bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Modulation of promoter occupancy dictates the transcriptional response to graded BMP signalling levels in the *Drosophila* embryo

Caroline Hoppe[1], Jonathan Bowles[1], Thomas G. Minchington[1], Catherine Sutcliffe[1], Priyanka Upadhyai[1,2], Magnus Rattray[1*] and Hilary L. Ashe[1*#]

Affiliations:

[1]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, M13 9PT, UK.

[2]Present address: Department of Medical Genetics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India.

*Correspondence: hilary.ashe@manchester.ac.uk, magnus.rattray@manchester.ac.uk

[#]Lead contact: hilary.ashe@manchester.ac.uk

Keywords: BMP signalling, Drosophila embryo, monoallellic transcription, Dpp, transcription kinetics, bursting, promoter occupancy, MS2, live imaging

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SUMMARY

Morphogen gradients specify cell fates during development, with a classic example being the BMP gradient's conserved role in embryonic dorsal-ventral axis patterning. Here we use quantitative imaging and computational modelling to determine how the BMP gradient is interpreted at single-cell resolution in the *Drosophila* embryo. We show that BMP signalling levels are decoded by modulating promoter occupancy, the time the promoter is active, predominantly through regulating the promoter activation rate. As a result, graded mRNA numbers are detected for BMP target genes in cells across their expression domains. Introducing a heterologous promoter into a BMP target gene changes burst amplitude but not promoter occupancy suggesting that, while the promoter sequence controls amplitude, occupancy depends on the amount of BMP signal decoded by the enhancer. We provide evidence that graded mRNA output is a general feature of morphogen gradient interpretation and discuss how this can impact on cell fate decisions.

INTRODUCTION

A gradient of Bone Morphogenetic Protein (BMP) signalling patterns ectodermal cell fates along the dorsal-ventral axis of vertebrate and invertebrate embryos (Bier and De Robertis, 2015; Hamaratoglu et al., 2014). In *Drosophila*, visualisation of Decapentaplegic (Dpp), the major BMP signalling molecule, reveals a shallow graded distribution in early embryos that subsequently refines to a peak of Dpp at the dorsal midline (Shimmi et al., 2005; Wang and Ferguson, 2005). BMP-receptor activation leads to phosphorylation of the Mad transcription factor, which associates with Medea (Med) to activate or repress target gene transcription (Hamaratoglu et al., 2014). A stripe of phosphorylated Mad (pMad) and Med centred at the dorsal midline has been visualised in the early *Drosophila* embryo (Dorfman and Shilo, 2001; Rushlow et al., 2001; Sutherland et al., 2003), similar to that observed for Dpp (Shimmi et al., 2005; Wang and Ferguson, 2005), although lower pMad levels are also detectable in a few adjacent dorsal-lateral cells (Rushlow et al., 2001). The BMP/pMad gradient activates different thresholds of gene activity, including the peak target genes *Race* and *hindsight* (*hnt*) and intermediate targets *u-shaped* (*ush*) and *tailup* (*tup*) (Ashe et al., 2000).

New insights into transcriptional activation have been obtained by studying this process in single cells using quantitative and live imaging approaches, including single molecule FISH (smFISH) and the MS2/MCP system (Pichon et al., 2018). The latter allowed the first direct visualisation of pulses or bursts of transcriptional activity (Chubb et al., 2006; Golding et al., 2005). Enhancers have been shown to regulate the frequency of transcriptional bursts, with strong enhancers generating more bursts than weaker enhancers (Fukaya et al., 2016; Larson et al., 2013; Larsson et al., 2019; Senecal et al., 2014). In addition, the detection of simultaneous bursts of transcription of two linked reporters by a single enhancer argues against the classic enhancer-promoter looping model (Fukaya et al., 2016).

