala et al., 2012). The model proposes that secreted Dpp/Scw binds to the extracellular matrix protein collagen IV (Col IV), which acts as a scaffold to promote formation of a Dpp/Scw-Sog-Tsg complex (Sawala et al., 2012; Wang et al., 2008). In this inhibitory complex, the Dpp/Scw ligand is unable to interact with its receptors but can diffuse dorsally (Ross et al., 2001; Eldar et al., 2002; Shimmi et al., 2005b; Sawala et al., 2012). Cleavage of Sog within this complex by Tld liberates Dpp/Scw, allowing the ligand to re-bind Col IV. In dorso-lateral regions, close to the *sog* expression domain, the Dpp/Scw-Sog-Tsg complex is reassembled, resulting in inhibition of signalling and diffusion of the complex towards the dorsal midline. At the dorsal midline and in the absence of Sog, however, Dpp/Scw is free to interact with receptors, resulting in a graded BMP signal across the dorsal ectoderm (Sawala et al., 2012; Wang et al., 2008; Winstanley et al., 2015).

Sog function is also important in *Drosophila* pupal wing vein patterning, including formation of the posterior crossvein (PCV), which depends on signalling by Dpp and Glass bottomed boat (Gbb) ligands (Serpe et al., 2005; Wharton et al., 1999; Yu et al., 1996). As in embryogenesis, Sog functions with a Tsg-like protein, Crossveinless (Cv), and a Tolloid-related (Tlr) metalloprotease to both locally inhibit BMP signalling and enhance it at a distance from the source in the pupal wing (Ralston and Blair, 2005; Serpe

Biology Open

1

et al., 2005; Vilmos et al., 2005; Shimmi et al., 2005a). In this model, Dpp/Gbb is transported from the longitudinal veins in a Dpp/Gbb-Sog-Cv complex to the presumptive PCV, where it is released from the inhibitory complex by Tlr-mediated Sog cleavage, enabling ligand-receptor interactions (Shimmi et al., 2005a; Serpe et al., 2005).

Sog and its vertebrate ortholog Chordin each contain four cysteine rich von Willebrand type C (vWC) domains which mediate protein interactions. These domains are 60–80 residues in length and have been identified in approximately 500 extracellular matrix proteins (Garcia Abreu et al., 2002; O'Leary et al., 2004; Zhang et al., 2007). Sog/Chordin vWC1 is separated from vWC2/3/4 domains by a 'stem' region comprising four Chordin specific (CHRD) domains (Francois et al., 1994). Structures have been solved for the human Procollagen IIA and the zebrafish Crossveinless-2 vWC1 domains (O'Leary et al., 2004; Xu et al., 2017; Zhang et al., 2008); however, there is currently no experimental structure for these domains in Sog, or the Sog/Chordin specific 4x CHRD 'stem' region.

Sog secretion is critical to its function, and a previous study has reported a potential role for palmitoylation during this process (Kang and Bier, 2010). Palmitoylation is a lipid modification that can influence protein interactions, membrane association, and trafficking between sub-cellular compartments (Bannan et al., 2008; Kang and Bier, 2010; Linder and Deschenes, 2003). The *Drosophila* palmitoyl-transferase Huntingtin-interacting protein 14 (dHIP14) was identified as an interacting partner of Sog, and Sog was shown to be palmitoylated in tissue culture cells (Giot et al., 2003; Kang and Bier, 2010). Mis-expression of dHIP14 in embryos and wings reduced BMP activity, similar to Sog overexpression phenotypes. In addition, mutation of cysteines 27 and 28 of Sog prevented the dHIP14-mediated increase in Sog secretion in tissue culture, suggesting that these two residues are the primary palmitoylation targets (Kang and Bier, 2010).

In this study, we generate a *sog* knockout (KO) 'reintegration-ready' fly stock that we use to test the effect of mutating the palmitoylation sites *in vivo*. Our data show that these residues are not critical for Sog function. In addition, we combine EM imaging
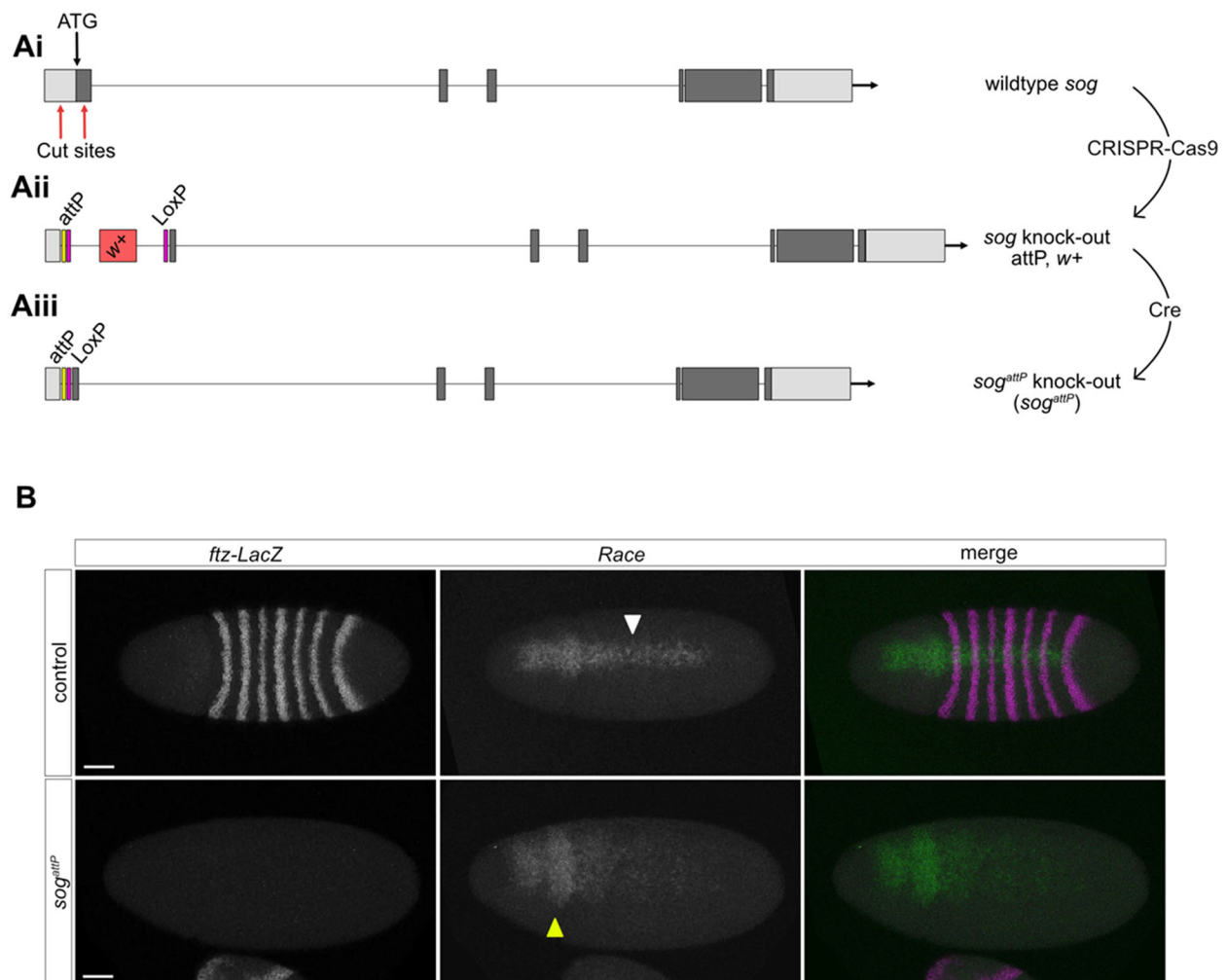

Fig. 1. A CRISPR-Cas9 generated *sog* KO mutant. (Ai) Cartoon shows the positions of the CRISPR-Cas9 cut sites (red arrows), which are located in the 5'UTR and downstream of the ATG and signal sequence in the first protein coding exon. (Aii) CRISPR-Cas9 with HDR was used to insert an attP sequence, two LoxP sequences, and a *white+* marker gene. (Aiii) Cre-Lox recombination removes the *white+* marker gene shown in Aii. The genome of the resulting fly line does not include the *sog* start codon or signal sequence, which are replaced with attP and LoxP sequences. (B) RNA *in situ* hybridisation for the BMP target gene *Race*. *Race* is expressed in heterozygous *sog^attP*/FM7c *ftz-lacZ* embryos, the white arrowhead indicates *Race* expression in the presumptive amnioserosa (top panel). Heterozygotes are identified by *lacZ* expression from the FM7c *ftz-lacZ* balancer. *Race* expression is lost in the presumptive amnioserosa and the 'head spots' are broader (yellow arrowhead) in *sog^attP* embryos. Scale bars: 50 µm.

and AlphaFold prediction to model Sog structure, which reveals a curved compact shape. This Sog structure, along with the *sog* KO fly stock that we describe, will facilitate a complete molecular dissection of this critical extracellular BMP regulator.

RESULTS

Generation of a *sog* KO with CRISPR

To facilitate analysis of Sog *in vivo*, we used CRISPR genome editing to generate a *sog* KO line in which the translation start codon of the endogenous *sog* locus on the X chromosome was replaced with a phiC31 recombination landing site (Baena-Lopez et al., 2013) (Fig. 1A). Specifically, CRISPR-Cas9 mediated homology directed repair (HDR) was used to delete 800 bp including the ATG and signal sequence (Fig. 1Ai) and replace these with an attP recombination site. The resulting *Drosophila* stock facilitates simple insertion of modified *sog* sequences, such as point mutants

(Fig. 2A), for expression under the endogenous *sog* promoter. In addition, regulatory elements located within *sog* introns, for instance the *sog* primary enhancer, remain intact (Markstein et al., 2002). The *white* gene was included in the HDR template and used as a marker to identify successful CRISPR events (Fig. 1Aii), before removal by Cre-Lox recombination (Fig. 1Aiii).

The resulting *sog*^{attP} stock is maintained with an X-chromosomal balancer, and the absence of non-balancer males in the stock is consistent with a loss of Sog function. In addition, insertion of the attP recombination sequence at the *sog* locus was confirmed by sequencing. We used single molecule FISH (smFISH) to quantitate the amount of transcription from the *sog*^{attP} locus. The smFISH probes detect both *sog* transcription and mRNAs in early nuclear cycle (nc) 14 wild-type and *sog*^{attP} embryos (Fig. S1). To estimate *sog* mRNA number/cell, we assigned *sog* mRNAs to the closest nucleus as these embryos are only starting to cellularise. This
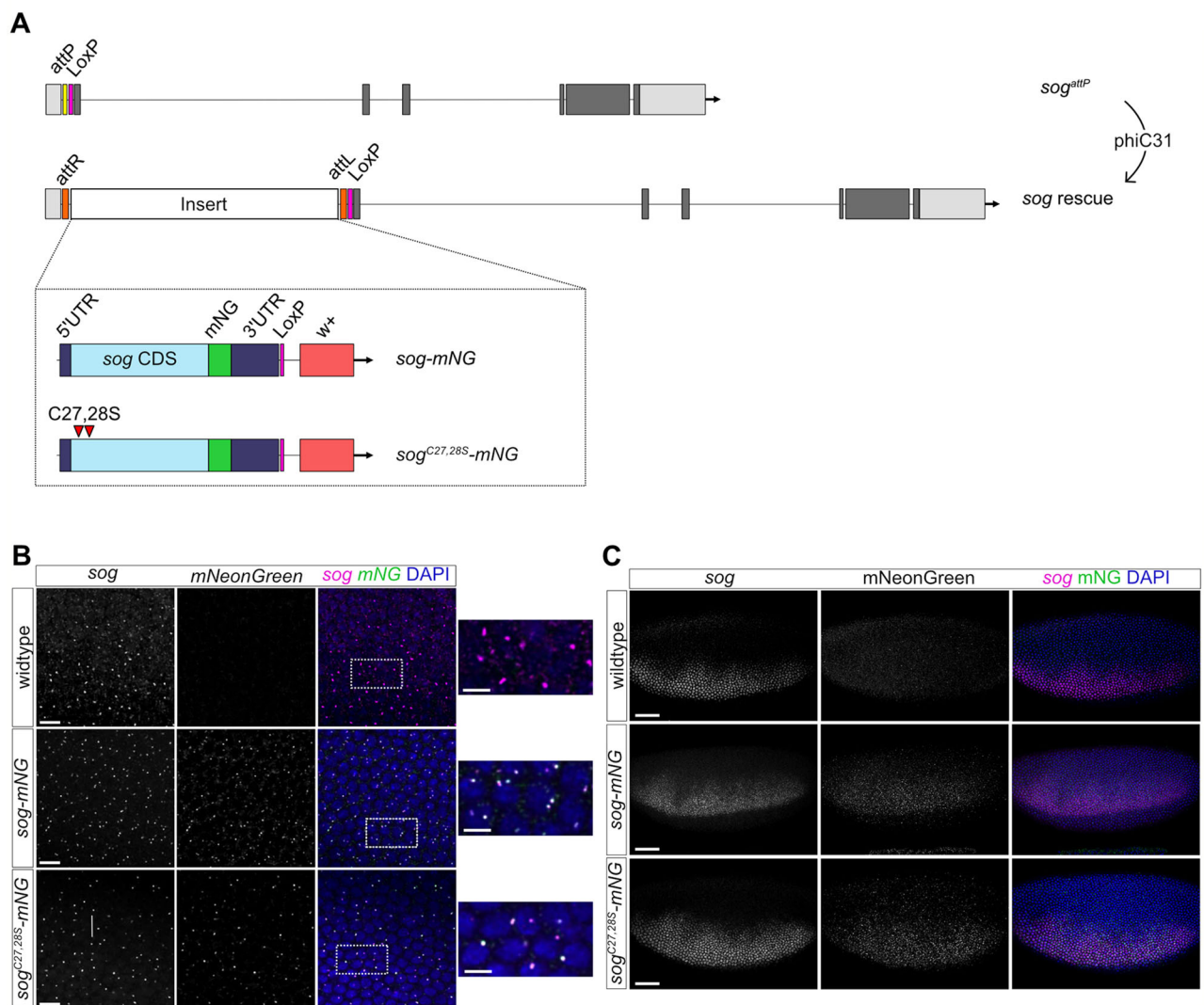

Fig. 2. Insertion of specific *sog* coding sequences into the endogenous *sog* locus. (A) Specific *sog* coding sequences fused to a C-terminal mNG tag were inserted into the genome by phiC31 recombination at the inserted attP landing site of the *sog*^{attP} line. Wild-type Sog and a Cys27,28Ser mutant were reintegrated. (B) Embryos (nc14) of the indicated genotypes showing fluorescent RNA *in situ* hybridisation staining with *sog* and *mNG* probes (magenta and green, respectively). Scale bars: 10 μm. Expanded views of the areas outlined in the merged images are shown, scale bars: 5 μm. Colocalisation of *sog* (magenta) and *mNG* (green) transcription sites is indicated by white spots in the merged images. (C) *sog* smFISH (magenta) and mNG immunostaining (green) of the indicated nc14 embryos. A merged image with DAPI staining (blue) is shown, with single images of the *sog* and mNG channels for clarity. Scale bars: 50 μm.

analysis reveals that there is a ~2.5-fold reduction in the peak numbers of *sog* mRNAs/cell in the *sog*^{attP} embryos (Fig. S1). The presence of *sog* mRNAs is consistent with the deletion removing sequences downstream of the *sog* promoter, although the reduced mRNA number suggests that there is an effect on transcription and/or mRNA stability, potentially due to nonsense-mediated decay.

A scan for cryptic start codons at the modified *sog* locus identified one large and several smaller open reading frames (ORFs) (Fig. S2). Although the large ORF encodes *sog* sequences starting within the first vWC domain, the signal sequence is absent. No cryptic signal sequence in this truncated Sog ORF was predicted using various software tools, e.g. Phobius webserver (Käll et al., 2004; 2007), that were able to predict endogenous Sog's signal sequence/transmembrane domain (data not shown). As Sog is a secreted protein (Francois et al., 1994), no Sog function is predicted following the deletion and attP insertion made in the *sog* locus.

To confirm loss of Sog function in *sog*^{attP} embryos, RNA fluorescence *in situ* hybridisation (FISH) was used to visualise expression of the peak BMP target gene *Race* (Fig. 1B). We also probed for *lacZ* mRNAs as *ftz-lacZ* is present on the balancer chromosome. The absence of *lacZ* expression indicates that the embryos are *sog*^{attP} males. In these *sog*^{attP} embryos *Race* expression is expanded in the anterior and lost in the presumptive amnioserosa (Fig. 1B), consistent with that described in *sog* mutant embryos (Ashe and Levine, 1999). These data, and the lethality of *sog*^{attP} males, support successful removal of the *sog* translation start site and its replacement with an attP recombination sequence to generate a *sog*^{attP} KO allele.

Reintegration of transgenes at the *sog* locus

The presence of an attP landing site at the *sog* locus facilitates targeted insertion of specific *sog* coding sequences into the genome (Fig. 2A). A previous study proposed that palmitoylation at two cysteines, at positions 27 and 28, is important for Sog secretion and stability of a membrane bound form of Sog (Kang and Bier, 2010). To test how the disruption of palmitoylation affects Sog function *in vivo* we used the *sog*^{attP} line we generated to integrate a *sog* cDNA in which Cys27 and 28 are mutated to Ser. Wild-type and palmitoylation mutant versions of the *sog* cDNA, to which a C-terminal mNeonGreen (mNG) tag was added (referred to as *sog-mNG* and *sog*^{C27,28S}-*mNG*, respectively), were integrated into the endogenous locus (Fig. 2A). In total, ~12.6 kb of DNA was inserted at the *sog* locus, including the *sog* CDS, *white+* marker, and LoxP sites (Fig. S3). Although endogenous *sog* sequences remain downstream of the integration site, cryptic initiation within the reintegration sequences and readthrough is not predicted to result in a Sog ORF longer than the truncated one described above. Therefore, if transcription of the remaining endogenous *sog* locus occurs due to cryptic initiation, the mRNA is only predicted to encode a truncated Sog ORF lacking a signal sequence (Fig. S2). This truncated Sog lacks activity based on the phenotype of the *sog*^{attP} embryos and lethality of the *sog*^{attP} males, as described above.

Both male and female flies carrying only the reintegrated *sog-mNG* and *sog*^{C27,28S}-*mNG* sequences are viable (see later). Transcription of the integrated *sog* sequences in lateral stripes in the embryos was confirmed by FISH using *sog* and *mNG* probes (Fig. 2B). In this experiment, the control embryos carry an unmodified *sog* locus, so a signal is only detected with the *sog* probe (Fig. 2B). However, both the *sog* and *mNG* probes detect co-localised signals in the *sog-mNG* and *sog*^{C27,28S}-*mNG* embryos,

as expected for transcription of the reintegrated sequences. We next used smFISH to test for any differences in *sog* expression between *sog-mNG* and *sog*^{C27,28S}-*mNG* early stage 6 embryos compared to wild-type (Fig. S4A). We found no significant difference in the number of *sog* expressing cells between embryos of these genotypes (Fig. S4B). We were unable to quantitate absolute *sog* mRNA numbers at this stage due to their clustering. However, quantitation based on fluorescence intensity in equivalent areas of the expression domain suggests that there are no significant differences in *sog* expression levels in the reintegration embryos (Fig. S4C). The same result was obtained by analysing the fluorescence intensity along a line through the whole expression domain (data not shown). Finally, mNG immunostaining in combination with *sog* smFISH showed accumulation of Sog-mNG and Sog^{C27,28S}-mNG protein (Fig. 2C).