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Based on the simultaneous activation of more than one promoter by an enhancer and the behaviour of super enhancers, a new model of transcriptional activation has been proposed, which invokes compartmentalisation of transcription factors, coregulators and Pol II in dynamic phase separated condensates (Hnisz et al., 2017). Intrinsically disordered regions in transcription factors and coactivators, including subunits of the Mediator complex and the chromatin reader BRD4, promote formation of hubs or condensates at genomic loci, which concentrate Pol II to promote activation (Boija et al., 2018; Cho et al., 2018; Chong et al., 2018; Sabari et al., 2018).

To provide insight into morphogen gradient interpretation at single cell resolution, we have used live imaging and quantitative analysis to determine the kinetics of endogenous target gene activation in response to the BMP gradient in the *Drosophila* embryo. These data reveal that BMP signalling modulates the fraction of time the promoter of target genes is active. Mechanistically, we provide evidence that the enhancer decodes the BMP signal to regulate the rate the promoter switches on, regardless of the promoter sequence present. In contrast, the promoter predominantly regulates burst amplitude. Overall these data reveal how a signalling gradient is decoded with different transcriptional kinetics to impart positional information on cells.

RESULTS

Monoallelic transcription and graded mRNA outputs in response to the BMP gradient

In order to visualise the transcriptional activity of Dpp target genes in the early *Drosophila* embryo we first used nascent FISH. While the classic expression patterns (Ashe et al., 2000) are detected for these genes (Fig. S1A), for a proportion of nuclei, the Dpp target gene is only transcribed by a single allele. To facilitate visualisation of the number of active alleles in each nucleus within the expression domain, the nuclei were false coloured based on allelic activity (Fig. 1Ai). Higher magnification images of a subset of nuclei expressing both alleles (biallelic) or only a single allele are shown in Fig. 1Aii. We refer to the latter nuclei as monoallelic, meaning that only one allele is active rather than one being stably inactivated, as observed in imprinting for example (Khamlichi and Feil, 2018). Quantitation shows that around one quarter of active nuclei are monoallelic for the four tested Dpp target genes (Fig. 1B). The false coloured images reveal that the monoallelic nuclei are predominantly localised around the edge of the expression domain (Fig. 1Ai). Consistent with this, quantitation shows that monoallelic nuclei are located significantly further from the midline of the expression domain compared to those nuclei transcribing both alleles (Fig. 1C). This distribution suggests that monoallelic transcription is a consequence of limiting activator levels.

As the FISH data detected differences in the number of active alleles within nuclei across the gene expression domains, we next addressed how this affects mRNA number in individual cells. To this end, we used smFISH with *ush* exonic probes and single molecule inexpensive FISH (smiFISH) (Tsanov et al., 2016) with *ush* intronic probes to quantify mRNA number and visualise transcription foci, respectively (Fig. 1D). *ush* first becomes transcribed in nuclear cleavage cycle 14

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(nc14) with the number of transcripts per cell increasing with age (Fig. 1Ei). Analysis reveals that the proportion of monoallelic nuclei is highest when the gene is first switched on and then decreases (Fig. S1B), consistent with the proportion observed using FISH (Fig. 1B). Cells with 2 active alleles have a higher mRNA number than cells showing monoallelic transcription (Fig. 1Eii). However, the maximum number of mRNAs in biallelic cells is less than double that detected in monoallelic cells suggesting the latter may have been transcribing both alleles at an earlier time. A low number of mRNAs is also detected in cells without an active allele, also consistent with earlier transcription of at least one allele (Fig. 1Eii). Visualisation of *ush* mRNA number per cell based on position in early, mid and late nc14 embryos reveals that there is a mRNA gradient similar to that of Dpp, with highest levels at the dorsal midline that diminish in more dorsolateral cells (Fig. 1F). In late nc14 embryos there is a ~10-fold difference in mRNA number per cell between cells located at the centre and edges of the expression domain (Fig. 1F).

Analysis of *hnt* and *tup* smFISH data also reveal that the mRNA number per cell increases with developmental age while the proportion of monoallelic cells decreases (Fig. S1C-F). In both cases a gradient of mRNA is detected across the expression domain, again with large differences in transcript number per cell at positions near the middle or edge of the expression domain (Fig. S1D, F). Visualisation of *ush* and *tup* transcript numbers across the expression domain by individual cell widths mirrored at the midline reveals that the mRNA number per cell is similar for the first 4 cells on either side of the midline and then declines (Fig. S1Gi). These data show that >60% of the total *ush* or *tup* mRNAs in the expression domain are transcribed by these 8 central cells, even though they represent less than one third of the expression domain (Fig. S1Gii). It has been shown previously that the early peak of pMad in stage 5 embryos is 8-10 cells wide (Dorfman and Shilo, 2001; Mizutani et al., 2005) (see Discussion). Together these data show that there is a mRNA gradient of Dpp target genes in the dorsal ectoderm that reflects the Dpp gradient.

Next we tested the hypothesis that nuclei at the edge of the expression domain can only activate one allele due to limiting levels of Dpp signalling and therefore pMad activator. We increased Dpp levels by introducing a transgene with *dpp* under the control of the *even-skipped* stripe 2 enhancer (st2-*dpp*) (Ashe et al., 2000) and visualised transcription foci using smFISH (Fig. 1G). The proportion of monoallelic nuclei located in a region equivalent to the edge of the wildtype (wt) *ush* expression domain was determined (Fig. 1G, Hi). These data show that there are significantly less monoallelic *ush* nuclei compared to the same region of a wt embryo (Fig. 1Hii). This supports the idea that the failure of nuclei on the edge of the expression domain to activate both alleles is due to limiting Dpp/pMad levels.

Monoallelic transcription is a general feature of gradient interpretation

To determine if monoallelic transcription is a general feature of gradient activation, we analysed *snail* (*sna*), *short gastrulation* (*sog*) and *brinker* (*brk*), which are target genes of the Dorsal gradient

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Reeves and Stathopoulos, 2009). These genes also show monoallelic transcription (Fig. 2Ai-ii), in around 25% of nuclei within the expression domain (Fig. 2B). Monoallelic nuclei are predominantly located at the edges of the expression domain (Fig. 2C), although this is less pronounced for *sna* transcription, potentially due to Sna auto-repression (Boettiger and Levine, 2013) (see Discussion).

For *sog* and *brk*, the ventral border of the expression pattern is established by Sna repression, whereas the dorsal border is due to limiting Dorsal (Reeves and Stathopoulos, 2009). Therefore, given the above data that suggest monoallelic transcription reflects low activator levels, we predict that there would be more monoallelic transcription on the dorsal edge of the expression domain. Quantitation of the number of monoallelic nuclei on the dorsal and ventral sides separately reveals that there is a significantly higher proportion on the dorsal side of the *sog* and *brk* expression domains (Fig. 2D). In contrast, there is no significant difference between the two edges of the symmetric *sna* or Dpp target gene expression domains in terms of the relative percentage of monoallelic nuclei (Fig. 2E). These data are consistent with some nuclei activating only a single allele, depending on their position with respect to the gradient, due to limiting activator. The presence of monoallelic expression on the ventral side of *sog* and *brk* likely reflects asynchronous repression of each allele (see Discussion).

Temporal dynamics of transcriptional activation in response to the BMP gradient

To complement the above snapshot data, we used the MS2 system (Garcia et al., 2013; Lucas et al., 2013) to visualise the temporal dynamics of BMP gradient interpretation during early embryogenesis. We used CRISPR genome engineering to introduce 24 copies of the MS2 stem loops into the endogenous 5' UTR of the *ush* and *hnt* genes (Fig. 3A). Conventional in situ hybridisation showed *ush* and *hnt* expression patterns equivalent to those observed in wt embryos (Fig. S2A), indicating that insertion of the loops does not affect the expression patterns. To visualise transcription dynamics, females maternally expressing one copy of MCP-GFP and Histone-RFP were crossed to males carrying the *ush* or *hnt* gene with MS2 stem loops, so the resulting embryos have a single allele carrying the MS2 sequence. Confocal imaging of these embryos allows the bright fluorescent signal associated with the nascent transcription site to be recorded, as a measure of transcriptional activity, for each expressing nucleus (Fig. 3B).