Cysteines 27 and 28, putative palmitoylation targets, are not essential for Sog function

As a fly stock homozygous for *sog*^{C27,28S}-*mNG* was successfully established and maintained, cysteines 27 and 28 are not essential for Sog function. Due to difficulties associated with detecting palmitoylation *in vivo* we were unable to directly compare palmitoylation levels of wild-type and the mutant Sog. However, as the *sog*^{C27,28S} mutant was less able to inhibit BMP activity in a tissue culture assay (Kang and Bier, 2010), we investigated whether these mutations reduce viability *in vivo*. To test this, the survival of embryos to pupal and adult stages was quantified (Fig. S5A). *sog-mNG* or *sog*^{C27,28S}-*mNG* embryos were raised at 25°C and the number of pupae and eclosed adults counted. The proportion of pupae and adults show some lethality at each of these stages for both the *sog-mNG* or *sog*^{C27,28S}-*mNG* lines. This could be due to the presence of the mNG tag or differences in the reintegration locus compared to wild-type (see Discussion). Despite a trend towards lower survival rates for the *sog*^{C27,28S}-*mNG* allele, there is no significant difference between the number of *sog*^{C27,28S}-*mNG* and *sog-mNG* embryos that developed into pupae and successfully eclosed as adults.

To further test the functionality of the *sog-mNG* and *sog*^{C27,28S}-*mNG* sequences, the extent to which these alleles can rescue the *sog*^{S6} loss-of-function allele or the *sog*^{attP} KO allele was assayed. *sog-mNG* and *sog*^{C27,28S}-*mNG* males were crossed to *sog*^{S6}/FM7c or *sog*^{attP}/FM7c females (Fig. S5Bi), and the numbers of female offspring with either the *sog-mNG* or *sog*^{C27,28S}-*mNG* allele versus the FM7c balancer were counted. No significant difference in the ability of the *sog-mNG* or *sog*^{C27,28S}-*mNG* alleles to rescue either *sog* mutant allele relative to wild-type was observed (Fig. S5Bii). Although the different viability assays appear to have different sensitivities (Fig. S5A,B), together the data are consistent with the *sog*^{C27,28S} mutations having only a very minor effect, if any, on Sog function.

BMP signalling readouts in *sog-mNG* and *sog*^{C27,28S}-*mNG* embryos

Although *sog-mNG* and *sog*^{C27,28S}-*mNG* flies are viable, we investigated whether there are minor effects on Dpp gradient formation and interpretation. Sog functions in the early *Drosophila* embryo to concentrate BMP ligands dorsally, resulting in a stripe of the activated pMad transducer at the dorsal midline (Montanari et al., 2022). Therefore, pMad distribution was visualised in early stage 6 embryos by immunostaining and the width of the pMad stripe was measured at 50% embryo length
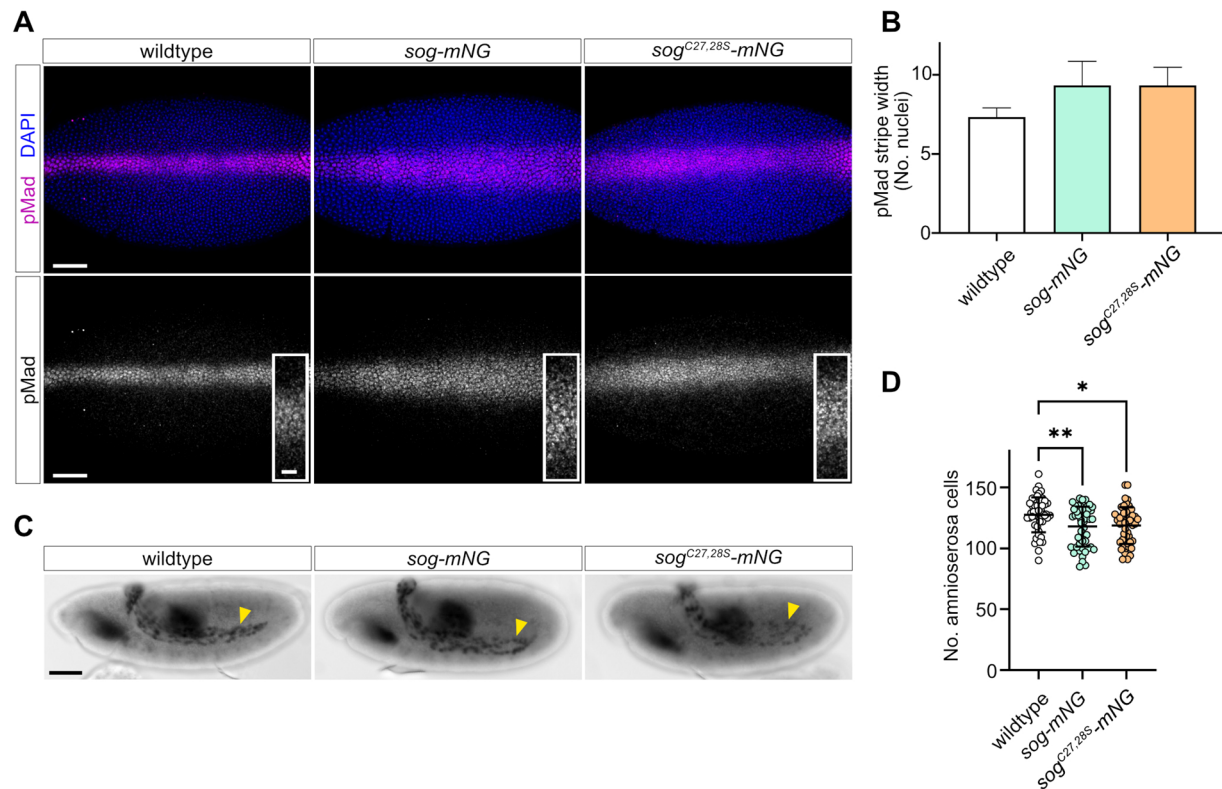
Fig. 3. BMP signalling readouts in *sog-mNG* and *sog^C27,28S^-mNG* embryos. (A) pMad immunostaining in early stage 6 wild-type control, *sog-mNG*, and *sog^C27,28S^-mNG* embryos, scale bars: 50 μm. Insets show a higher magnification view of the central region of each pMad stripe, scale bar: 10 μm. (B) Mean pMad width (in number of nuclei) at ~50% embryo length was measured for each embryo. No significant differences in mean pMad width were found by one-way ANOVA (with Tukey's multiple comparisons test, $P>0.05$). For each genotype, $n=3$. Error bars represent mean+s.d. (C) Hnt immunostaining of stage 11 embryos shows amnioserosa cells (yellow arrowheads) and midgut staining. Scale bar: 50 μm. (D) Quantification of amnioserosa cell numbers in embryos of each genotype is shown. *sog-mNG* and *sog^C27,28S^-mNG* embryos have significantly fewer amnioserosa cells than wild-type controls (one-way ANOVA, Tukey's multiple comparisons test, *sog-mNG* versus control $P=0.006$, *sog^C27,28S^-mNG* versus control $P=0.012$), $n=50$.

(Fig. 3A,B). The pMad stripes in *sog-mNG* and *sog^C27,28S^-mNG* embryos are generally broader than those in wild-type control embryos; however, these differences are not significant (Fig. 3B).

Peak BMP/pMad signalling specifies amnioserosa cell fate. Therefore, to test whether subtle differences in pMad stripe width affected amnioserosa specification, embryos were stained for the amnioserosa cell marker Hindsight (Hnt, Fig. 3C). Both *sog-mNG* and *sog^C27,28S^-mNG* embryos have a small but significant reduction in the number of amnioserosa cells compared to wild-type embryos. However, the *sog-mNG* and *sog^C27,28S^-mNG* embryos have a similar reduction in the number of amnioserosa cells, suggesting that the C27,28S mutations do not affect embryonic BMP-dependent cell fate decisions. Together, these data suggest that BMP signal reception is marginally affected in both *sog-mNG* and *sog^C27,28S^-mNG* embryos, however this level of disruption is tolerated during development.

Quantitative analysis of BMP target gene expression

As the number of amnioserosa cells was reduced in *sog-mNG* and *sog^C27,28S^-mNG* embryos, we used smFISH and quantitative analysis to assess effects on transcription of BMP target genes. smFISH was performed for the BMP target genes *Race* and *u-shaped* (*ush*), which respond to peak and intermediate levels of BMP signalling, respectively (Fig. 4A,B) (Ashe et al., 2000). Both *sog-mNG* and *sog^C27,28S^-mNG* embryos show similar *ush* expression patterns to wild-type (unedited) embryos, indicating

that there is a BMP gradient, consistent with Sog function (Ashe et al., 2000). However, while the number of mature *ush* mRNAs in *sog-mNG* embryos is similar to that in wild-type embryos, *sog^C27,28S^-mNG* embryos have, on average, around half the number of *ush* mRNAs (Fig. 4C; Fig. S6).

Race expression levels in *sog-mNG* embryos are, in general, lower than in controls, but *Race* is restricted to the dorsal midline as in control embryos (Fig. 4D; Fig. S6). The mean number of *Race* mRNAs across the three biological repeat embryos is slightly lower in *sog-mNG* relative to control embryos, whereas there is an even greater reduction in *Race* expression in *sog^C27,28S^-mNG* embryos (Fig. 4D). In addition, the levels of *Race* expression observed in *sog^C27,28S^-mNG* embryos show more variation: while one *sog^C27,28S^-mNG* embryo has a weak stripe of *Race* expression along the dorsal midline (Fig. 4Di), it is almost absent in the others (Fig. S6). As *Race* expression in *sog^C27,28S^-mNG* embryos is weaker than in *sog-mNG* embryos, this suggests that *sog^C27,28S^-mNG* may be less able to promote peak BMP signalling. Overall, this highly sensitive assay of BMP target gene transcription identifies subtle deficiencies in the responses, particularly with *sog^C27,28S^-mNG*, even though these do not have major effects on viability.

The *sog^C27,28S^* mutant shows weakly penetrant PCV patterning defects

Sog also regulates BMP signalling during pupal wing vein patterning, including PCV patterning. Therefore, we used this as
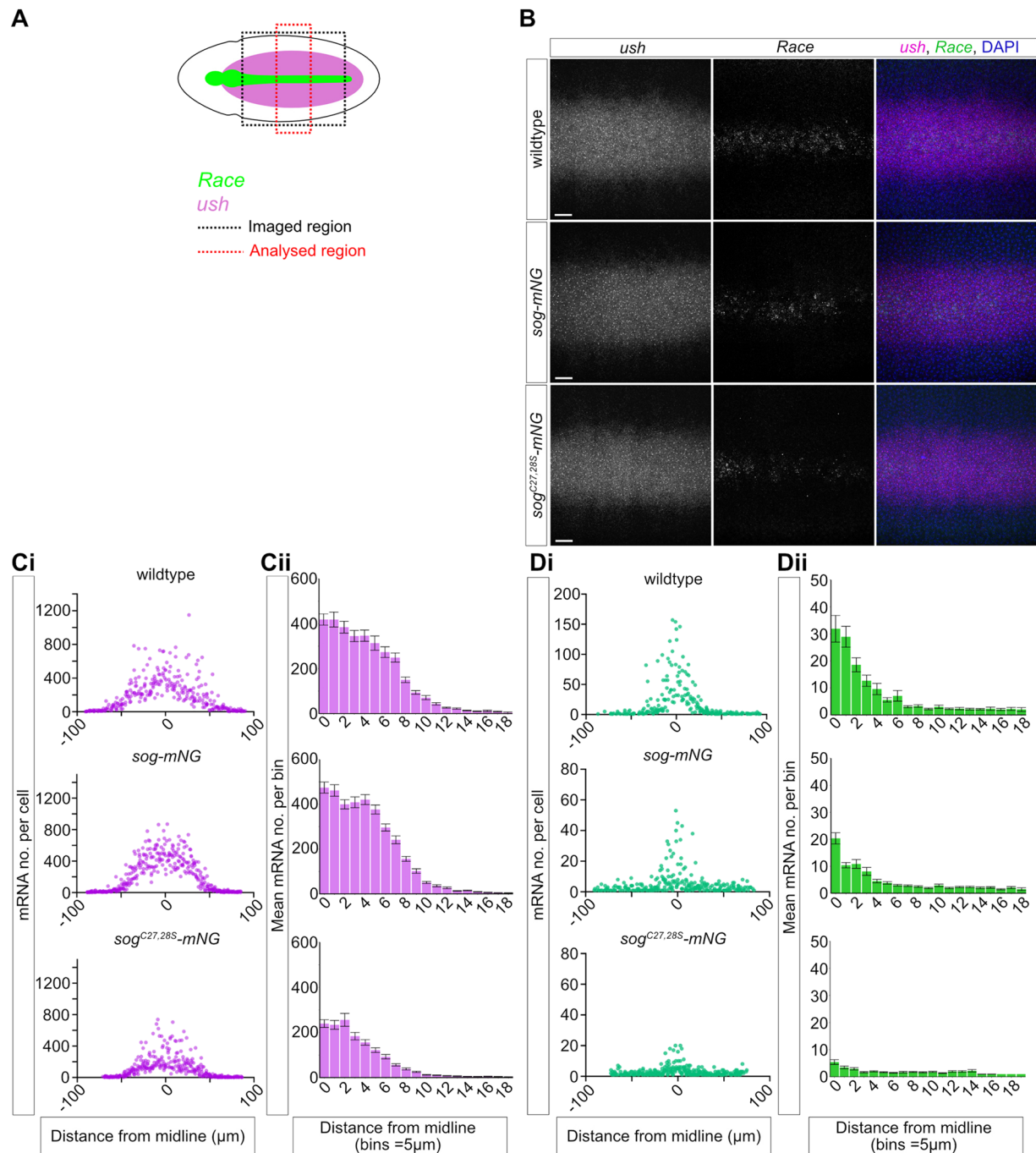
Fig. 4. $sog^{C27,28S}$-mNG embryos have disrupted BMP target gene expression. (A) Cartoon of a stage 6 *Drosophila* embryo showing the *Race* and *ush* expression patterns in green and magenta, respectively. The black box indicates the imaged region of the embryo, the red box represents the portion of the images used for analysis. (B) smFISH detection of *ush* (magenta) and *Race* (green) mRNAs in wild-type, *sog-mNG*, and *sog^{C27,28S}-mNG* stage 6 embryos. Scale bars: 20 μm. (Ci) The numbers of *ush* transcripts per cell were quantified and plotted against the distance from the dorsal midline (μm, midline=0). Analysis was performed for three embryos of each genotype, the results for one representative embryo of each genotype are shown here (see Fig. S6 for the other embryos). (Cii) Counts of *ush* mRNAs per cell for the three replicates of each genotype were combined and binned according to distance from the midline (bins=5 μm, each bin approximately corresponds to a nuclear width). Error bars represent mean±s.e.m. (Di,ii) Data shown are as for Ci,ii, but for *Race*. Note the change in the axis scale for the wild-type control embryo in Di. Error bars represent mean±s.e.m.

an alternative developmental context to test whether the requirement for palmitoylation of Sog may be context dependent. The wings of adult female flies, raised at either 18°C or 25°C, were examined for defects in PCV specification and patterning (Fig. 5). A low proportion of $sog^{C27,28S}$-mNG wings displayed a mutant PCV phenotype: a small extension to the distal side of the PCV (Fig. 5A). Ectopic PCV development was observed in a slightly higher

proportion of flies that developed at 18°C compared to 25°C, suggesting that the phenotype is exacerbated by mild cold temperature stress (Fig. 5B,C). Disruption to the PCV only in $sog^{C27,28S}$-mNG wings at 18°C suggests that mutation of cysteines 27 and 28 has mildly impacted Sog function or levels, resulting in reduced BMP signal refinement during PCV patterning (Antson et al., 2022) (see Discussion).
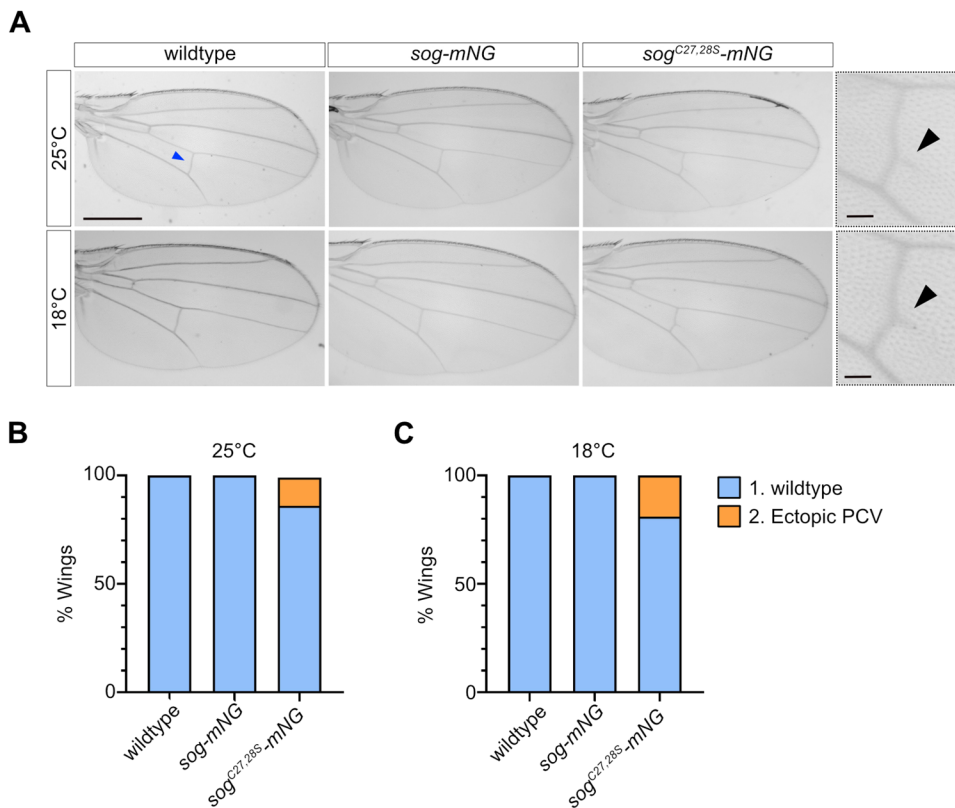
Fig. 5. The *sog^{C27,28S}* mutant shows weakly penetrant PCV patterning defects. (A) Wings from female wild-type control, *sog-mNG*, and *sog^{C27,28S}-mNG* adults raised at 18°C and 25°C (scale bar: 500 µm). The wing PCV (blue arrowhead) was classed as wild-type or ectopic PCV. Higher magnification views of *sog^{C27,28S}-mNG* wings with a mild ectopic PCV phenotype are shown adjacent to images of the whole wing (scale bars: 50µm). The black arrowheads indicate ectopic extension to the distal side of the PCV. (B, C) Percentage of wings in each phenotypic class at 25°C (B) and 18°C (C). Sample sizes for flies raised at 25°C are as follows: control n=56, *sog-mNG* n=66, *sog^{C27,28S}-mNG* n=67. Sample sizes for flies raised at 18°C: control n=64, *sog-mNG* n=65, *sog^{C27,28S}-mNG* n=67. Fisher's exact test (two-sided) finds a significant association between genotype and PCV phenotype in flies raised at 25°C ($P<0.01$) and at 18°C ($P<0.01$). Statistical tests were performed on raw data.