Embryos were imaged prior to the onset of nc14 to allow accurate timing of the initial activation of *ush* and *hnt* relative to the start of nc14 (Video S1 and S2 for *ush* and *hnt* transcription, respectively). We imaged the bulk of the expression domain for *ush*, whereas for *hnt* we imaged the central and posterior part; active nuclei are false coloured in a still from the video (Fig. S2Bi). As the *ush* expression domain is largely uniform along the anterior-posterior (AP) axis, we have focused on the anterior part for subsequent analysis (Fig. S2Bii). *hnt* expression is more modulated along the AP axis (Ashe et al. 2000), therefore we have analysed nuclei in the central region (Fig. S2Bii), corresponding to the presumptive amnioserosa.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

To measure the transcriptional activity of the expression domain, the mean fluorescence was analysed at each time point during nc14, showing that *hnt* has lower transcriptional activity than *ush* (Fig. 3C, Fig. S2C). Both the transcription onset time, based on the first time a fluorescent signal is detected, and the time taken to reach maximal transcriptional activity are delayed for *hnt* relative to *ush* (Fig. 3C, S2C-D). The transcription onset times for *ush* and *hnt* in each nucleus relative to its AP or dorsal-ventral (DV) position show little modulation along the AP axis (Fig. S2E, F). However, the onset times of *ush* and to a lesser extent that of *hnt* expression are delayed in nuclei further from the dorsal midline (Fig. S2E-F). The sum fluorescence of a nucleus, representing the total amount of transcriptional activity, is found to be highest in nuclei closer to the dorsal midline, experiencing peak BMP signalling levels, for both *ush* and *hnt* (Fig. 3D). Resolving the differences in *ush* and *hnt* transcriptional activity further, based on nuclear position, reveals that it is highest in nuclei at the dorsal midline at all time points, then reduces in nuclei towards the edges of the expression domain (Fig. S2G).

As the fluorescence signals for *hnt* and *ush* vary between expressing nuclei, we performed K-means clustering analysis based on all expressing nuclei. For a representative *ush* embryo these data show that the nuclei partition into 3 clusters, which broadly map to the centre, intermediate area and edges of the expression domain (Fig. 3E). Visualisation of the individual fluorescent traces for all nuclei within these clusters as heatmaps shows that nuclei from the middle of the expression domain have a faster onset time and higher fluorescence output than nuclei in the intermediate region with a further reduction in cells at the edge of the expression domain (Fig. 3E). Similar findings are obtained for *hnt*, with the nuclei partitioning into 2 clusters broadly based on their position. Central nuclei, receiving peak Dpp signalling, have faster onset times and higher fluorescent outputs than those on the edge (Fig. 3F). The low transcriptional activity at the edge of the *ush* and *hnt* expression domains observed with the MS2 system (Fig. 3E, F) is consistent with the reduced mRNA numbers detected in these cells by smFISH (Fig 1F, S1D).

Different BMP signalling levels alter transcriptional burst kinetics

Given the different transcriptional behaviours of nuclei, we used a memory adjusted Hidden Markov Model to infer bursting parameters (Fig. 4A, S3Ai) from the transcriptional traces, based on a two state promoter model (Lammers et al., 2019) (Fig. S3Aii). Representative *ush* and *hnt* traces for nuclei from the centre of the expression domain receiving peak Dpp signalling and the inferred promoter states are shown in Fig. 4B-C, revealing different promoter activity profiles for the two Dpp target genes. Traces for nuclei at other positions in the expression domain are shown in Fig. S3B and C.

We used *ush* fluorescent traces from all nuclei within the centre, intermediate and edge clusters to infer the global kinetic parameters for each cluster. For *hnt* we separated cells into 3 clusters to better understand the transcriptional response to differing levels of Dpp signalling. As

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the *ush* clusters are largely partitioned on expression level, we used mean expression to separate *hnt* expressing cells into 3 clusters. For both *ush* and *hnt*, decreasing levels of Dpp signalling between the centre, intermediate and edge clusters is associated with reduced promoter occupancy, equivalent to the fraction of time the promoter is active, k_{on} and burst frequency (Fig. 4D, E). The reduction in k_{on} indicates that the promoter off period (1/ k_{on}) increases (Fig. S3D, E). In contrast, there is no statistical difference in k_{off} and the linked duration of promoter activity (1/ k_{off}) between the centre and intermediate *ush* and *hnt* clusters (Fig. 4D, E, S3D, E). While k_{off} is unchanged for the edge *hnt* cluster, an increase is observed in edge nuclei for *ush*, consistent with a reduced burst duration in the presence of very low Dpp signalling levels (Fig. 4D, S3D). For both *ush* and *hnt*, burst size and the Pol II initiation rate, k_{ini} (hereafter referred to as amplitude) also decrease as signalling levels are reduced (Fig. 4D, E). Based on the *ush* and *hnt* parameters the theoretical burst profiles of nuclei receiving peak Dpp signalling can be compared (Fig. 4F). These show that while *ush* is transcribed in relatively low amplitude, long duration bursts, *hnt* exhibits high amplitude bursts of very high frequency and short duration (Fig. 4F).

Dpp concentration determines promoter occupancy

As many burst parameters change in response to different levels of Dpp signalling, we next addressed which parameter is the major determinant of the transcriptional response. To this end, we inferred burst parameters at single cell resolution and determined the degree of correlation with the mean fluorescence intensity for each nucleus expressing *ush* (Fig. 5A). Promoter occupancy shows the highest correlation with the mean fluorescence intensity (expression level), such that it almost perfectly predicts the expression level of every active nucleus (Fig. 5B). k_{on} is also strongly correlated, more so than k_{off} (Fig. 5C, D), suggesting that promoter occupancy predicts expression, predominantly through changes in k_{on}. Consistent with this, burst frequency and amplitude show weaker correlations with mean expression (Fig. 5E, F). Similar findings are obtained for *hnt* bursting parameters at single cell resolution, with promoter occupancy most correlated with mean fluorescence, followed by k_{on} and amplitude, whereas k_{off} is poorly correlated (Fig. S4A-F).

To further address how BMP signalling affects transcriptional bursting, we imaged *ush* transcription in the presence of ectopic signalling by introducing a single copy of the st2-*dpp* transgene (Ashe et al., 2000) (Video S3). For the analysis we focused on cells in the region where st2-*dpp* is expressed (Fig. 6Ai). The ectopic *dpp* results in an expanded *ush* expression pattern (Fig. 6Ai) with higher total fluorescence signals detected compared to wt (Fig. 6Aii compare to 3D). The *ush* transcription onset time is slightly earlier in the presence of st2-*dpp* (Fig. 6B) and the mean fluorescence is increased, although the time at which maximum fluorescence is reached is similar to wt (Fig. 6C). We next used the memory adjusted Hidden Markov Model to infer burst parameters, after dividing the expression domain into 3 regions based on expression level. These regions are broadly similar to those in wt embryos as, although the *ush* expression domain is

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

broader in st2-*dpp* embryos, we have focused our analysis on a region that is only around 3 cells wider on each edge. A representative trace for the centre cluster is shown in Fig. 6D. The global parameters reveal that in the intermediate regions promoter occupancy and k_{on} are increased relative to wt (Fig. 6E). No change of promoter occupancy in the centre nuclei, and an increase to this level in intermediate nuclei, suggests that occupancy is already close to saturation in wt cells receiving the highest Dpp signalling. Frequency and k_{off} show no change in centre and intermediate nuclei, although both respond to higher Dpp in edge nuclei that normally receive very low Dpp (Fig. 6E). These data suggest that promoter occupancy and not frequency predominantly integrates higher levels of BMP signalling. In addition, st2-*dpp* increases *ush* burst amplitude and therefore burst size (Fig. 6E), consistent with amplitude being responsive to Dpp levels. Together, these data are consistent with the analysis of *ush* and *hnt* in wt embryos and further support the conclusion that Dpp signalling promotes higher promoter occupancy, predominantly through increasing k_{on}, and to a lesser extent amplitude to generate a stronger transcriptional response.