Sog has a curved shape

The data presented above demonstrate the utility of our *sog^{attP}* line for testing and elucidating the effect of specific *sog* mutations. One limitation for targeted mutagenesis of *sog*, however, is the absence of structural information. Therefore, to investigate Sog structure, we purified Sog (with C-terminal His and V5 tags) from the conditioned media of a stable, Sog-expressing, HEK293 EBNA cell line by affinity purification followed by two rounds of size exclusion chromatography (SEC) (Fig. 6A; Fig. S7). Sog purity was assessed by SDS-PAGE and western blot analysis where, after an initial round of SEC, a prominent doublet band was typically observed (Fig. S7C). This doublet band is likely to represent full-length Sog and Sog lacking the N terminal vWC1 domain due to Tld cleavage, as the cleavage product is detected by a His antibody (Fig. S7C) and expression of Chordin in the same cell line results in co-purification of a Tolloid cleavage product lacking the vWC1 domain (Troilo et al., 2014). A second round of SEC was included but could not completely separate the lower molecular weight species (Fig. 6A,B). Negative stain transmission electron microscopy (TEM) was used to investigate the three-dimensional (3D) structure of the purified Sog protein. During single particle analysis, two-dimensional (2D) classification aligned Sog particle images and produced 2D class averages that were used to generate and refine a Sog 3D model (Fig. 6C,D). The final 3D reconstruction, with an estimated resolution of 22.8 Å (Fig. 6D,E), reveals that Sog has an asymmetric, curved shape and dimensions of 13.6 nm×9.9 nm×9.2 nm (Fig. 6D). Due to the similarity in size, the Sog cleavage product cannot be separated from the full-length protein during image analysis, therefore the cleavage product will also have a contribution to the 3D model.

Comparing AlphaFold predictions to experimental data

Next, we probed the AlphaFold protein structure database, a recently developed resource based on the machine learning prediction of protein structures to atomic resolution (Jumper et al., 2021; Varadi et al., 2022), to analyse the predicted Sog structure and investigate how it fits within the Sog 3D reconstruction described above. For reference, the domain organisation of Sog is shown in Fig. 7A. Given the presence of the hydrophobic N-terminally located Sog transmembrane domain/signal peptide, it is likely that much of the Sog N-terminus is cleaved prior to secretion into the perivitelline space. The Sog N-terminus has therefore been removed from the AlphaFold model, at a site between R79 and H80 previously identified as a putative cleavage target for separation of a hydrophobic N-terminal signal sequence/ transmembrane domain from mature extracellular Sog (Shimmi and O'Connor, 2003). AlphaFold generates a pLDDT score, which is a useful metric from which to infer confidence in the local predicted protein structure. AlphaFold predicts the folds of each Sog vWC domain, with the majority of residues predicted with a high pLDDT score of 70–90 (Fig. 7B). Furthermore, AlphaFold predicts a novel fold for each of the four CHRD specific domains (Fig. 7A) with high confidence (pLDDT >90 and 70–90) (Fig. 7B). Residues linking vWC domains to other vWC or CHRD domains, however, are predicted with low pLDDT scores (50–70 and <50), indicating potentially disordered and/or flexible regions.

A predicted aligned error (PAE) score is also calculated for each residue pair by AlphaFold (Fig. 7C) (Jumper et al., 2021; Varadi et al., 2022). This score is a measure of the confidence with which the positions of amino acid pairs are predicted, thereby indicating the confidence of relative domain positions. The relative positions of each CHRD domain, vWC1 and the 4x CHRD region, and vWC2 and vWC3 are predicted with high confidence. Other inter-domain distances, for instance between vWC3 and vWC4, are predicted with lower confidence, suggesting some flexibility in the full-length protein. The Tld cleavage sites in both Sog and its ortholog Chordin
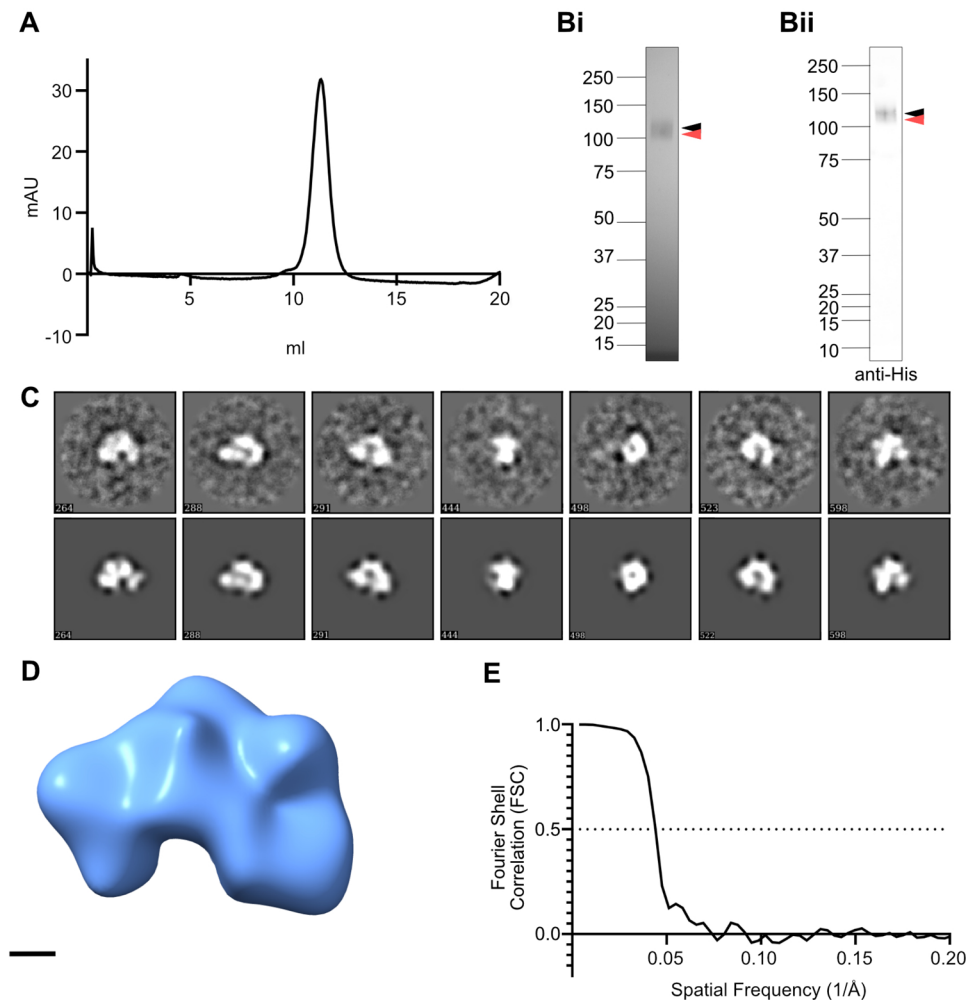
Fig. 6. Sog has a curved shape. (A) SEC trace for Sog purification. The Sog sample was subject to two rounds of SEC to minimise the contribution of a Tld cleavage product. The SEC trace for the second SEC round is presented here, see also Fig. S7 for the first SEC purification. (B) Reduced SDS-PAGE gel (i), and anti-His western blot (ii), corresponding to the SEC trace in A, showing purified Sog. Full length Sog is indicated by a black arrowhead. Full length Sog co-purifies with an Sog N-terminal cleavage product (red arrowhead). (C) A selection of 2D class averages (top panel) generated during the final refinement and corresponding re-projections (bottom panel) of the Sog model shown in D. Numbers in the bottom left of the boxes are reference numbers arbitrarily assigned to classes and projections by the analysis software. Box size: 36 nm. (D) Final 3D model of Sog. Scale bar: 2 nm. (E) Fourier shell correlation (FSC) calculations were used in Eman2.2 (Tang et al., 2007) to estimate the model resolution using a cut off of 0.5.

are located within these flexible interdomain regions (Fig. 7A; Fig. S8) (see Discussion).

The asymmetric nature of the AlphaFold model, with the bulky 4x CHRD region on one side (Fig. 7B), resembles the Sog 3D negative stain EM model (Fig. 6D). Indeed, overlay of the AlphaFold model with the Sog 3D EM volume, using ChimeraX (Pettersen et al., 2021), demonstrates how the 4× CHRD region might sit within the larger arm of the Sog EM model, as well as the arrangement of the other domains within the density (Fig. 7D). For further insight into the hydrodynamic properties of Sog, purified Sog protein was subject to sedimentation velocity analytical ultracentrifugation (AUC) (Fig. 7E). c(S) analysis calculated a sedimentation coefficient of 5.23 (S20,W), and a frictional ratio of 1.75, reflecting the relatively large size of Sog and indicating an elongated shape (Fig. 7F). To further test the level of agreement between the Sog AlphaFold structure and purified Sog, the sedimentation coefficient and frictional ratio of the N-terminally cleaved Sog AlphaFold model was predicted with US-SOMO (Brookes and Rocco, 2018; Rai et al., 2005; Brookes, et al., 2010a, b). A sedimentation coefficient of 5.14 (S20,W) and frictional ratio of 1.59 were predicted for the Sog AlphaFold model (Fig. 7F), lending further confidence to the accuracy of this predicted atomic structure. Together, similarities in shape between the AlphaFold Sog prediction and the Sog EM model, and between experimentally and theoretically derived hydrodynamic parameter values, suggest an overall domain organisation that provides a framework for future studies.

DISCUSSION

In this study, we have generated a *sog* KO line that allows simple reintegration of altered *sog* cDNAs. While the Sog 3D model that we have described will allow targeted mutagenesis in future studies, here we showed the utility of our *sog*[attP] KO stock by using it to investigate the effect of mutating two residues implicated in Sog palmitoylation. Palmitoylation at cysteines 27 and 28 has previously been suggested to play an important role in membrane targeting and Sog secretion (Kang and Bier, 2010). However, our data show that mutation of these two residues to serine *in vivo* resulted in viable flies, consistent with these residues not being essential for Sog function.

Our data reveal that the wild-type *sog-mNG* reintegration embryos accumulate less *Race* mRNAs and have a minor reduction in the number of amnioserosa cells compared to wild-type embryos. As *Race* expression and amnioserosa fate are dependent on peak signalling, we speculate that there is a subtle defect in BMP gradient formation arising from slightly reduced shuttling of BMP heterodimers to the dorsal midline in the *sog-mNG* embryos. Although no significant difference in the width of the pMad stripe in the reintegration embryos was observed, the reduced *Race* mRNA numbers suggest that there is a minor defect in pMad levels. Co-staining of wild-type and *sog-mNG* or *sog*[C27,28S]-*mNG* embryos with the pMad antibody, along with sophisticated quantitation of the staining intensities (Gavin-Smyth et al., 2013; Umulis et al., 2010) will allow this to be addressed in future studies.
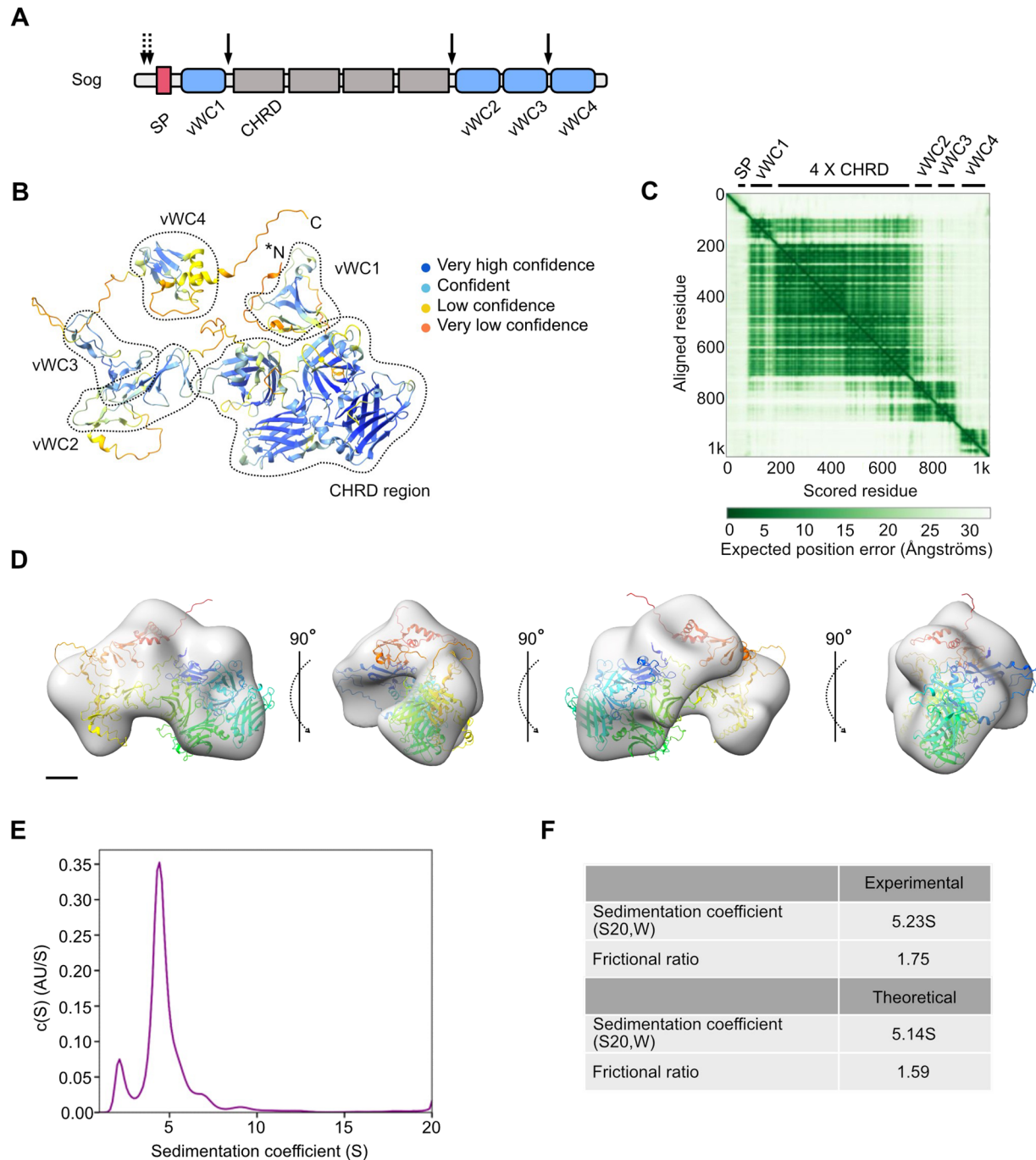
Fig. 7. Sog negative stain EM model is consistent with the AlphaFold prediction of Sog structure. (A) Sog comprises four vWC domains (blue) and four CHRD domains (grey). The N-terminal Sog hydrophobic domain is predicted to be a transmembrane domain (TMD) or signal peptide (red, SP). Solid arrows indicate Tld cleavage sites, dashed arrows indicate the location of the putative palmitoylation sites cysteines 27 and 28. (B) The prediction of the Sog structure by AlphaFold (Jumper et al., 2021; Varadi et al., 2022). N- and C-termini are indicated, the model shown here has been N-terminally truncated between R79 and H80 (N*). Sog vWC and CHRD domains are circled (figure adapted from AlphaFold entry Q24025). The predicted model is coloured according to per residue confidence (pLDDT) score. A score of 100 indicates maximum confidence. pLDDT scores are as follows: dark blue, >90; blue, 70–90; yellow, 50–70; orange, <50. (C) Predicted aligned error (PAE) (figure from AlphaFold entry Q24025). Domain boundaries are indicated by black bars along the 'scored residue' (X-) axis. (D) AlphaFold Sog model, N-terminally truncated after the hydrophobic domain at the predicted cleavage site, fits within the Sog EM model volume. Scale bar: 2 nm. (E) Sedimentation coefficient distribution produced by c(S) analysis of AUC data. After accounting for the sample buffer, the sedimentation coefficient is calculated to be 5.23 S (S20,W). (F) A comparison of experimental and theoretical sedimentation coefficients and frictional ratios. Theoretical values were calculated with US-SOMO software (Brookes et al., 2010a,b; Brookes and Rocco, 2018; Rai et al., 2005) for the predicted Sog structure by AlphaFold.

There are various potential explanations for the mildly reduced peak outputs in the *sog-mNG* reintegration embryos. Fusion of the C-terminal mNeonGreen tag to Sog may have slightly reduced its function. It is also possible that we have perturbed the timing of expression as we reintegrated the *sog* cDNA. As the cDNA is much shorter than the sequences present in the endogenous locus that

contains long introns, there is potentially more rapid accumulation of *sog* mRNAs, which may be particularly important in the short nc13 (Sandler et al., 2018). However, we did not detect any global effects on *sog* expression levels relative to wild-type embryos at the end of nc14, which may be expected if there is earlier accumulation of full length *sog* mRNAs. Introduction of the *sog* cDNA in the reintegration embryos also prevents expression of a truncated *sog* transcript that includes intron-derived sequence. The encoded short Sog protein has been shown to suppress early BMP signalling and prevent ectopic target gene expression during nc13 (Sandler et al., 2018). However, as we observe a reduction in *Race* mRNAs at nc14, this phenotype does not seem compatible with a loss of Short Sog expression. Reintegration also introduces additional sequences including an attR scar in the 5′UTR, which could affect translational regulation of the *sog* mRNA. In addition, the endogenous *sog* locus sequences (starting at the end of the CRISPR deletion) are present downstream of the reintegration sequences. While we find no evidence that the remnants of the endogenous locus can encode a secreted functional Sog protein, reintegration plasmid sequences can be removed in future studies by Cre recombination.

Using the *sog-mNG* flies as controls, we identified two minor phenotypes associated with *sog^{C27,28S}-mNG* embryos and wings. There is a reduction in both *Race* and *ush* expression in the *sog^{C27,28S}-mNG* embryos, suggesting that there are slightly lower levels of the pMad activator. We speculate that this is due to a further minor reduction in BMP shuttling by the mutant Sog, due to a slight decrease in either its levels or activity, resulting in a lower concentration of pMad at the dorsal midline. Although *sog^{C27,28S}-mNG* embryos have ~4-fold reduction in *Race* mRNA numbers relative to the *sog-mNG* embryos, there is no difference in the number of amnioserosa cells specified. As *Race* expression is exquisitely sensitive to reductions in peak Dpp/Scw signalling (Ashe and Levine, 1999; Rusch and Levine, 1997), it is possible that other peak BMP targets show a more modest reduction in mRNA numbers. Consistent with this, the intermediate target gene *ush* is only reduced ~2-fold in the *sog^{C27,28S}-mNG* embryos. We also observed a weakly penetrant ectopic PCV phenotype in *sog^{C27,28S}-mNG* wings, as a minor proportion have a small extension to the distal side of the PCV. This phenotype is consistent with reduced pMad, potentially due to lower activity of the *sog^{C27,28S}* mutant in promoting BMP transport to the PCV (Antson et al., 2022). While both the embryonic target gene and wing PCV phenotypes suggest slightly reduced pMad levels, overall these are very mild phenotypes and the flies with *sog^{C27,28S}* are viable.

While our data suggest that C27 and C28 in Sog are not critical, previously it was shown that overexpression of the dHIP14 palmitoyl transferase in the anterior of the early embryo or wing reduced pMad and disrupted wing vein patterning, respectively, similar to the phenotypes associated with *sog* overexpression (Kang and Bier, 2010). It is possible that dHIP14 overexpression has pleiotropic effects. Recent evidence suggests that the inhibitory Smad Daughters Against Dpp (Dad) is palmitoylated, which is important for its function (Li et al., 2017). While *dad* is not expressed in the early embryo, perhaps effects on Dad palmitoylation contribute to the wing phenotypes observed on dHIP14 overexpression.

As the low levels of Sog *in vivo* prevented us from directly measuring its level of palmitoylation for the wild-type and C27,28S mutant, we cannot rule out another palmitoylation target site in Sog. A cys residue within the predicted TMD/SP is another putative target (Kang and Bier, 2010). However, conditioned media collected from cells expressing the *sog^{C27,28S}* mutant was less able to inhibit Dpp signalling compared to wild-type Sog in a tissue

culture assay, consistent with reduced Sog secretion (Kang and Bier, 2010). This result suggests that the C27,28S mutation would be sufficient to reveal some defect in BMP signalling regulation *in vivo*, if Sog palmitoylation at these residues is important. Together, our data suggest that Sog secretion is much less dependent on C27, 28 and palmitoylation *in vivo* than in tissue culture cells.