The enhancer decodes the BMP signal to regulate promoter occupancy

As the above data suggest that BMP signalling level predominantly regulates promoter occupancy, we next addressed the role of the promoter in the transcriptional response by replacing the *ush* promoter with that of *hnt* in the endogenous locus (*hnt>ush*) (Fig. 7A). This line also contains 24 copies of the MS2 stem loops in the *ush* 5'UTR as described above so that the effect of changing the promoter on burst kinetics can be determined. Analysis of the fluorescent signals for *hnt>ush* (Video S4) reveals that the cumulative expression pattern, comprised of every cell that activates transcription at one or more time points, is similar but slightly narrower compared to wt *ush* (Fig. S5A). The times of transcription onset and at which maximum fluorescence is reached for *hnt>ush* are equivalent to those observed for *ush* (Fig. 7B-C, Fig. S5B). As *hnt* has a later onset time than *ush* (Fig. S2D) and changing the *ush* promoter to that of *hnt* has no effect on onset time (Fig. 7B), this suggests that onset time is largely dictated by the enhancer, with only fine-tuning by the promoter. It is also evident from the 3 *hnt>ush* biological replicates that introducing a heterologous promoter increases variation in the fluorescent signals (Fig. S5B).

Clustering of the cells in the *hnt>ush* expression domain and analysis of the fluorescent traces reveals that cells in each *hnt>ush* cluster have higher fluorescence compared to wt (Fig. S5C). We next used these clusters to infer global bursting parameters from the model. A representative trace for each cluster is shown in Fig. 7D and S5D. The global parameters show that amplitude and therefore burst size are significantly higher for *hnt>ush* embryos relative to wt (Fig. 7E). In contrast, there is no significant change in promoter occupancy, k_{on}, k_{off} or frequency (Fig. 7E). This suggests that the promoter predominantly regulates burst amplitude, whereas promoter occupancy is not determined by the actual promoter sequence itself. Given the data above that promoter occupancy is established by the level of BMP signalling, the simplest

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

interpretation is that promoter occupancy is dictated by the enhancer, depending on the amount of signal/activator, regardless of the promoter present.

Using the *hnt>ush* parameters, simulation of the burst profile in the centre region shows that the *hnt>ush* traces represent a hybrid profile between that of the short duration, high amplitude *hnt* traces and those of *ush* that are longer and lower amplitude (Fig. 7F). Together, these data suggest that the enhancer controls promoter occupancy but the amplitude of the response depends on the nature of the promoter.

DISCUSSION

Here we analyse the transcriptional burst kinetics of the endogenous *hnt* and *ush* genes at single cell resolution and show that cells interpret different levels of BMP signalling by modulating promoter occupancy, predominantly through altering k_{on}. *hnt* transcription occurs in very short bursts with high frequency and amplitude, whereas *ush* bursts are less frequent but longer duration (~10 fold longer than *hnt* for cells at the midline). *hnt* shows much lower promoter occupancy than *ush*, providing a molecular explanation for the observed threshold responses of these genes to the BMP gradient (Fig. 7G). Our data indicate that *hnt* requires high BMP signalling for its activation, as lower signalling levels are insufficient to maintain the promoter in an active state, resulting in a narrow expression pattern. In contrast, low signalling levels allow sufficient promoter occupancy for *ush*, which therefore has a broader expression pattern. We conclude that k_{on} and promoter occupancy, which are unchanged when the heterologous *hnt* promoter is tested, are dictated by features of the enhancer and dependent on the level of signal received. This is consistent with other studies that have found the enhancer to regulate k_{on} (Fukaya et al., 2016; Lammers et al., 2019; Larson et al., 2013; Larsson et al., 2019; Senecal et al., 2014). Our promoter swap data suggest that the promoter regulates burst amplitude. The *hnt* promoter is associated with a higher initiation rate than the *ush* promoter, and insertion of the *hnt* promoter into the *ush* locus increases burst amplitude. This may relate to the presence of a TATA box in the *hnt* promoter as TATA has been linked to high initiation rates previously (Corrigan et al., 2016), whereas the *ush* promoter has an initiator but lacks a TATA box. However, other differences between the *ush* and *hnt* promoters also exist, including that the *hnt* promoter has a higher degree of Pol II promoter proximal pausing than *ush* (Saunders et al., 2013). Therefore, further studies are required to determine the contribution of different promoter features to burst amplitude.