Our 3D model of Sog constructed from negative stain EM data reveals a curved shape similar to the 'horseshoe' shape of its vertebrate ortholog Chordin. Due to the known BMP-Chordin vWC domain binding affinities, BMP dimers have been predicted to cooperatively bind Sog vWC1 and vWC4, and Chordin vWC1 and vWC3 domains (Larraín et al., 2000; Troilo et al., 2014; Zhang et al., 2007). In addition, Sog vWC1 and vWC4 have also been shown to interact with Col IV (Sawala et al., 2012). Sog and Chordin have therefore been predicted to adopt a curved conformation that would position the N- and C-termini in close proximity (Larraín et al., 2000; Sawala et al., 2012; Troilo et al., 2014). For Chordin, this 'horseshoe-like' shape is supported by a 3D reconstruction generated by single particle analysis of negative stain EM data (Troilo et al., 2014). The curved shape of the Sog EM and AlphaFold structures is therefore consistent with models of cooperative BMP binding, and simultaneous vWC1- and vWC4-Col IV interactions.

In comparison to Chordin, the model of Sog presented here shows Sog to be slightly more compact, with dimensions of 13.6×9.9×9.2 nm versus 15×13×8 nm of Chordin (Troilo et al., 2014). While Tld can only cleave Sog when it is bound to BMP, Chordin alone is processed by Tld (Marqués et al., 1997; Peluso et al., 2011; Piccolo et al., 1997). It has been suggested that a BMP induced conformational change in Sog is required for Sog cleavage by Tld (Marqués et al., 1997; Peluso et al., 2011). The PAE scores for AlphaFold Sog suggest that it is less flexible than Chordin, which together with the more compact shape of Sog could contribute to the requirement for this conformational change. The lower confidence in the relative positions of C-terminal vWC domains of the AlphaFold model is consistent with a level of Sog flexibility, potentially facilitating a ligand dependent conformational change, and permitting Tld access to target residues. The predicted greater flexibility of Chordin interdomain regions, where Tld cleavage sites are located, could therefore reflect the absence of required co-factors for Tld-mediated Chordin processing. The dependency of Sog on BMP ligand binding for Tld cleavage appears to underpin the 'shuttling' function of Sog during *Drosophila* embryogenesis (Peluso et al., 2011). In contrast, a 'source-sink' model of BMP gradient formation is most likely to operate during vertebrate embryo dorsal-ventral patterning (Pomreinke et al., 2017; Tuazon et al., 2020; Zinski et al., 2017). Future studies probing how the structures of Sog and Chordin differ will illuminate how these proteins use the same interacting proteins but different mechanisms to generate a BMP gradient.

MATERIALS AND METHODS
Sog purification
Lipofectamine 3000 reagent (Thermo Fisher Scientific) and Xfect (Takara Bio) transfection reagents were used to transfect HEK293 EBNA cells (Baldock lab stock, not recently authenticated) with pCep-Pu-Sog-V5-His to establish a stable cell line. pCep-Pu-Sog-V5-His cells were maintained in growth media (10% FBS, DMEM:F-12 Hams, 1% penicillin-streptomycin (P/S), 1% Glu) containing 1 µg/ml puromycin. For protein expression, cells were cultured in HYPERflasks (Corning) with expression media (DMEM: F-12 Hams, 1% P/S, 1% Glu, 50 mM L-Arginine). Collected conditioned media was stored at −20°C.

Recombinant Sog protein was isolated from collected conditioned media using affinity chromatography. 1 ml HisTrap Excel columns (Cytiva) were used to pull down Sog via the C-terminal His-tag. For this, 2 mM imidazole was added to filtered conditioned media, and a loading buffer was used (10 mM Tris, 800 mM NaCl, pH7.4) to equilibrate the His-trap column. Filtered conditioned media was pumped over the column at 4°C. The column was then washed with 20 column volumes (CV) loading buffer with 10 mM imidazole. Protein was eluted from the column with elution buffer (10 mM Tris, 800 mM NaCl, 500 mM imidazole, pH7.4).

Affinity chromatography fractions were subject to SEC for further purification with a Superdex 200 Increase 10/30 gel filtration column (Cytiva). The column was washed with 1.5 CV filtered and degassed Milli-Q water (Millipore), followed by 1.5 CVs SEC buffer (20 mM Tris, 800 mM NaCl, pH 7.4) to equilibrate. Affinity chromatography fractions were passed across the column at 0.5 ml/min. Protein elution was monitored using UV absorbance (280 nm). Eluted 0.5 ml fractions were collected and screened by SDS-PAGE under reducing conditions with Instant Bluestain (Abcam), and by western blot with anti-His [clone AD1.1.10] (1:1000, R&D Systems, cat. #MAB050, RRID: AB_357353) primary antibody and IRDye 800CW Donkey anti-Mouse (1:10,000, LI-COR Biosciences, cat. #926-32212, RRID: AB_621847Li-COR) secondary antibody. Purified Sog was stored at −80°C, before undergoing a second round of SEC (as above) to improve sample purity for production of negative stain EM grids. Collected fractions were screened by SDS-PAGE under reducing conditions with Colloidal Coomassie stain, and by western blot (as above). The Sog sample used for AUC underwent one round of SEC with the following SEC buffer: 800 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4.

Negative stain electron microscopy

A 5 nm (approximately) layer of carbon was deposited onto a mica sheet using a Cressington coating system 308R. The carbon layer was floated onto the carbon side of the TEM grids which were left to dry overnight at room temperature. Carbon coated grids were glow discharged at 25 mA for 1 min. After SEC, Sog (∼13 µg/ml) was adhered to carbon coated EM grids for 1 min and stained with 2% uranyl acetate. The stain was wicked away with filter paper before drying. Negatively stained protein molecules were imaged in low dose mode on an FEI Tecnai 12 Biotwin Transmission Electron Microscope at 100 kV. Images were captured with a Gatan Orius SC1000 camera at 30,000×magnification. Data collection parameters were as follows: Defocus=0.5–1 µm; Exposure length=0.6 s, pixel size=2.8 Å.

Data collected from negative-stain EM grids were processed using Eman2.2 software (Tang et al., 2007). Particles were manually picked and processed with a box size of 128 pixels. Contrast transfer function (CTF) parameters were estimated in Eman2.2 and particle images underwent phase flipping to correct the CTF. A soft Gaussian mask, with an outer radius of 52 nm was applied to selected particles. 2D class averaging was performed to iteratively align and average the particles into 100 reference free class averages. Class averages were discarded based on particle contrast against background noise. Of the remaining classes, 17 were selected that represented different views of Sog to generate an initial model. The initial model underwent two iterative refinements, using a 'gold standard' refinement procedure to produce a final model. The Fourier shell correlation (FSC) of two separately refined half-maps was used to estimate the resolution of the final model, with the 0.5 threshold.

AUC

A purified Sog sample (0.1 mg/ml) in the same buffer as used in SEC was characterized by sedimentation velocity AUC using a Beckman XL-A analytical ultracentrifuge with an An60Ti 4-hole rotor running at 45,000 rpm at 20°C. The sedimenting boundary was monitored at 230 nm for 300 scans. Data were analysed by continuous model-based distribution C(s) of Lamm equation solutions method using SEDFIT software (Schuck, 2000), and the resulting sedimentation coefficients were corrected to standard conditions using SEDNTERP software (Philo, 1997). For theoretical predictions of hydrodynamic properties for the predicted Sog structure by AlphaFold, US-SOMO software (Rai et al., 2005; Brookes, et al., 2010a,b; Brookes and Rocco, 2018) was used.

Fly stocks

All stocks were grown and maintained on standard fly food media (yeast 50 g/l, glucose 78 g/l, maize 72 g/l, agar 8 g/l, 10% nipagen in EtOH 27 ml/l and proprionic acid 3 ml/l). The following fly lines were used in this study: $y^1 w^{67c23}$ (BDSC Stock 6599), w^{1118}; $PBac\{y[+mDint2] GFP[E.3xP3]=vas-Cas9\}VK00027$ (BDSC Stock 51324); $y^1 w^{67c23}$; $MKRS, P\{ry+t7.2=hsFLP\}86E/TM6B, P\{w+mC=Crew\}DH2, Tb1$ (BDSC Stock 1501); $y^1 sog^{S6}/FM7c, sn^+$ (BDSC Stock 2497); $brk^{M68}/FM7c$-ftz-lacZ (Jazwinska et al., 1999). $y^1 w^{67c23}$ flies were used as controls throughout.

CRISPR-Cas9 genome editing and phiC31 reintegration

The sog^{attP} CRISPR Drosophila line was generated by CRISPR-Cas9 with HDR in a two-step CRISPR approach (Baena-Lopez et al., 2013; Hoppe and Ashe, 2021). For a detailed outline of the strategy see (Hoppe and Ashe, 2021). PAM sites (NGG) located either side of the sog start codon were identified using the CRISPR OptimalTarget Finder tool on the flyCRISPR website (Gratz et al., 2014). Two guide RNA sequences were designed 3 nucleotide upstream of the selected PAM sites to target these sites for Cas9 nuclease digestion and the creation of double stranded breaks. For oligonucleotide sequences encoding sense and antisense strands for guide sequence see Table S1. Complementary guide oligonucleotides were annealed and inserted into the pU6-Bbs1-gRNA plasmid (RRID: Addgene_45946; Gratz et al., 2013) as previously described (Hoppe and Ashe, 2021). Homology arm (HA) sequences were obtained from Drosophila genomic DNA (BL51324) by PCR. Homology arms were inserted into the pTV^{Cherry} donor plasmid (Drosophila Genomics Resource Center, DGRC_1338) (Baena-Lopez et al., 2013) at SpeI and KpnI restriction sites. Donor plasmids and gRNA plasmids were injected into Cas9 expressing embryos (BL51324) at the University of Cambridge Fly Injection Facility. Flies that developed from injected embryos were crossed according to (Hoppe and Ashe, 2021), using $y^1 w^{67c23}$ and brk^{M68}/FM7c-ftz-lacZ (Jazwinska et al., 1999) stocks. The white+ marker was removed by crossing sog^{attP} females with males that carried FM7c on the X and Cre-recombinase on the third chromosome. The sog^{attP} CRISPR mutation results in a deletion of 800 bp (ChrX:15,625,466, - 15,626,265, dm6), including the endogenous sog start codon and signal sequence, that is replaced by 103 nucleotides containing attP and LoxP sites.

Reintegration plasmids were generated from the RIV^{White} plasmid (gift from the Vincent lab). A partial sog 5′UTR (p5′UTR) sequence (source: pBS-Sog-CDS; Ashe and Levine, 1999), followed by the sog CDS, mNeonGreen and the sog 3′UTR (source: pBS-Sog-CDS; Ashe and Levine, 1999), was inserted between the RIV^{White} attB and LoxP sites. To summarise the construction of these plasmids, the partial 5′UTR and the sog CDS were ligated together in a pAc5.1/V5-His (Thermo Fisher Scientific, V411020) vector, as were the mNeonGreen and sog 3′UTR sequences, using In-fusion cloning (Takara Biosciences). A linker sequence was added downstream of the sog CDS. The p5′UTR-sog CDS-linker, and mNeonGreen-3′UTR sequences were inserted into RIV^{White} using In-Fusion cloning. To create the $sog^{C27,28S}$ mutant, Cys 27 and 28 were replaced by two Ser residues with In-Fusion cloning. Reintegration vectors were injected with a phiC31 encoding plasmid into embryos of the sog^{attP} CRISPR stock. Female flies that developed from the injected embryos were crossed to $y^1 w^{67c23}$ males and w+ offspring were crossed to brk^{m68}/FM7c ftz-lacZ to balance. Successful phiC31 mediated recombination was confirmed by sequencing genomic DNA. Flies in which successful phiC31 recombination events had occurred were backcrossed to each other to make homozygous stocks. The w+ marker gene was not removed from the sog-mNG and $sog^{C27,28S}$-mNG stocks generated here. This could be done by Cre-recombinase if necessary for future work. See Table S1 for primers and oligonucleotide sequences. SnapGene viewer software (from Insightful Science; available at snapgene.com) was used to scan DNA sequences for ORFs.

Viability and lethality assays

For viability assays, 50 embryos were placed on an apple juice agar plate which was transferred into a food bottle. Embryos were incubated at 25°C in the bottle. The number of pupae and adults were counted. For the sog recue assay, virgins of $y^1 sogS6$/FM7c, sn+ (BDSC Stock 2497) or sog^{attP}/FM7c

were crossed to male *yw*, *sog-mNG;;* or *sog*[C27,28S]*;;* flies. The number of FM7c and non-FM7c female offspring were counted to assess the degree of rescue from each genotype in the presence of the mutant *sog* allele.

Wing dissection

Adult flies were incubated at 18°C or 25°C for 24 h, before transfer to a fresh vial. Flies were then allowed to lay eggs before being discarded. Embryos were allowed to develop to adulthood at the designated temperature condition. Wings were then removed from adult females, placed on a slide and washed briefly in isopropanol. Wings were mounted in DPX mounting media (Fisher D/5319/05) under a No.1 coverslip. Samples were imaged with a light microscope (Zeiss Axioskop) with a 5X objective (Zeiss CP-Achromat 5X/0.12). Images were acquired with Agilent OpenLab 2.2.2 software. For analysis, statistical tests were performed on raw count data in RStudio.

In situ hybridisation and immunofluorescence

Embryos (2–4 h) were collected and stained by RNA *in situ* hybridisation with *sog*-digoxygenin-UTP, *Race*-Biotin-UTP, *lacZ*-digoxygenin-UTP or mNeonGreen-biotin-UTP probes as described (Hoppe et al., 2020; Kosman et al., 2004). An mNeonGreen-biotin-UTP probe was synthesised as previously described (Kosman et al., 2004) with primers listed in Table S1. Antibodies used were mouse anti-biotin (1:250, Roche, cat. #1297597), sheep anti-digoxigenin Fab fragments antibody, AP conjugated (1:200, Roche, cat. #11093274910 RRID:AB514497), donkey anti-mouse IgG secondary antibody, Alexa Fluor 647 (1:500, Thermo Fisher Scientific, cat. #A-31571, RRID:AB162542), and donkey anti-sheep IgG secondary antibody, Alex Fluor 488 (1:500, Thermo Fisher Scientific, cat. #A-11015, RRID: AB_2534082). For pMad immunostaining, anti-Smad3 (phospho S423+S425) [EP823Y] (1:500, Abcam, cat. #ab52903, RRID: AB_882596) primary antibody and Donkey anti-rabbit IgG secondary antibody, Alexa Fluor 647 (1:500, Thermo Fisher Scientific, cat. #A-31573, RRID: AB_2536183) were used. To stain embryo nuclei, samples were incubated with DAPI (1:1000, NEB 4083). Samples were mounted in ProLong™ Diamond Antifade Mountant (Thermo Fisher Scientific, P36961).

smFISH (Stellaris), smiFISH and immunofluorescence

For smFISH, 2–4 h or 1–3 h embryos were processed as described (Hoppe et al., 2020) with anti-*sog* Stellaris (Fig. 2C, Fig. S1), or *sog* single molecule inexpensive FISH (smiFISH) (Fig. S4) (Tsanov et al., 2016), *ush* Stellaris, *lacZ* Stellaris, and *Race* smiFISH probes (Tsanov et al., 2016). *ush* probe sequences are available from (Hoppe et al., 2020), while *sog* Stellaris, *sog* smiFISH, *lacZ* Stellaris, and *Race* smiFISH probe sequences are provided in Table S2. *Race* and *sog* smiFISH probes were annealed to a 570-conjugated Y-FLAP and Z-FLAP, respectively (Tsanov et al., 2016). For immunostaining against mNeonGreen and Spectrin, mouse anti-mNeonGreen [32F6] (1:500, ChromoTek, cat. #32f6-100, RRID: AB_2827566), mouse anti-Spectrin (1:50, DSHB, cat. #3A9 (323 or M10-2), RRID: AB_528473) and donkey anti-mouse IgG secondary antibody, Alexa Fluor 488 (1:1000, Thermo Fisher Scientific, cat. #A-21202, RRID: AB_141607) primary and secondary antibodies were used. To stain nuclei, samples were incubated with DAPI (1:1000, NEB 4083). Samples were mounted in ProLong™ Diamond Antifade Mountant (Thermo Fisher Scientific, P36961).

Imaging stained embryos

Fixed embryos stained with *Race* and *lacZ* RNA *in situ* hybridisation probes were imaged with a Leica TCS SP5 AOBS inverted microscope using a HCX PL APO lambda blue 20.0×0.70 IMM UV oil objective. The following confocal settings were used: pinhole 1 airy unit, scan speed 400 Hz unidirectional, 512×512 pixel format, Z step size of 1.5 µm at 8 bit and 1.3× zoom. Images shown are maximum intensity projections. Images were deconvolved with Huygens Professional software (SVI, Scientific Volume Imaging, RRID:SCR 014237).

sog Stellaris early nc14 embryos, co-stained with *lacZ* stellaris and anti-Spectrin antibody, were imaged in a Leica TCS SP8 AOBS confocal microscope with a HC PL APO CS2 40×/1.30 oil objective with 0.75× zoom to screen for embryos that did not stain with *lacZ* smFISH. The settings used were as follows, pinhole 1 airy unit, scan speed 400 Hz bidirectional, format 2048×512 pixels, at 8 bit. Images were collected using hybrid detectors

using the white light laser with 647 nm (20%), 548 nm (20%), 405 nm (6%). To age embryos using the anti-Spectrin antibody stain, at the centre of the embryo stacks of 5–10 images (Z step size=0.3 µm) were collected with hybrid detectors using the white light laser with 548 nm (20%), 488 nm (12%) 405 nm (6%). To collect images for analysis embryos were imaged with a HCX PL Apo 63×/1.40 oil objective and 0.75×zoom. The confocal settings used were as follows, pinhole 1 airy unit, scan speed 400 Hz bidirectional, format 2048×512 pixels, at 8 bit and Z step size 0.3 µm. Images were collected using hybrid detectors using the white light laser with 548 nm (20%), 405 nm (6%) with 3X line accumulation. Raw images were deconvolved with Huygens Remote Manager software v3.7.1 (SVI). Images shown are maximum intensity projections.

sog smiFISH stage 6 embryos were imaged on a Leica TCS SP8 AOBS confocal microscope with a HCX PL Apo 63×/1.40 oil objective with 1×zoom. The confocal settings used were as follows, pinhole 1 airy unit, scan speed 400 Hz bidirectional, format 2048×2048 pixels, at 8 bit and Z step size 0.3 µm. Images were collected using hybrid detectors using the white light laser with 548 nm (20%), 405 nm (9.8%) with 3×line accumulation. Images shown are maximum intensity projections. Raw images were deconvolved with Huygens Remote Manager software v3.7.1 (SVI). Samples subject to FISH with *sog*-digoxygenin and mNeonGreen-biotin RNA probes were imaged on a Leica TCS SP8 AOBS confocal microscope with a HCX PL Apo 63×/1.40 oil objective and 0.75× zoom. The confocal settings used were as follows, pinhole 1 airy unit, scan speed 400 Hz unidirectional, format 2048×700 pixels, at 8 bit, and Z step size 0.35 µm. Images were collected using hybrid detectors using the white light laser with 647 nm (14%), 488 nm (14%), 405 nm (14%), and 4X line averaging. Cropped squares from the 2048×700 images are presented and the images shown are single Z slices.

sog smiFISH and anti-mNeonGreen immunofluorescence samples were imaged on a Leica TCS SP8 AOBS confocal microscope with a HC PL APO CS2 40×/1.30 oil objective and 0.75× zoom. The confocal settings used were as follows, pinhole 1 airy unit, scan speed 400 Hz, format 2048×2048 pixels, at 8 bit and 0.35 µm Z step size. Images were collected using hybrid detectors using the white light laser with 647 nm (22%), 548 nm (23%), 488 nm (20%), 405 nm (14%), and 6X line averaging. Images were deconvolved with Huygens Professional software. Images shown are maximum intensity projections (~3.5 µm depth).

Embryos stained for pMad were imaged on a Leica TCS SP8 AOBS confocal microscope with a HCPL APO CS2 40×/1.30 oil objective and 0.75× zoom. The confocal settings were as follows; pinhole of 1 airy unit, unidirectional scanning at 400 Hz, 1024×1024 pixel format at 8 bit and Z step size of 0.35 µm. Images for these samples were collected with hybrid detectors using a white light laser with 488 nm (20%), 647 nm (27%), 405 nm (12%) and 4X line averaging. Raw images were deconvolved with Huygens Remote Manager software v3.7.1 (SVI).

ush smFISH and *Race* smiFISH samples were imaged on a Leica TCS SP8 AOBS confocal microscope with a HCX PL Apo 63×/1.40 oil objective and 1X zoom. The confocal settings used were as follows, pinhole 1 airy unit, scan speed 400 Hz bidirectional, format 3144×3144 pixels, at 8 bit and Z step size 0.3 µm. Images were collected using hybrid detectors using the white light laser with 647 nm (22%), 548 nm (22%), 405 nm (6.1%) with 3X line accumulation. Raw images were deconvolved with Huygens Remote Manager software v3.7.1 (SVI).

smFISH/smiFISH image analysis

To age *sog*[attP] and *y*[1] *w*[67c23] control embryos stained with anti-*sog* and anti-*lacZ* smFISH and anti-Spectrin antibody, maximum intensity projections of images were made, and the length of the cell membrane ingression (Spectrin antibody stain) was measured (Calvo et al., 2021). Embryos with cell membranes of 3.5–5.5 µm were used for analysis. Only male embryos were analysed. *y*[1] *w*[67c23] males were identified by the number of *sog* transcription sites, and *sog*[attP] males were identified by the absence of *lacZ* smFISH stain.

Quantitative analysis of *ush* and *Race*, and *sog* smFISH/smiFISH images was performed in Imaris 9.2 (Bitplane, Oxford Instruments). For efficiency, analysis of *ush*/*Race* images was performed for only the central 1048×3144 area of the 3144×3144 images. For *sog* smFISH data, the entire 2048×512

images were analysed. For quantification of *ush*, *Race*, or *sog* mRNA number, individual mRNAs were detected with the 'spots' function. Spots of diameter 0.3 µm (X/Y direction) and 0.8 µm (Z direction) were used. Nuclear locations were determined using the 'surfaces' function to identify nuclei based on DAPI staining. To determine the number of spots per cell, spots were assigned to surfaces using the spotMe_V2.py python script (Vinter et al., 2021) (script available at https://github.com/TMinchington/sass). Output from this analysis was processed in Rstudio to remove duplicated nuclei values and 'NAs'. Data for the number of mRNAs for each of the three embryos analysed were pooled and divided into 5 µm bins (approximate cell size) to permit calculation of the mean number of mRNAs at a given distance from the dorsal/expression domain midline. To account for imaging of embryos that were not perfectly lateral, data embryos stained with anti-*sog* smFISH, were cropped −60 and +80 µm from the expression domain midline, as these were the boundaries shared by all embryos. Binned data were plotted in Rstudio, while data for individual embryos were plotted in GraphPad Prism 9 (RRID:SCR 002798).

For analysis of *sog* expression in *sog-mNG* and *sogC27,28S-mNG* and control embryos, *sog* smiFISH images were analysed in Fiji ImageJ (Schindelin et al., 2012). To quantify the height of the *sog* expression domain, maximum intensity projections of the imaging stacks were made, and the maximum vertical number of *sog* positive cells was counted in a 25 µm region of interest (ROI) located 50 µm to the posterior of the cephalic furrow. To quantify *sog* expression level within *sog* positive cells in sum of slices projected images, a 40×40 µm ROI was drawn 50 µm away from the cephalic furrow and bordering the ventral edge of the *sog* expression domain. The mean grey value of the ROI was calculated. To correct for background fluorescence, the mean grey value along a 20 µm line situated outside the sog expression domain was calculated. This value was subtracted from the mean grey value within the ROI. Statistical analysis was performed in GraphPad Prism 9 (RRID:SCR 002798).

pMad stain analysis

To quantify pMad distribution in stage 6 embryos, the width of the pMad stripe was measured as the number of pMad positive nuclei in maximum intensity projected images. Analysis was performed in Fiji ImageJ (Schindelin et al., 2012). The data were plotted and the statistical test was performed in GraphPad Prism (RRID:SCR 002798).

Amnioserosa counts

Fixed embryos were stained with mouse anti-Hnt 1G9 (1:40, DSHB Cat# 1g9, RRID: AB_528278) and Anti-Mouse IgG (H+L) AP Conjugate S372B 1:500 (Promega) by standard techniques (Kosman et al., 2004). Stage 11 embryos were imaged on a Leica DM600B microscope with a 20x objective using brightfield. Total amnioserosa cells were counted on ImageJ using the Cell Counter plugin. 50 embryos across three biological repeats were analysed per genotype. Counts were plotted in GraphPad Prism 9 (RRID:SCR 002798).

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: S.L.F., C.B., H.L.A.; Investigation: S.L.F., C.S.; Writing - original draft: S.L.F., H.L.A.; Writing - review & editing: S.L.F., C.S., C.B., H.L.A.t

References

Antson, H., Tõnissoo, T. and Shimmi, O. (2022). The developing wing crossvein of Drosophila melanogaster: a fascinating model for signaling and morphogenesis. *Fly (Austin)* **16**, 118-127. doi:10.1080/19336934.2022.2040316

Arora, K., Levine, M. S. and O'Connor, M. B. (1994). The screw gene encodes a ubiquitously expressed member of the TGF-beta family required for specification of dorsal cell fates in the Drosophila embryo. *Genes Dev.* **8**, 2588-2601. doi:10.1101/gad.8.21.2588

Ashe, H. L. and Levine, M. (1999). Local inhibition and long-range enhancement of Dpp signal transduction by Sog. *Nature* **398**, 427-431. doi:10.1038/18892

Ashe, H. L., Mannervik, M. and Levine, M. (2000). Dpp signaling thresholds in the dorsal ectoderm of the Drosophila embryo. *Development* **127**, 3305. doi:10.1242/dev.127.15.3305

Baena-Lopez, L. A., Alexandre, C., Mitchell, A., Pasakarnis, L. and Vincent, J.-P. (2013). Accelerated homologous recombination and subsequent genome modification in Drosophila. *Development* **140**, 4818. doi:10.1242/dev.100933

Bandyopadhyay, A., Yadav, P. S. and Prashar, P. (2013). BMP signaling in development and diseases: a pharmacological perspective. *Biochem. Pharmacol.* **85**, 857-864. doi:10.1016/j.bcp.2013.01.004

Bannan, B. A., Van Etten, J., Kohler, J. A., Tsoi, Y., Hansen, N. M., Sigmon, S., Fowler, E., Buff, H., Williams, T. S., Ault, J. G. et al. (2008). The drosophila protein palmitoylome characterizing palmitoyl-thioesterases and DHHC palmitoyl-transferases. *Fly* **2**, 198-214. doi:10.4161/fly.6621

Brookes, E. and Rocco, M. (2018). Recent advances in the UltraScan SOlution MOdeller (US-SOMO) hydrodynamic and small-angle scattering data analysis and simulation suite. *Eur. Biophys. J.* **47**, 855-864. doi:10.1007/s00249-018-1296-0

Brookes, E., Demeler, B. and Rocco, M. (2010a). Developments in the US-SOMO bead modeling suite: new features in the direct residue-to-bead method, improved grid routines, and influence of accessible surface area screening. *Macromol. Biosci.* **10**, 746-753. doi:10.1002/mabi.200900474

Brookes, E., Demeler, B., Rosano, C. and Rocco, M. (2010b). The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. *European Biophysics Journal : EBJ* **39**, 423-435. doi:10.1007/s00249-009-0418-0

Calvo, L., Ronshaugen, M. and Pettini, T. (2021). smiFISH and embryo segmentation for single-cell multi-gene RNA quantification in arthropods. *Commun. Biol.* **4**, 1-12. doi:10.1038/s42003-021-01803-0

Dorfman, R. and Shilo, B. Z. (2001). Biphasic activation of the BMP pathway patterns the Drosophila embryonic dorsal region. *Development* **128**, 965-972. doi:10.1242/dev.128.6.965

Eldar, A., Dorfman, R., Weiss, D., Ashe, H., Shilo, B.-Z. and Barkai, N. (2002). Robustness of the BMP morphogen gradient in Drosophila embryonic patterning. *Nature* **419**, 304-308. doi:10.1038/nature01061

Ferguson, E. L. and Anderson, K. V. (1992). Decapentaplegic acts as a morphogen to organize dorsal-ventral pattern in the Drosophila embryo. *Cell* **71**, 451-461. doi:10.1016/0092-8674(92)90514-D

Francois, V., Solloway, M., O'Neill, J. W., Emery, J. and Bier, E. (1994). Dorsal-ventral patterning of the Drosophila embryo depends on a putative negative growth factor encoded by the short gastrulation gene. *Genes Dev.* **8**, 2602-2616. doi:10.1101/gad.8.21.2602

Garcia Abreu, J., Coffinier, C., Larraín, J., Oelgeschläger, M. and De Robertis, E. M. (2002). Chordin-like CR domains and the regulation of evolutionarily conserved extracellular signaling systems. *Gene* **287**, 39-47. doi:10.1016/S0378-1119(01)00827-7

Gavin-Smyth, J., Wang, Y. C., Butler, I. and Ferguson, E. L. (2013). A genetic network conferring canalization to a bistable patterning system in Drosophila. *Curr. Biol.* **23**, 2296-2302. doi:10.1016/j.cub.2013.09.055

Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E. et al. (2003). A Protein Interaction Map of Drosophila melanogaster. *Science* **302**, 1727-1736. doi:10.1126/science.1090289

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J. and O'Connor-Giles, K. M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194**, 1029. doi:10.1534/genetics.113.152710

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M. and Oconnor-Giles, K. M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics* **196**, 961-971. doi:10.1534/genetics.113.160713

Holley, S. A., Neul, J. L., Attisano, L., Wrana, J. L., Sasai, Y., Oconnor, M. B., DeRobertis, E. M. and Ferguson, E. L. (1996). The Xenopus dorsalizing factor noggin ventralizes Drosophila embryos by preventing DPP from activating its receptor. *Cell* **86**, 607-617. doi:10.1016/S0092-8674(00)80134-8

Hoppe, C. and Ashe, H. L. (2021). CRISPR-Cas9 strategies to insert MS2 stem-loops into endogenous loci in Drosophila embryos. *STAR Protocols* **2**, 100380. doi:10.1016/j.xpro.2021.100380

Hoppe, C., Bowles, J. R., Minchington, T. G., Sutcliffe, C., Upadhyai, P., Rattray, M. and Ashe, H. L. (2020). Modulation of the promoter activation rate dictates the transcriptional response to graded BMP signaling levels in the Drosophila embryo. *Dev. Cell* **54**, 727-741.e7. doi:10.1016/j.devcel.2020.07.007

Jazwinska, A., Rushlow, C. and Roth, S. (1999). The role of brinker in mediating the graded response to Dpp in early Drosophila embryos. *Development* **126**, 3323-3334. doi:10.1242/dev.126.15.3323

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589. doi:10.1038/s41586-021-03819-2

Käll, L., Krogh, A. and Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027-1036. doi:10.1016/j.jmb.2004.03.016

Käll, L., Krogh, A. and Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35** Suppl. 2, W429-W432. doi:10.1093/nar/gkm256

Kang, K.-H. and Bier, E. (2010). dHIP14-dependent palmitoylation promotes secretion of the BMP antagonist Sog. *Dev. Biol.* **346**, 1-10. doi:10.1016/j.ydbio.2010.06.024

Kosman, D., Mizutani, C. M., Lemons, D., Cox, W. G., McGinnis, W. and Bier, E. (2004). Multiplex detection of RNA expression in Drosophila embryos. *Science* **305**, 846. doi:10.1126/science.1099247

Larraín, J., Bachiller, D., Lu, B., Agius, E., Piccolo, S. and De Robertis, E. M. (2000). BMP-binding modules in chordin: a model for signalling regulation in the extracellular space. *Development (Cambridge, England)* **127**, 821-830. doi:10.1242/dev.127.4.821

Li, W., Li, W., Zou, L., Ji, S., Li, C., Liu, K., Zhang, G., Sun, Q., Xiao, F. and Chen, D. (2017). Membrane targeting of inhibitory Smads through palmitoylation controls TGF-β/BMP signaling. *Proc. Natl Acad. Sci. USA* **114**, 13206-13211. doi:10.1073/pnas.1710540114

Linder, M. E. and Deschenes, R. J. (2003). New insights into the mechanisms of protein palmitoylation. *Biochemistry* **42**, 4317-4320.

Madamanchi, A., Mullins, M. C. and Umulis, D. M. (2021). Diversity and robustness of bone morphogenetic protein pattern formation. *Development (Camb.)* **148**, dev192344. doi:10.1242/dev.192344

Markstein, M., Markstein, P., Markstein, V. and Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. *Proc. Natl Acad. Sci. USA* **99**, 763. doi:10.1073/pnas.012591199

Marqués, G., Musacchio, M., Shimell, M. J., Wünnenberg-Stapleton, K., Cho, K. W. and O'Connor, M. B. (1997). Production of a DPP activity gradient in the early Drosophila embryo through the opposing actions of the SOG and TLD proteins. *Cell* **91**, 417-426. doi:10.1016/S0092-8674(00)80425-0

Mason, E. D., Konrad, K. D., Webb, C. D. and Marsh, J. L. (1994). Dorsal midline fate in Drosophila embryos requires twisted gastrulation, a gene encoding a secreted protein related to human connective tissue growth factor. *Genes Dev.* **8**, 1489-1501. doi:10.1101/gad.8.13.1489

Matsuda, S., Harmansa, S. and Affolter, M. (2016). BMP morphogen gradients in flies. *Cytokine Growth Factor Rev.* **27**, 119-127. doi:10.1016/j.cytogfr.2015.11.003

Montanari, M. P., Tran, N. V. and Shimmi, O. (2022). Regulation of spatial distribution of BMP ligands for pattern formation. *Dev. Dyn.* **251**, 178-192. doi:10.1002/dvdy.397

O'Leary, J. M., Hamilton, J. U., Deane, C. M., Valeyev, N. V., Sandell, L. J. and Downing, A. K. (2004). Solution structure and dynamics of a prototypical chordin-like cysteine-rich repeat (von Willebrand factor type C module) from collagen IIA. *J. Biol. Chem.* **279**, 53857-53866. doi:10.1074/jbc.M409225200

Peluso, C. E., Umulis, D., Kim, Y.-J., O'Connor, M. B. and Serpe, M. (2011). Shaping BMP morphogen gradients through enzyme-substrate interactions. *Dev. Cell* **21**, 375-383. doi:10.1016/j.devcel.2011.06.025

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. and Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70-82. doi:10.1002/pro.3943

Piccolo, S., Agius, E., Lu, B., Goodman, S., Dale, L. and DeRobertis, E. M. (1997). Cleavage of chordin by Xolloid metalloprotease suggests a role for proteolytic processing in the regulation of Spemann organizer activity. *Cell* **91**, 407-416.

Philo, J. S. (1997). An improved function for fitting sedimentation velocity data for low-molecular-weight solutes. *Biophys. J.* **72**, 435-444. doi:10.1016/S0006-3495(97)78684-3

Pomreinke, A. P., Soh, G. H., Rogers, K. W., Bergmann, J. K., Bläßle, A. J. and Müller, P. (2017). Dynamics of BMP signaling and distribution during zebrafish dorsal-ventral patterning. *ELife*, **6**, e25861. doi:10.7554/eLife.25861

Raftery, L. A. and Sutherland, D. J. (2003). Gradients and thresholds: BMP response gradients unveiled in Drosophila embryos. *Trends Genet.* **19**, 701-708. doi:10.1016/j.tig.2003.10.009

Rai, N., Nöllmann, M., Spotorno, B., Tassara, G., Byron, O. and Rocco, M. (2005). SOMO (SOlution MOdeler): differences between X-Ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics. *Structure* **13**, 723-734. doi:10.1016/j.str.2005.02.012

Ralston, A. and Blair, S. S. (2005). Long-range Dpp signaling is regulated to restrict BMP signaling to a crossvein competent zone. *Dev. Biol.* **280**, 187-200. doi:10.1016/j.ydbio.2005.01.018

Ross, J. J., Shimmi, O., Vilmos, P., Petryk, A., Kim, H., Gaudenz, K., Hermanson, S., Ekker, S. C., O'Connor, M. B. and Marsh, J. L. (2001). Twisted gastrulation is a conserved extracellular BMP antagonist. *Nature* **410**, 479-483. doi:10.1038/35068578

Rusch, J. and Levine, M. (1997). Regulation of a dpp target gene in the Drosophila embryo. *Development* **124**, 303-311. doi:10.1242/dev.124.2.303

Rushlow, C., Colosimo, P. F., Lin, M. C., Xu, M. and Kirov, N. (2001). Transcriptional regulation of the Drosophila gene zen by competing Smad and Brinker inputs. *Genes Dev.* **15**, 340-351. doi:10.1101/gad.861401

Sandler, J. E., Irizarry, J., Stepanik, V., Dunipace, L., Amrhein, H. and Stathopoulos, A. (2018). A developmental program truncates long transcripts to temporally regulate cell signaling. *Dev. Cell* **47**, 773-784.e6. doi:10.1016/j.devcel.2018.11.019

Sawala, A., Sutcliffe, C. and Ashe, H. L. (2012). Multistep molecular mechanism for Bone morphogenetic protein extracellular transport in the Drosophila embryo. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11222-11227. doi:10.1073/pnas.1202781109

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682. doi:10.1038/nmeth.2019

Schmierer, B. and Hill, C. S. (2007). TGFβ-SMAD signal transduction: molecular specificity and functional flexibility. *Nat. Rev. Mol. Cell Biol.* **8**, 970-982. doi:10.1038/nrm2297

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606-1619. doi:10.1016/S0006-3495(00)76713-0

Serpe, M., Ralston, A., Blair, S. S. and Connor, M. B. (2005). Matching catalytic activity to developmental function: Tolloid-related processes Sog in order to help specify the posterior crossvein in the Drosophila wing. *Development* **132**, 2645. doi:10.1242/dev.01838

Shilo, B.-Z., Haskel-Ittah, M., Ben-Zvi, D., Schejter, E. D. and Barkai, N. (2013). Creating gradients by morphogen shuttling. *Trends Genet.* **29**, 339-347. doi:10.1016/j.tig.2013.01.001

Shimmi, O. and O'Connor, M. B. (2003). Physical properties of Tld, Sog, Tsg and Dpp protein interactions are predicted to help create a sharp boundary in Bmp signals during dorsoventral patterning of the Drosophila embryo. *Development (Cambridge, England)* **130**, 4673-4682. doi:10.1242/dev.00684

Shimmi, O., Ralston, A., Blair, S. S. and O'Connor, M. B. (2005a). The crossveinless gene encodes a new member of the Twisted gastrulation family of BMP-binding proteins which, with Short gastrulation, promotes BMP signaling in the crossveins of the Drosophila wing. *Dev. Biol.* **282**, 70-83. doi:10.1016/j.ydbio.2005.02.029

Shimmi, O., Umulis, D., Othmer, H. and O'Connor, M. B. (2005b). Facilitated transport of a Dpp/Scw heterodimer by Sog/Tsg leads to robust patterning of the Drosophila blastoderm embryo. *Cell* **120**, 873-886. doi:10.1016/j.cell.2005.02.009

Srinivasan, S., Rashka, K. E. and Bier, E. (2002). Creation of a Sog morphogen gradient in the Drosophila embryo. *Dev. Cell* **2**, 91-101. doi:10.1016/S1534-5807(01)00097-1

St Johnston, R. D. and Gelbart, W. M. (1987). Decapentaplegic transcripts are localized along the dorsal-ventral axis of the Drosophila embryo. *EMBO J.* **6**, 2785-2791. doi:10.1002/j.1460-2075.1987.tb02574.x

Sutherland, D. J., Li, M. F., Liu, X. Q., Stefancsik, R. and Raftery, L. A. (2003). Stepwise formation of a SMAD activity gradient during dorsal-ventral patterning of the Drosophila embryo. *Development* **130**, 5705-5716. doi:10.1242/dev.00801

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. and Ludtke, S. J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38-46. doi:10.1016/j.jsb.2006.05.009

Troilo, H., Zuk, A. V., Tunnicliffe, R. B., Wohl, A. P., Berry, R., Collins, R. F., Jowitt, T. A., Sengle, G. and Baldock, C. (2014). Nanoscale structure of the BMP antagonist chordin supports cooperative BMP binding. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13063-13068. doi:10.1073/pnas.1404166111

Tsanov, N., Samacoits, A., Chouaib, R., Traboulsi, A. M., Gostan, T., Weber, C., Zimmer, C., Zibara, K., Walter, T., Peter, M. et al. (2016). smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* **44**, e165. doi:10.1093/nar/gkw784

Tuazon, F. B., Wang, X., Andrade, J. L., Umulis, D. and Mullins, M. C. (2020). Proteolytic restriction of chordin range underlies BMP gradient formation. *Cell Reports* **32**, 108039. doi:10.1016/j.celrep.2020.108039

Umulis, D. M., Serpe, M., O'Connor, M. B. and Othmer, H. G. (2006). Robust, bistable patterning of the dorsal surface of the Drosophila embryo. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11613-11618. doi:10.1073/pnas.0510398103

Umulis, D., O'Connor, M. B. and Blair, S. S. (2009). The extracellular regulation of bone morphogenetic protein signaling. *Development* **136**, 3715-3728. doi:10.1242/dev.031534

Umulis, D. M., Shimmi, O., O'Connor, M. B. and Othmer, H. G. (2010). Organism-scale modeling of early Drosophila patterning via bone morphogenetic proteins. *Dev. Cell* **18**, 260-274. doi:10.1016/j.devcel.2010.01.006

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A. et al. (2022). AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439-D444. doi:10.1093/nar/gkab1061

Vilmos, P., Sousa-Neves, R., Lukacsovich, T. and Lawrence Marsh, J. (2005). crossveinless defines a new family of Twisted-gastrulation-like modulators of bone morphogenetic protein signalling. *EMBO Rep.* **6**, 262-267. doi:10.1038/sj.embor.7400347

Vinter, D. J., Hoppe, C., Minchington, T. G., Sutcliffe, C. and Ashe, H. L. (2021). Dynamics of hunchback translation in real time and at single mRNA resolution in the Drosophila embryo. *Development* **148**, dev.196121.

Wang, Y.-C. and Ferguson, E. L. (2005). Spatial bistability of Dpp-receptor interactions during Drosophila dorsal-ventral patterning. *Nature* **434**, 229-234. doi:10.1038/nature03318

Wang, X. M., Harris, R. E., Bayston, L. J. and Ashe, H. L. (2008). Type IV collagens regulate BMP signalling in Drosophila. *Nature* **455**, U72-U49. doi:10.1038/nature07214

Wang, R. N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S. et al. (2014). Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis.* **1**, 87-105. doi:10.1016/j.gendis.2014.07.005

Wharton, K. A., Cook, J. M., Torres-Schumann, S., de Castro, K., Borod, E. and Phillips, D. A. (1999). Genetic analysis of the bone morphogenetic protein-related gene, gbb, identifies multiple requirements during Drosophila development. *Genetics* **152**, 629-640. doi:10.1093/genetics/152.2.629

Winstanley, J., Sawala, A., Baldock, C. and Ashe, H. L. (2015). Synthetic enzyme-substrate tethering obviates the Tolloid-ECM interaction during Drosophila BMP gradient formation. *Elife* **4**, e05508. doi:10.7554/eLife.05508

Xu, E.-R., Blythe, E. E., Fischer, G. and Hyvönen, M. (2017). Structural analyses of von Willebrand factor C domains of collagen 2A and CCN3 reveal an alternative mode of binding to bone morphogenetic protein-2. *J. Biol. Chem.* **292**, 12516-12527. doi:10.1074/jbc.M117.788992

Yu, K., Sturtevant, M. A., Biehs, B., Francois, V., Padgett, R. W., Blackman, R. K. and Bier, E. (1996). The Drosophila decapentaplegic and short gastrulation genes function antagonistically during adult wing vein development. *Development* **122**, 4033. doi:10.1242/dev.122.12.4033

Zhang, J.-L., Huang, Y., Qiu, L.-Y., Nickel, J. and Sebald, W. (2007). von Willebrand factor type C domain-containing proteins regulate bone morphogenetic protein signaling through different recognition mechanisms. *J. Biol. Chem.* **282**, 20002-20014. doi:10.1074/jbc.M700456200

Zhang, J., Qiu, L., Kotzsch, A., Weidauer, S., Patterson, L., Hammerschmidt, M., Sebald, W. and Mueller, T. D. (2008). Crystal structure analysis reveals how the chordin family member crossveinless 2 blocks BMP-2 receptor binding. *Dev. Cell* **14**, 739-750. doi:10.1016/j.devcel.2008.02.017

Zinski, J., Bu, Y., Wang, X., Dou, W., Umulis, D. and Mullins, M. C. (2017). Systems biology derived source-sink mechanism of BMP gradient formation. *ELife,* **6**, e22199. doi:10.7554/eLife.22199
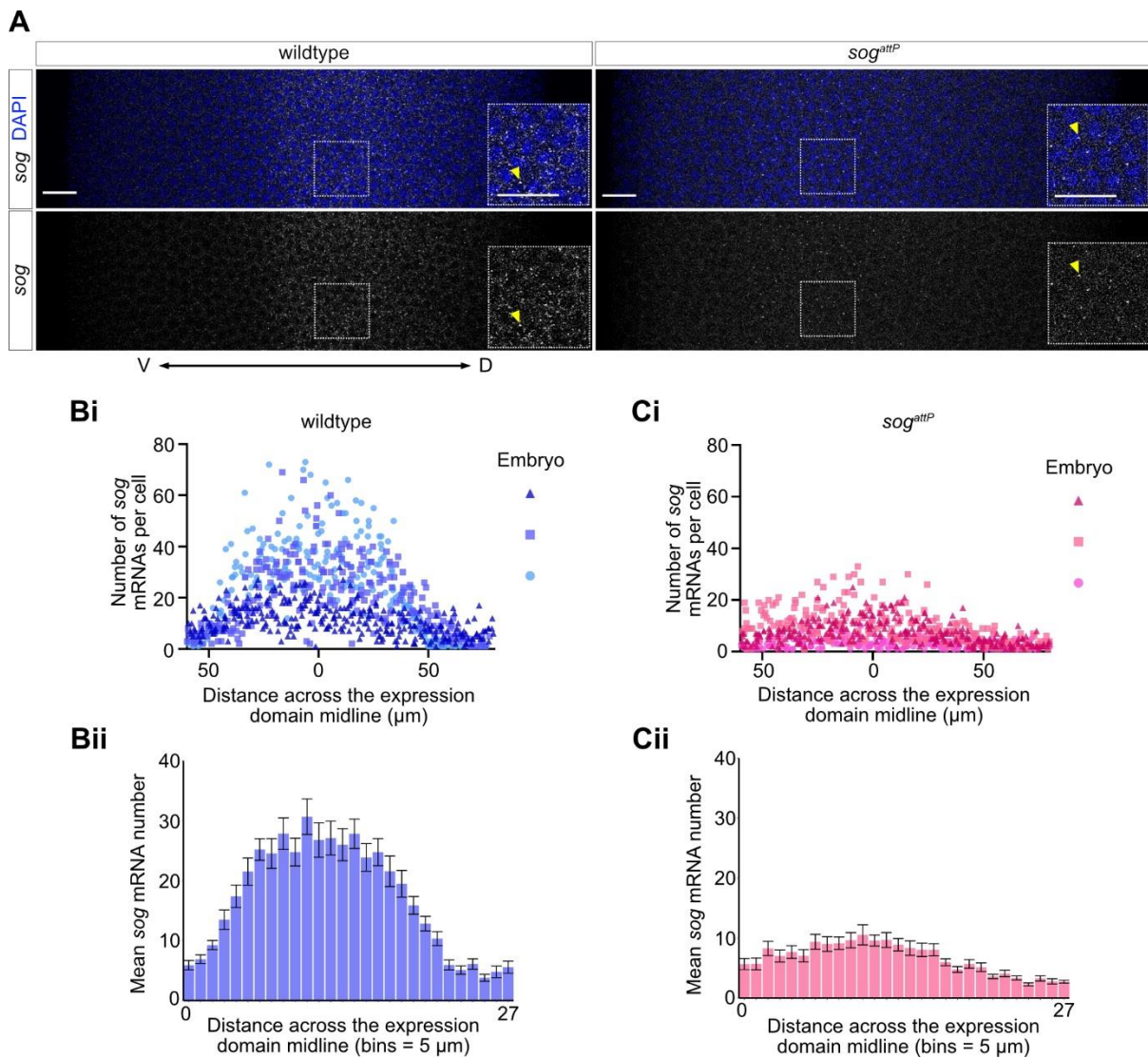
Fig. S1. *sog* is transcribed at low levels in early nc14 *sog^attP* embryos.

A) Confocal images (maximum intensity projections) showing *sog* mRNAs in early nc14

wildtype and *sog^attP* male embryos stained with *sog* smFISH probes and the DAPI nuclear

marker. A merged image and single smFISH channel are shown for clarity. Spectrin

antibody staining was used to measure membrane ingression to age the embryos (data not

shown). Scale bar = 15 µm. Insets are enlarged sections of the *sog* expression domain

(box), scale bar = 15 µm). Yellow arrowheads indicate transcription sites.

Bi) The numbers of *sog* transcripts per cell in three wildtype embryos were quantified and

are plotted against distance from the *sog* expression domain midline (µm, - 60 = ventral

edge). Data for each embryo is presented on the same graph in a shade of blue with a

triangle, square, or circle symbol.

Bii) Data for each wildtype embryo were combined and binned according to distance across

the *sog* expression domain (bins = 5 µm, each bin approximately corresponds to a nuclear

width, bin 0 = ventral edge). Mean *sog* mRNA number per bin is plotted against bin number. Error bars = mean ± SEM.

Ci) As for Bi, but for *sog* mRNA counts in three *sog*^{attP} embryos. Each embryo is represented by a shade of pink and either a triangle, square or circle symbol.

Cii) As for Bii, but for *sog*^{attP} embryos.
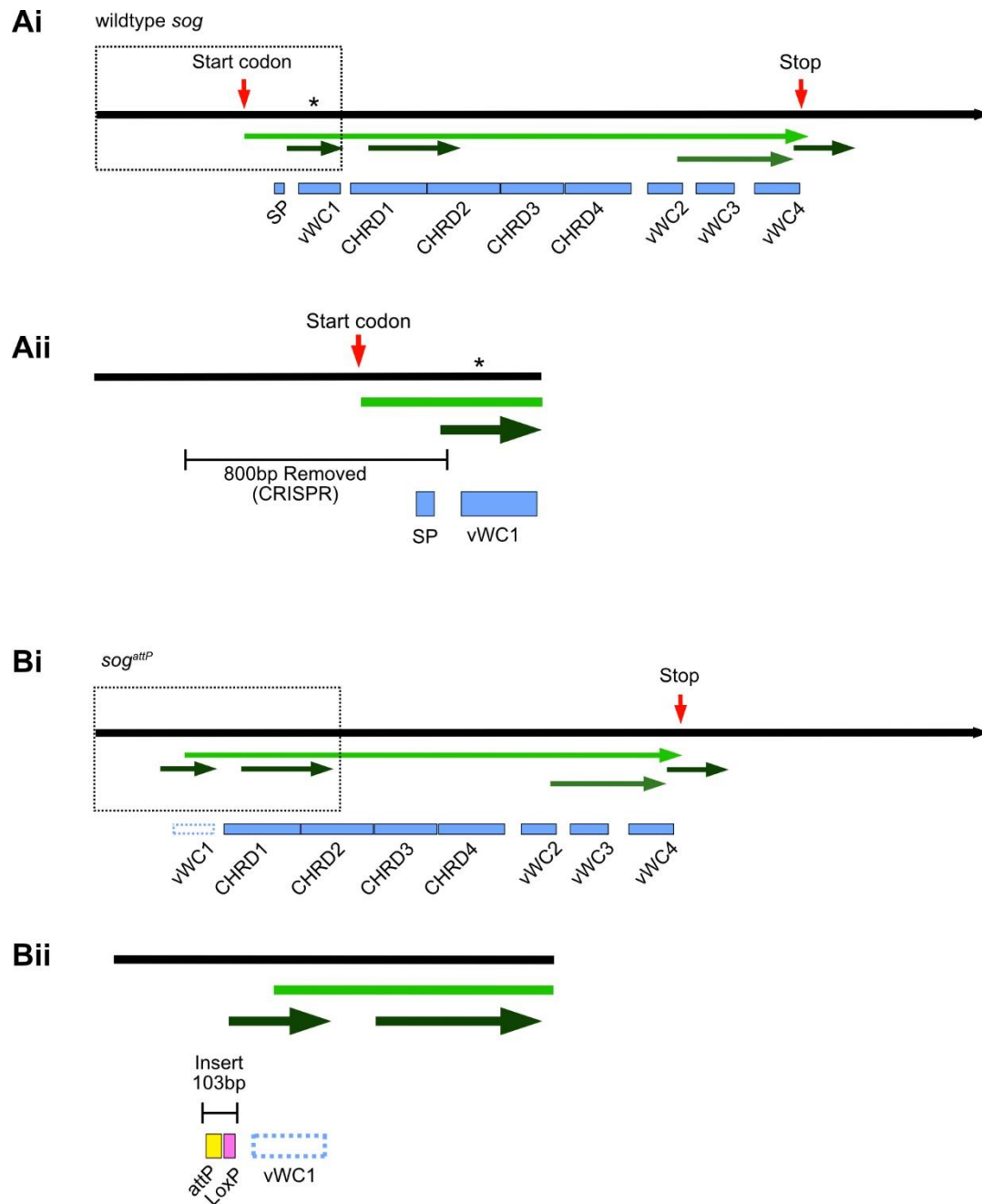
Fig. S2. **Potential ORFs in the wild type *sog* and CRISPR modified *sog*^attP^
transcripts.** Ai) Schematic showing the wildtype *sog* transcript (isoform E, (Larkin et al.,
2021)). ORFs are shown by green arrows, with the shade of green indicative of the
reading frame. A red arrow indicates the endogenous start and stop codons of the *sog*
coding sequence. The asterisk marks the position where a truncated Sog ORF initiates
following the CRISPR deletion (see Bi).

Aii) An enlarged view of the outlined region in Ai) showing the 800 bp sequence removed by CRISPR-Cas9 and the locations of the predicted signal peptide (SP) and vWC1 domain (blue boxes).

Bi) Schematic showing the locations of predicted ORFs (green arrows) in the predicted sog^{attP} transcript (sog isoform E).

Bii) An enlarged view of the outlined region in Bi) showing the location of the 103 bp insert, containing attP and LoxP sequences (yellow and pink boxes, respectively), introduced into the sequence by CRISPR-Cas9 mediated homology directed repair, and the vWC1 domain. Due the CRISPR-Cas9 genome editing, none of the predicted ORFs contain the Sog SP, and therefore any translated protein would not enter the secretory pathway.
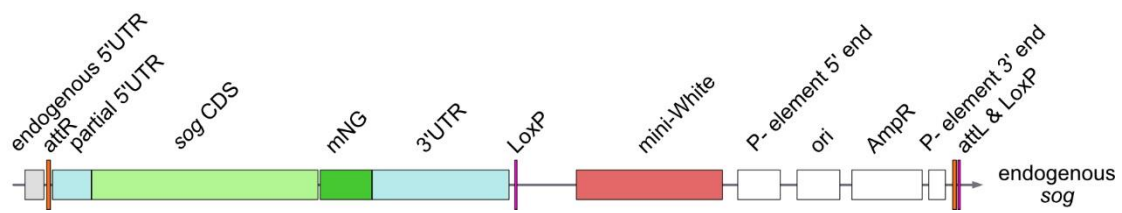
Fig. S3. The *sog* locus after CRISPR-Cas9 editing and reintegration of *sog-mNG* sequences.

The *sog* CDS and entire RIV white plasmid (Baena-Lopez et al., 2013) were inserted into the genome by phiC31 recombination, resulting in insertion of a 12,565 bp sequence. Due to the location of the CRISPR-Cas9 cut sites, some of the *sog* 5'UTR was removed in the initial CRISPR event. The grey box in the schematic represents the remaining endogenous 5'UTR, while the light blue box labelled 'partial 5'UTR' represents the 5'UTR sequence used to replace the 5'UTR sequences which were initially removed. The LoxP sequences (pink) located after the *sog* 3'UTR and P-element 3' end sequences permit removal of 6,081 bp after the 3'UTR by Cre-Lox recombination. However, this was not performed for this study.
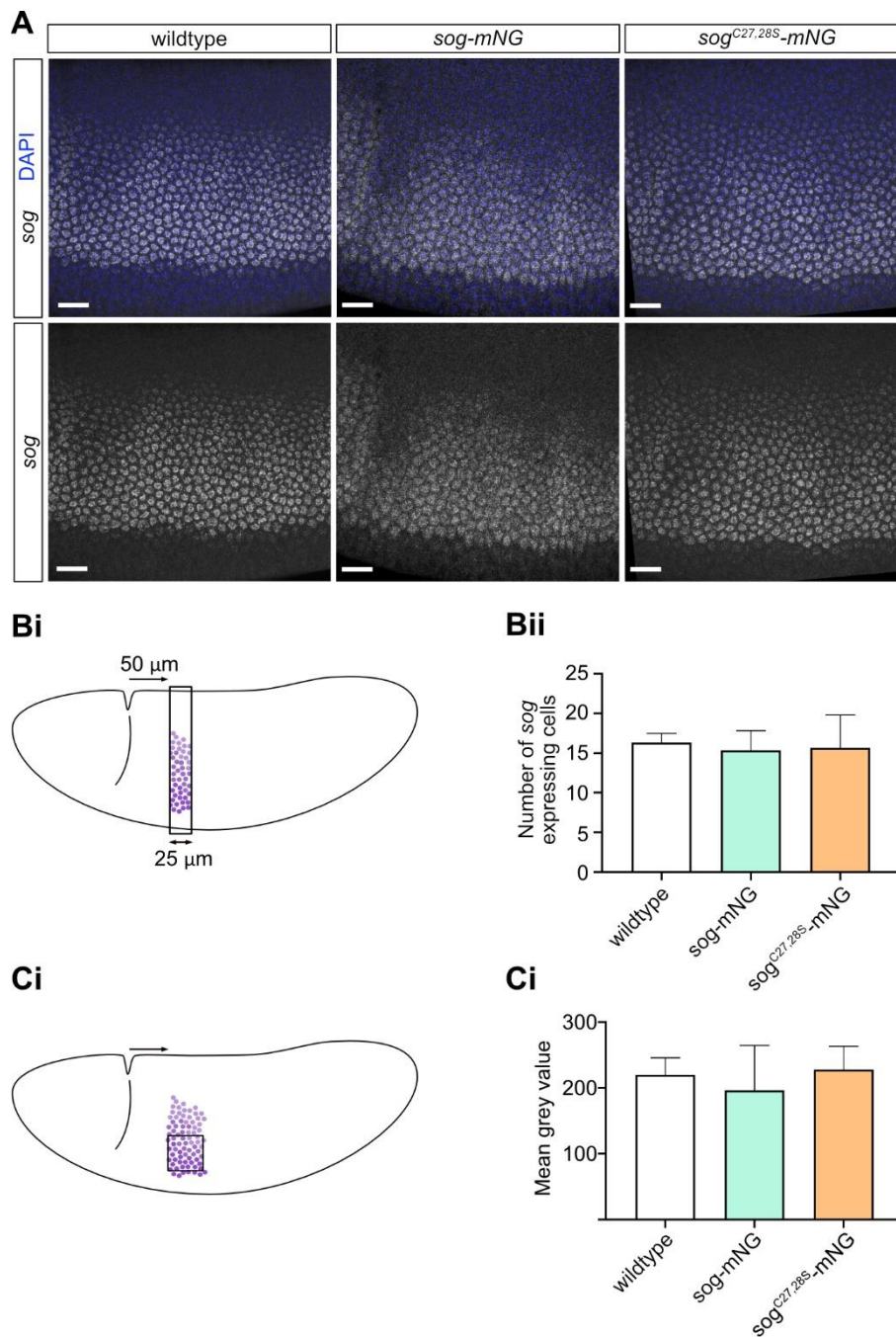
Fig. S4. *sog* expression in nc14 embryos is similar between genotypes.

A) *sog* mRNA was detected in wildtype, *sog-mNG*, and *sog^{C27,28S}-mNG* early stage 6 embryos by smiFISH. Scale bar = 20 μm.

Bi) Cartoon showing the position of a region of interest (ROI) that was used to measure the height of the expression domain based on the maximum number of *sog* expressing cells. The ROI is located 50 μm posterior of the cephalic furrow.

Bii) Graph shows the maximum number of *sog* expressing cells in the ROI depicted in Bi. No significant difference in the height of the *sog* expression domain was found (One-Way ANOVA

with Tukey's multiple comparisons test, P>0.9). n = 3 embryos for each genotype, error bars show standard deviation.

Ci) Cartoon showing a 40x40 µm ROI, 50 µm to the posterior of the cephalic furrow, that was used to estimate *sog* expression level in early stage 6 embryos. Mean fluorescence intensity (based on grey value), of the ROI was measured, as clustering of *sog* mRNAs at this stage prevents the counting of absolute mRNA numbers as is possible in early nc14 embryos (see Fig S1).

Cii) Background subtracted fluorescence intensity (grey) values of the ROI are shown. No significant difference was detected between genotypes (One-Way ANOVA, Tukey's multiple comparisons test, P>0.7). n = 3 embryos for each genotype, error bars show standard deviation.
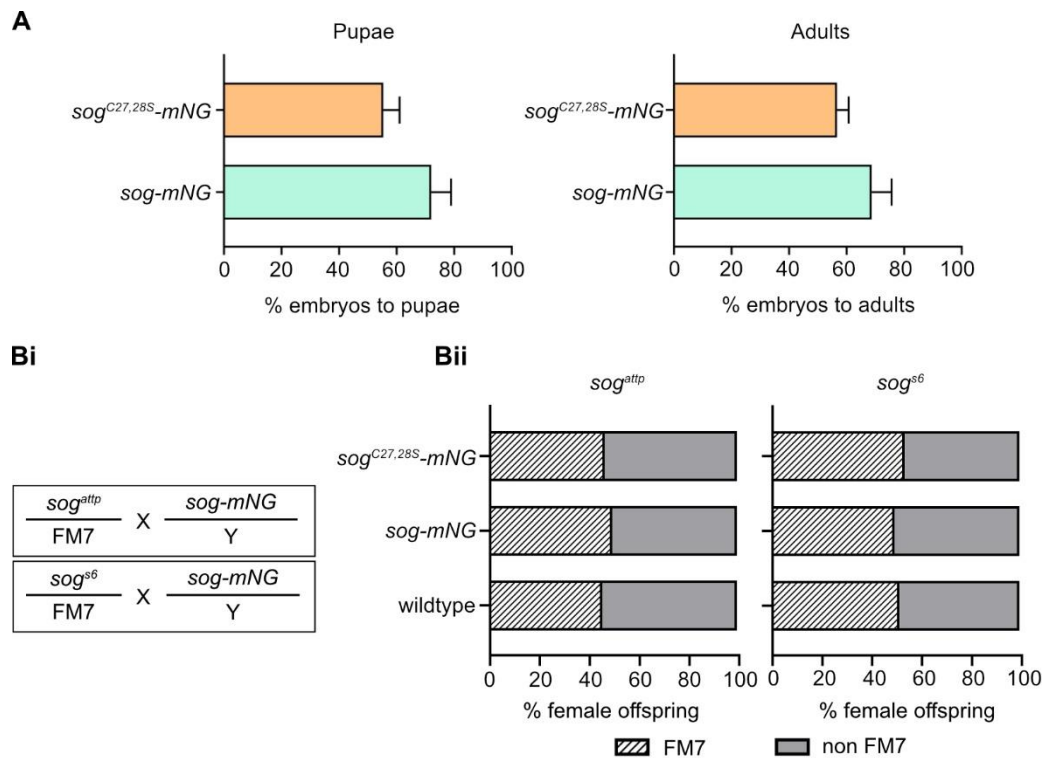
Fig. S5. Assessing viability of *sog-mNG* and *sog*^{C27,28S}*-mNG* flies.

A) Graphs show the percentage survival of *sog-mNG* and *sog*^{C28,28S}*-mNG* embryos to pupal and adult stages. No significant difference was found between the proportion of *sog-mNG* and *sog*^{C27,28S}*-mNG* flies that survived to pupae (p=0.14) and adulthood (p=0.21) (t-test, unpaired, two-tailed. t-tests were performed on raw data).

Bi) Overview of the crosses used to compare the ability of the *sog-mNG* and *sog*^{C27,28S}*-mNG* alleles to rescue either the strong *sog*^{S6} loss of function or *sog*^{attP} null alleles. Wildtype males were also tested as a control, only the *sog-mNG* allele is shown in the crossing scheme for simplicity. The number of female FM7 and non-FM7 progeny were counted. For female progeny that do not carry FM7, the *sog* allele has successfully rescued the *sog*^{attP} or *sog*^{S6} loss of function alleles.

Bii) The percentage of FM7 (striped bars) and non-FM7 (grey bars) progeny from the crosses described in Bi are shown. There is no significant association between the number of FM7/non-FM7 offspring and the genotype of fathers for crosses with *sog*^{attP} (X^2 (df = 2, N = 1764) = 2.07, p = 0.36) and *sog*^{S6} (X^2 (df = 2, N = 1596) = 1.71, p = 0.43) mothers. Statistical test was performed on raw data. Percentage frequency of the phenotypes is shown in the figure for simplicity.
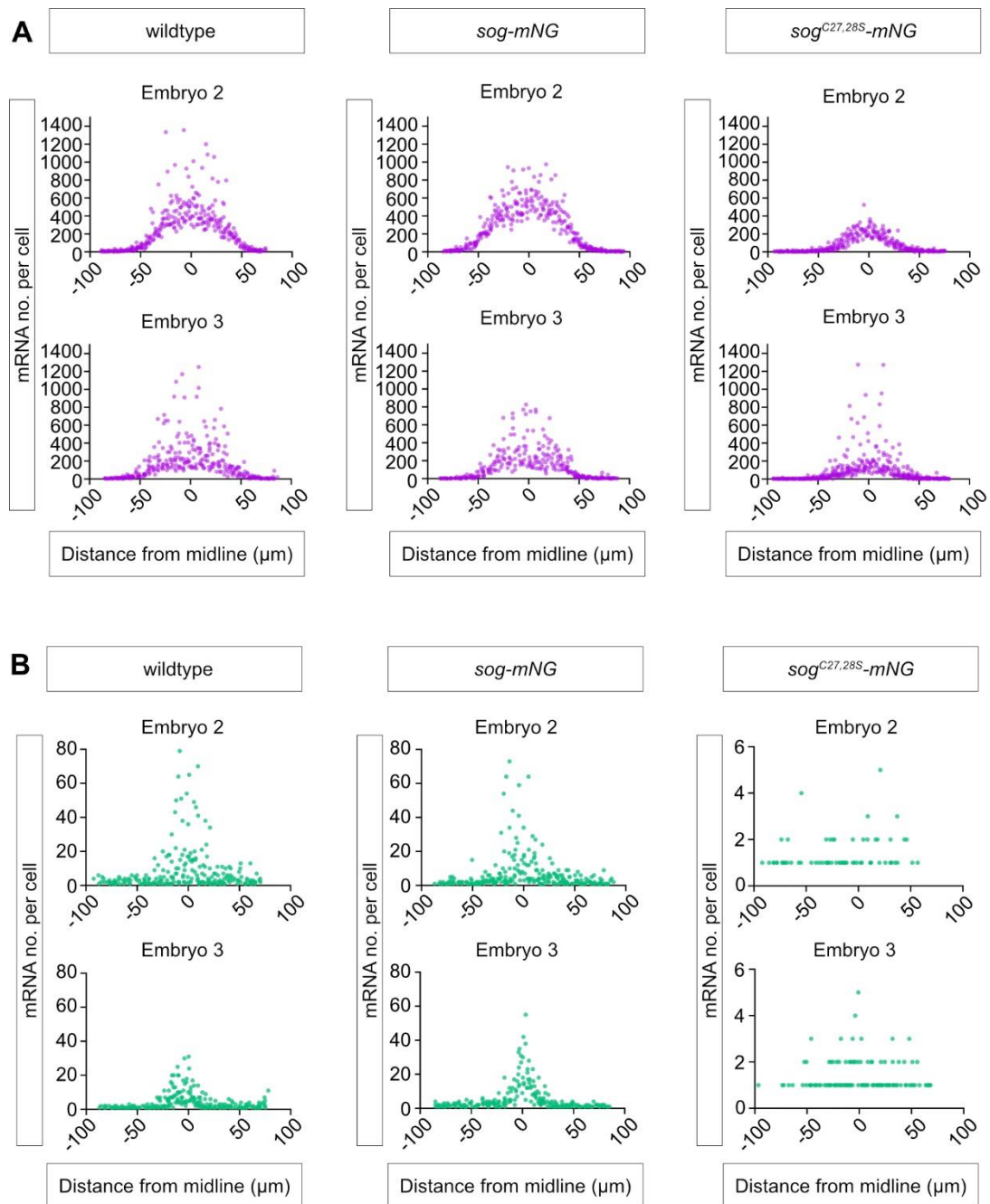
Fig. S6. Quantification of *ush* and *Race* expression.

(A) The numbers of *ush* transcripts per cell plotted against distance from the dorsal midline (µm, 0 = midline) for the additional biological repeats relating to Fig.4.

(B) As in (A), but for *Race* mRNAs.
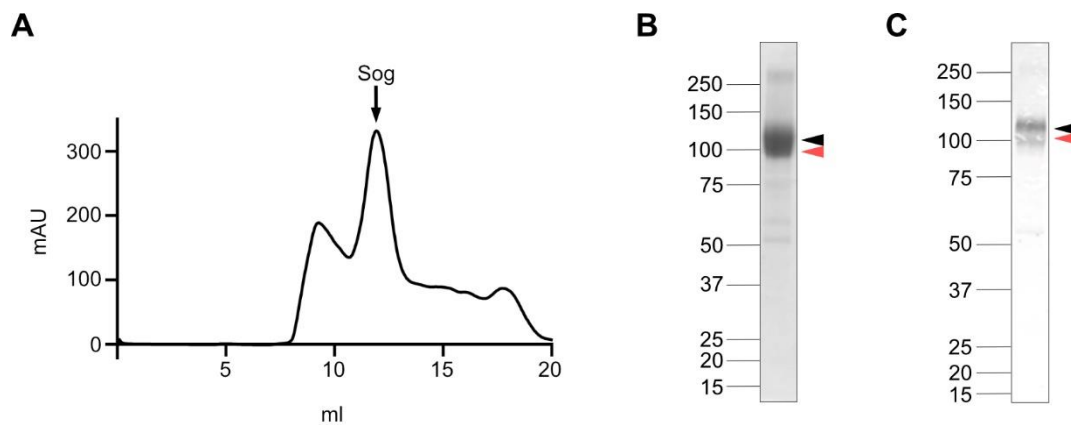
Fig. S7. Purification of Sog by size exclusion chromatography.

A) Following affinity chromatography using a His-tag, Sog was purified by SEC. The peak Sog-containing fraction, indicated by an arrow, was subjected to a second round of SEC (main text, Fig.6) prior to electron microscopy.

B) Reduced SDS-PAGE gel showing Sog from the peak fraction in A. The black arrow indicates full-length Sog and the red arrow indicates a Sog cleavage product.

C) Anti-His western blot of the same sample shown in B. Arrows are as in B.
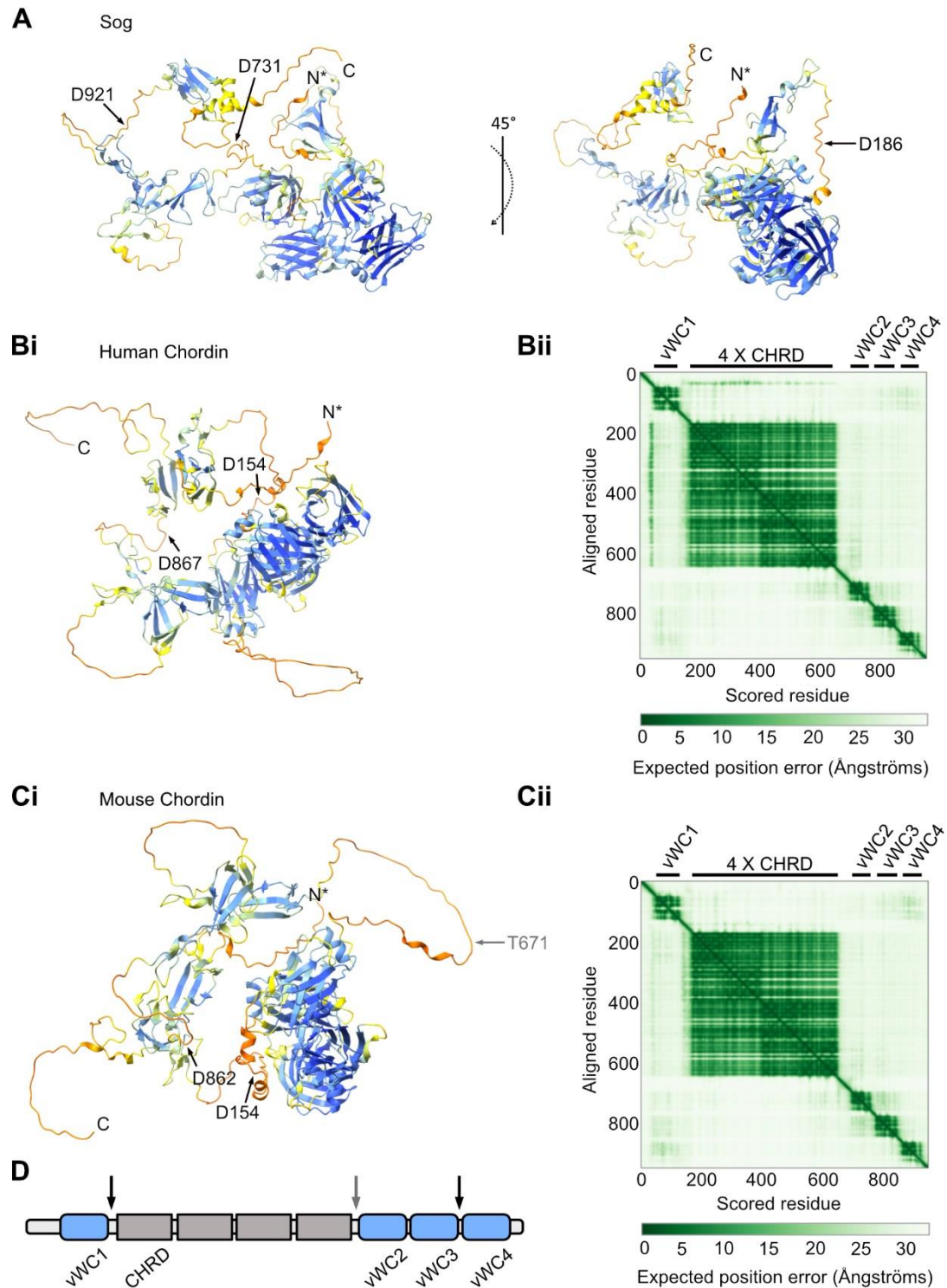
Fig. S8. Interdomain regions that contain the Sog/Chordin Tolloid cleavage sites are flexible.

A) AlphaFold2 structure prediction for Sog (Jumper et al., 2021; Varadi et al., 2022), colour coded according to per residue confidence (pLDDT) score (dark blue, very high confidence; light blue, confident; yellow, low confidence; orange, very low confidence). Tld cleaves Sog

before D921, D731, and D186 (arrows) (Peluso et al., 2011; Shimmi & O'Connor, 2003) . In the figure shown, the Sog N-terminus is truncated at R79 (N*).

Bi) AlphaFold2 structure prediction for human Chordin (UniProt Q9H2X0), coloured as for Sog in (A). BMP1/Tld-like protease cleavage sites are annotated (arrows) (Scott et al., 1999). The predicted Chordin structure has been truncated following the signal peptide at G26 (N*).

Bii) Predicted aligned error (PAE) plot for the AlphaFold human Chordin prediction. Approximate boundaries for vWC and CHRD (4 X CHRD) domains are indicated by black bars above the heatmap.

Ci) AlphaFold2 structure prediction for mouse Chordin (Uniprot Q9Z0E2), coloured as for human Chordin and Sog. Tld cleavage sites are indicated by arrows. The predicted structure has been truncated after the signal peptide at G26 (N*). Two cleavage sites (D154 and D862) are conserved in human Chordin. A third Tld cleavage site (T671, grey arrow) in mouse Chordin is also used in the presence of Tsg *in vitro* (Scott et al., 2001).

Cii) Predicted aligned error (PAE) plot for the AlphaFold mouse Chordin prediction.

D) Cartoon of Chordin (not to scale), showing the vWC and CHRD domains in relation to Tld cleavage sites present in both mouse and human Chordin (black arrows), and the cryptic Tld target site (grey arrow) found in mouse Chordin in the presence of Tsg.

References

Baena-Lopez, L. A., Alexandre, C., Mitchell, A., Pasakarnis, L., & Vincent, J.-P. (2013). Accelerated homologous recombination and subsequent genome modification in Drosophila. *Development*, *140*(23), 4818.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873*, *596*(7873), 583–589.

Larkin, A., Marygold, S. J., Antonazzo, G., Attrill, H., dos Santos, G., Garapati, P. V, Goodman, J. L., Gramates, L. S., Millburn, G., Strelets, V. B., Tabone, C. J., Thurmond, J., Consortium, F., Perrimon, N., Gelbart, S. R., Agapite, J., Broll, K., Crosby, M., dos Santos, G., … Lovato, T. (2021). FlyBase: updates to the Drosophila melanogaster knowledge base. *Nucleic Acids Research*, *49*(D1), D899–D907.

Peluso, C. E., Umulis, D., Kim, Y.-J., O'Connor, M. B., & Serpe, M. (2011). Shaping BMP Morphogen Gradients through Enzyme-Substrate Interactions. *Developmental Cell*, *21*(2), 375–383.

Scott, I. C., Blitz, I. L., Pappano, W. N., Imamura, Y., Clark, T. G., Steiglitz, B. M., Thomas, C. L., Maas, S. A., Takahara, K., Cho, K. W. Y., & Greenspan, D. S. (1999). Mammalian BMP-1/Tolloid-Related Metalloproteinases, Including Novel Family Member Mammalian Tolloid-Like 2, Have Differential Enzymatic Activities and Distributions of Expression Relevant to Patterning and Skeletogenesis. *Developmental Biology*, *213*(2), 283–300.

Scott, I. C., Blitz, I. L., Pappano, W. N., Maas, S. A., Cho, K. W. Y., & Greenspan, D. S. (2001). Homologues of Twisted gastrulation are extracellular cofactors in antagonism of BMP signalling. *Nature*, *410*(6827), 475–478.

Shimmi, O., & O'Connor, M. B. (2003). Physical properties of Tld, Sog, Tsg and Dpp protein interactions are predicted to help create a sharp boundary in Bmp signals during dorsoventral patterning of the Drosophila embryo. *Development (Cambridge, England)*, *130*(19), 4673–4682.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., … Velankar, S. (2022). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, *50*(D1).

Table. S1. Primer and oligonucleotide sequences.

Plasmid	Primer	Sequence (5'-3')
pU6-BbsI-chiRNA-guides	Guide 1 Forwards	[Phos]CTTCGAGTCGATCTCGTATGAGGA
	Guide 1 Reverse	[Phos]AAACTCCTCATACGAGATCGACTC
	Guide 2 Forwards	[Phos]CTTCGCATGCGCCGCTCATGTTCG
	Guide 2 Reverse	[Phos]AAACCGAACATGAGCGGCGCATGC
pTV-Cherry-homology-arms	Homolog arm 1 Forwards	CTAGCACATATGCAGGTACCTTTAAGATTGTCAGCATTGCA
	Homolog arm 1 Reverse	AGTTGGGGCACTACGGTACCATTACGACAACGCGACTTTT
	Homology arm 2 Forwards	CGAAGTTATCACTAGTAGTCCGACACGGGCAGGC
	Homology arm 2 Reverse	GGAGATCTTTACTAGTGCAACTCGGGAACATAATAG
pAC-p5UTR-Sog CDS-link	KpnI p5UTR F	GGTACCTCATACGAGATCGACTCTATTTTCC
	NotI SogCDS R	GCGGCCGCGCTGGAGGATCGCTGCT
	Linker sense	[Phos}GGCCGCCGGTGGTGGTGGAAGTGGCGGAGGTGGAGGCC
	Linker antisense	[Phos]CTCCACCTCCGCCACTTCCACCACCACCGGC
pAC-mNeonGreen-sog3UTR	Kpn1 NG F	GGTACCATGGTGAGCAAGGGCG
	NotI NG R	GCGGCCGCTTACTTGTACAGCTCGTCCATG
	pACng3UTRinfFwd	CAAGTAAGCGGCCGCGCGGCTCCACGTGACGGAT
	pAC3UTRinfRev0818	ACCTTCGAAGGGCCCATGGGTATATTTCGAATATATTTTGTCTATATTTC
pAC-p5UTR-sogCDS-link-mNeonGreen-3UTR	ApaI NG F	GGGCCCATGGTGAGCAAGGGCG
	BstBI 3UTR R	TTCGAAATGGGTATATTTCGAATATATTTTGTCTATATTTCAATTTA
RIV-p5UTR-sog-link-	RIV p5UTR inffwd 250418	GGCGCGTACTCCACGAATTCTCATACGAGATCGACTCTAT
	Linker rev20818	TCCACCTCCGCCACT

mNeonGreen-3UTR	Link-NGinfFwd0818	GTGGAAGTGGCGGAGGTGGAATGGTGAGCAAGGGC
	RIV 3UTR infrev 250418	GCGGCCGCTCCGGAGAATTCATGGGTATATTTCGAATATA
RIV-p5UTR-sogPALM-link-mNeonGreen-3UTR	RIV p5UTR inffwd 250418	GGCGCGTACTCCACGAATTCTCATACGAGATC GACTCTAT
	palm mut rev 161018	GCGGCGTCCTCGCTGTGAGAAGAGCTCCTTTCCAGGAGC
	palm mutfwd3 1218	TCTCACAGCGAGGACGCCGC
	RIV 3UTR infrev 250418	GCGGCCGCTCCGGAGAATTCATGGGTATATTTCGAATATA
Probe	**Primer**	**Sequence (5'-3')**
mNeonGreen-biotin-UTP	T3 promoter, forward.	ATTAACCCTCACTAAAGGGAATGGTGAGCAAGGGCGAGGAG GAT
	T7 promoter, reverse	GAATTAATACGACTCACTATAGGGATTACTTGTACAGCTCG TCCATGCC

Table. S2. Sequences of the *Race* smiFISH, *lacZ* Stellaris, and *sog* Stellaris and exonic smiFISH probes. *Race* smiFISH probes fused to Quasar 570-conjugated Y-FLAP, sog exonic smiFISH probes fused to Quasar 570-conjugated Z-FLAP, and *sog*, *ush* (Quasar 570 conjugated) and *lacZ* (Quasar 670 conjugated) Stellaris probes were used for smFISH.

Probe name	No.	Probe 5'-3'
race Y FLAP	1	TTACACTCGGACCTCGTCGACATGCATTATCCGTTAATTGCCCTAA
race Y FLAP	2	TTACACTCGGACCTCGTCGACATGCATTCCCCAATTAAAGGCTACT
race Y FLAP	3	TTACACTCGGACCTCGTCGACATGCATTTCGCGAAACAGTTCGCCA
race Y FLAP	4	TTACACTCGGACCTCGTCGACATGCATTTCCGTGGTGTGTACTACA
race Y FLAP	5	TTACACTCGGACCTCGTCGACATGCATTCGACCATGTCCACCGAAA
race Y FLAP	6	TTACACTCGGACCTCGTCGACATGCATTACCACCGCAGAAGTGTTT
race Y FLAP	7	TTACACTCGGACCTCGTCGACATGCATTTCGCTTTTGGTACACAGT
race Y FLAP	8	TTACACTCGGACCTCGTCGACATGCATTCGATCGTAAGTGCAACCT
race Y FLAP	9	TTACACTCGGACCTCGTCGACATGCATTTGGTTGGTTCGATCTGTT
race Y FLAP	10	TTACACTCGGACCTCGTCGACATGCATTACTCTGGAAGTCACTTCA
race Y FLAP	11	TTACACTCGGACCTCGTCGACATGCATTCTCAAGTGATTGCCCACA

race Y FLAP	12	TTACACTCGGACCTCGTCGACATGCATTGGATGACTCTGGGGTCAG
race Y FLAP	13	TTACACTCGGACCTCGTCGACATGCATTGGGCTAGCAGAAACAGTC
race Y FLAP	14	TTACACTCGGACCTCGTCGACATGCATTTACCGCCAAAGTAGCCAG
race Y FLAP	15	TTACACTCGGACCTCGTCGACATGCATTCTCCTTGACCAGCGCTTG
race Y FLAP	16	TTACACTCGGACCTCGTCGACATGCATTTACTCCTTGGCCTGTATC
race Y FLAP	17	TTACACTCGGACCTCGTCGACATGCATTCCTTGTTCAGATTCTCCA
race Y FLAP	18	TTACACTCGGACCTCGTCGACATGCATTGTTGGTTCGCTTGGCCAG
race Y FLAP	19	TTACACTCGGACCTCGTCGACATGCATTCAGGCAGCTTCGGTTTCC
race Y FLAP	20	TTACACTCGGACCTCGTCGACATGCATTTGATGTTGGAGCCATAGG
race Y FLAP	21	TTACACTCGGACCTCGTCGACATGCATTTTTCTTTTCGTTCTCGTC
race Y FLAP	22	TTACACTCGGACCTCGTCGACATGCATTCTCGGCGGATATCTCATT
race Y FLAP	23	TTACACTCGGACCTCGTCGACATGCATTTCCTTCATGAACTTGGCC
race Y FLAP	24	TTACACTCGGACCTCGTCGACATGCATTCTTGGTGGTATCACTGGC
race Y FLAP	25	TTACACTCGGACCTCGTCGACATGCATTATTGGTACGAGCGCCATT
race Y FLAP	26	TTACACTCGGACCTCGTCGACATGCATTAACTGGCGCTTGAGATCC
race Y FLAP	27	TTACACTCGGACCTCGTCGACATGCATTAGCCCAGTTTGGTTAGAG
race Y FLAP	28	TTACACTCGGACCTCGTCGACATGCATTGTCTTCAGGTAGAGCAGC
race Y FLAP	29	TTACACTCGGACCTCGTCGACATGCATTTCCAGCAGTTCGGCATAG
race Y FLAP	30	TTACACTCGGACCTCGTCGACATGCATTACTCCATGGCGGAGAGTG
race Y FLAP	31	TTACACTCGGACCTCGTCGACATGCATTCCTTGACCTTGGCGAAAT
race Y FLAP	32	TTACACTCGGACCTCGTCGACATGCATTGCTATCCTTGTAGTCGCA
race Y FLAP	33	TTACACTCGGACCTCGTCGACATGCATTTCATCCAACCAGGCCTCG
race Y FLAP	34	TTACACTCGGACCTCGTCGACATGCATTCGAAGGTGTCGTCCTCGT
race Y FLAP	35	TTACACTCGGACCTCGTCGACATGCATTGATGTCCTCCAGCTGCTG
race Y FLAP	36	TTACACTCGGACCTCGTCGACATGCATTAGCGGACGAATATCCGCG
race Y FLAP	37	TTACACTCGGACCTCGTCGACATGCATTAGCCATGGATCTGCTGGT
race Y FLAP	38	TTACACTCGGACCTCGTCGACATGCATTCCTCAGGCGGAAACGCAC
race Y FLAP	39	TTACACTCGGACCTCGTCGACATGCATTACCGCGTCACCATAGTGT
race Y FLAP	40	TTACACTCGGACCTCGTCGACATGCATTTGGGTCCTGTCTCGGAGA
race Y FLAP	41	TTACACTCGGACCTCGTCGACATGCATTGCCCAATAGGTGCATGGG
race Y FLAP	42	TTACACTCGGACCTCGTCGACATGCATTCACTGCTGTGCCCACATG
race Y FLAP	43	TTACACTCGGACCTCGTCGACATGCATTCGATGTCCGCAATCTCTG
race Y FLAP	44	TTACACTCGGACCTCGTCGACATGCATTCTTCTCCGGAAAGGGGGA
race Y FLAP	45	TTACACTCGGACCTCGTCGACATGCATTAGCGCTCACATCCACCAG
race Y FLAP	46	TTACACTCGGACCTCGTCGACATGCATTTAGCCCTGCTTTTCCATC
race Y FLAP	47	TTACACTCGGACCTCGTCGACATGCATTGGAACATTTTGAGTGGCG
race Y FLAP	48	TTACACTCGGACCTCGTCGACATGCATTGAAGAAGTCGTCGCCCAT
sog_570	1	ATTCGATGGCGTTCGATTTC
sog_570	2	CAAGCAGACGATCAGCAGTC
sog_570	3	AATTCGCGCAAAACTTTGCC
sog_570	4	TACATAACTCCGAAGGGTGG

sog_570	5	CCACACATTCACACTTGATG
sog_570	6	GGGCACTCGTTTTTGATATT
sog_570	7	GAGATGGGATCATCGCATTT
sog_570	8	TACATCCGTATCGTTTCGAT
sog_570	9	TAGCAACGCAGCGTAATGTT
sog_570	10	TGAGGAAATAGGAGGTGCGG
sog_570	11	TACATGGACTTCATTTCCTC
sog_570	12	CACATTCTGCGGATTGTAGG
sog_570	13	TTGTGGAACAGGAAACGGGC
sog_570	14	GCGATGAGGTGTAGAAGGAG
sog_570	15	AATTGAATGGCACGCGGACG
sog_570	16	GATAACACCCGCATCATCAA
sog_570	17	TGATAGACACTGAGAGTGCC
sog_570	18	GCAGAATGCGCTTGTAATCA
sog_570	19	GGAGGACAACATGGAGACGA
sog_570	20	AACTGAACAACTCCGTCTGC
sog_570	21	CATTGAAGACCAGGGTGAGA
sog_570	22	GCTCAATTTTCACACTCAGT
sog_570	23	CACACGTGGAATCTCATCGA
sog_570	24	GTATGGAAATGGGCGACGAC
sog_570	25	ACGCGACATCAGTCGAAGAT
sog_570	26	AGATGTGGGTACTTCTTGGA
sog_570	27	TCTGGAAGATTTCGCAGCTG
sog_570	28	ATCCATCGGTGTTCAAGTAG
sog_570	29	CAAACTGATGTTGGGCCTAT
sog_570	30	TGGTTGAAGTTGAAGCTCGG
sog_570	31	AACTTCTCCACACTACCAAT
sog_570	32	ATTGCACTCGTTGTCAATGG
sog_570	33	CGTTGAATTCCTCGAGCAAT
sog_570	34	AAGAAGCCTTCCAGATAGGA
sog_570	35	TGGAATGGACCTCCAGATAG
sog_570	36	AGCAGAAGCTGTTTGGAGTG
sog_570	37	AACATTGTTGTCCGTGTAGA
sog_570	38	CCGTTCACCAAATGGTTATC
sog_570	39	CGATTCGTTGTAGAAGCGTC
sog_570	40	CACATCTGACAGGAATCCTG
sog_570	41	CATTCACCATCACGCTGAAG
sog_570	42	GAGGATTGCGCTGAATAGTC
sog_570	43	CAGGAATGGATGCCAACTGG
sog_570	44	CTTCTTGTCTGGACGATAGG
sog_570	45	TCTTGTGTCCATTGGAACTG

sog_570	46	CAGCACATTGGGATTGTTCG
sog_570	47	TTCTCGTACACCTTGTTGAC
sog_570	48	TGGGACATCAGGATCGGATG
sog Z FLAP	1	CCAGCTTCTAGCATCCATGCCCTATAAGATTGTAGGTGGTGTACATGG
sog Z FLAP	2	CCAGCTTCTAGCATCCATGCCCTATAAGTTGTGGAACAGGAAACGGGC
sog Z FLAP	3	CCAGCTTCTAGCATCCATGCCCTATAAGGCGATGAGGTGTAGAAGGAG
sog Z FLAP	4	CCAGCTTCTAGCATCCATGCCCTATAAGATAACACCCGCATCATCAAC
sog Z FLAP	5	CCAGCTTCTAGCATCCATGCCCTATAAGTGATAGACACTGAGAGTGCC
sog Z FLAP	6	CCAGCTTCTAGCATCCATGCCCTATAAGGCAGAATGCGCTTGTAATCA
sog Z FLAP	7	CCAGCTTCTAGCATCCATGCCCTATAAGAACTGAACAACTCCGTCTGC
sog Z FLAP	8	CCAGCTTCTAGCATCCATGCCCTATAAGCATTGAAGACCAGGGTGAGA
sog Z FLAP	9	CCAGCTTCTAGCATCCATGCCCTATAAGGCTCAATTTTCACACTCAGT
sog Z FLAP	10	CCAGCTTCTAGCATCCATGCCCTATAAGCACACGTGGAATCTCATCGA
sog Z FLAP	11	CCAGCTTCTAGCATCCATGCCCTATAAGACGCGACATCAGTCGAAGAT
sog Z FLAP	12	CCAGCTTCTAGCATCCATGCCCTATAAGATCCATCGGTGTTCAAGTAG
sog Z FLAP	13	CCAGCTTCTAGCATCCATGCCCTATAAGCAAACTGATGTTGGGCCTAT
sog Z FLAP	14	CCAGCTTCTAGCATCCATGCCCTATAAGTGGTTGAAGTTGAAGCTCGG
sog Z FLAP	15	CCAGCTTCTAGCATCCATGCCCTATAAGAACTTCTCCACACTACCAAT
sog Z FLAP	16	CCAGCTTCTAGCATCCATGCCCTATAAGATTGCACTCGTTGTCAATGG
sog Z FLAP	17	CCAGCTTCTAGCATCCATGCCCTATAAGCGTTGAATTCCTCGAGCAAT
sog Z FLAP	18	CCAGCTTCTAGCATCCATGCCCTATAAGAAGAAGCCTTCCAGATAGGA
sog Z FLAP	19	CCAGCTTCTAGCATCCATGCCCTATAAGTTGGAGTGCTTGGAATGGAC
sog Z FLAP	20	CCAGCTTCTAGCATCCATGCCCTATAAGCGATTCGTTGTAGAAGCGTC
sog Z FLAP	21	CCAGCTTCTAGCATCCATGCCCTATAAGCACATCTGACAGGAATCCTG
sog Z FLAP	22	CCAGCTTCTAGCATCCATGCCCTATAAGGAGGATTGCGCTGAATAGTC
sog Z FLAP	23	CCAGCTTCTAGCATCCATGCCCTATAAGCTTCTTGTCTGGACGATAGG
sog Z FLAP	24	CCAGCTTCTAGCATCCATGCCCTATAAGTCTTGTGTCCATTGGAACTG
lacZ 670	1	CTTCTTGGGGATTCCAATAG
lacZ 670	2	AATGGGATAGGTCACGTTGG
lacZ 670	3	GAACAAACGGCGGATTGACC
lacZ 670	4	TTGCACCACAGATGAAACGC
lacZ 670	5	TTCAGACGGCAAACGACTGT
lacZ 670	6	CGCGTAAAAATGCGCTCAGG
lacZ 670	7	TCCTGATCTTCCAGATAACT
lacZ 670	8	GAGACGTCACGGAAAATGCC
lacZ 670	9	ATCGCTGATTTGTGTAGTCG
lacZ 670	10	CACCCTGCCATAAAGAAACT
lacZ 670	11	CTCATCGATAATTTCACCGC
lacZ 670	12	ACGTTCAGACGTAGTGTGAC
lacZ 670	13	TAACGCCTCGAATCAGCAAC
lacZ 670	14	TGACCATGCAGAGGATGATG

lacZ 670	15	CACGGCGTTAAAGTTGTTCT
lacZ 670	16	AGCGGATGGTTCGGATAATG
lacZ 670	17	GGGTTTCAATATTGGCTTCA
lacZ 670	18	GATCATCGGTCAGACGATTC
lacZ 670	19	GATCACACTCGGGTGATTAC
lacZ 670	20	GGATCGACAGATTTGATCCA
lacZ 670	21	CGCGTACATCGGGCAAATAA
lacZ 670	22	AAGCCATTTTTTGATGGACC
lacZ 670	23	TATTCGCAAAGGATCAGCGG
lacZ 670	24	GAAACCGCCAAGACTGTTAC
lacZ 670	25	CCTGTAAACGGGGATACTGA
lacZ 670	26	CATACAGAACTGGCGATCGT
lacZ 670	27	AAACTGCTGCTGGTGTTTTG
lacZ 670	28	CGCTATGACGGAACAGGTAT
lacZ 670	29	GTAGTTCAGGCAGTTCAATC
lacZ 670	30	ACCCAGCTCGATGCAAAAAT
lacZ 670	31	CTGACTGGCGGTTAAATTGC
lacZ 670	32	TTGTTTTTTATCGCCAATCC
lacZ 670	33	ACTTACGCCAATGTCGTTAT
lacZ 670	34	AATAAGGTTTTCCCCTGATG
lacZ 670	35	CAACGGTAATCGCCATTTGA
lacZ 670	36	GGAAGACGTACGGGGTATAC
lacZ 670	37	GTGGGCCATAATTCAATTCG
lacZ 670	38	TCAGTTGCTGTTGACTGTAG
lacZ 670	39	GGAAACCGTCGATATTCAGC
lacZ 670	40	CACCAGACCAACTGGTAATG
lacZ 670	41	GCCCGGTTATTATTATTTTT
lacZ 670	42	CTTACGCGAAATACGGGCAG
lacZ 670	43	TCCTTCACAAAGATCCTCTA
lacZ 670	44	TTGTCCAATTATGTCACACC
lacZ 670	45	AGTTCCATAGGTTGGAATCT
lacZ 670	46	CATTAAAGGCATTCCACCAC
lacZ 670	47	ATGGCATTTCTTCTGAGCAA
lacZ 670	48	CTCTTCTTTTTTGGAGGAGT