The lack of a contribution of burst duration (1/ k_{off}) to decoding BMP signalling is in stark contrast to the interpretation of Notch signalling in *Drosophila* and *C. elegans*, whereby Notch alters the duration, but not frequency, of transcription bursts (Falo-Sanjuan et al., 2019; Lee et al., 2019). Increasing gene expression through high k_{on} rates can decrease the noise level, whereas lengthening burst duration is associated with more noise (Wong et al., 2018). In addition, regulation of burst frequency may allow genes to respond with more sensitivity to activator

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

concentration than when burst duration is modulated (Li et al., 2018). Therefore, perhaps regulation of BMP target genes by promoter occupancy, via k_{on}, has the advantage of allowing more sensitive regulation with less noise. Our findings for decoding BMP signalling are similar to the strategy described for modulation of gap gene transcription during AP patterning, where the key regulatory parameter is also the fraction of time the promoter is active (Zoller et al., 2018). It remains to be determined whether other signals will be interpreted through changes in promoter occupancy or duration.

The phase separation model of transcriptional control proposes that transcription factors, Mediator and other coactivators form dynamic condensates associated with activation (Hnisz et al., 2017). The Smads interact with Mediator subunits (Zhao et al., 2013) and Smad3 can form condensates in vitro and in cells (Zamudio et al., 2019). The CBP histone acetyltransferase is a Smad transcriptional coactivator (Ashe et al., 2000; Waltzer and Bienz, 1999) and modification of transcription regulators, including by acetylation, has been implicated in formation of phase-separated transcription condensates (Hnisz et al., 2017). Therefore, based on these data, it is likely that pMad-Medea, CBP and Mediator form a transcription hub that allows gene activation. Live imaging has provided evidence for groups of closely spaced Pol II, referred to as convoys, which elongate along a gene together. Knockdown of a Mediator subunit reduced the promoter on time, lowered the number of Pol II molecules in the convoy and increased spacing between them, suggesting that Mediator is important for quick succession of initiation events (Tantale et al., 2016). Therefore, we suggest that the higher pMad levels associated with increased BMP signalling will recruit more Mediator, resulting in the target promoter being active for longer and a larger Pol II convoy, explaining the effect of BMP signalling on promoter occupancy and amplitude.

The different burst kinetics of BMP target gene transcription in cells within the expression domain provides an explanation for the observed monoallelic expression (Fig. 7G). Cells on the edge of the expression domain have low burst frequency and duration, resulting in typically only one allele being active. Similarly, stochastic transcriptional bursting events from one allele have been suggested to explain rare cases of random monoallelic expression observed for less than 1% of genes in mouse fibroblasts and human $CD8^+$ T cells (Reinius and Sandberg, 2015), with supporting evidence for this obtained for poorly expressed genes in the mouse kidney (Symmons et al., 2019). Our study highlights how a gene can show monoallelic or biallelic expression within the same expression domain, depending on cellular position with respect to graded signalling levels. Monoallelic transcription has also been reported for zygotic *hunchback* (*hb*) transcription, which is activated by the Bicoid gradient, particularly at the anterior tip and posterior border of the expression domain (Lucas et al., 2013; Porcher et al., 2010). As we also detect one active allele of Dorsal target genes in some cells, we suggest that monoallelic transcription with a concomitant reduction in mRNA number, is a general feature of gradient interpretation for cells receiving low signal.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sna transcription, however, differs from that of the other Dorsal targets *brk* and *sog* as we detect monoallelic *sna* nuclei more evenly distributed throughout the expression domain. There is unusual homogeneity in the number of *sna* mRNAs in each cell, due to a rapid transcription rate and autorepression (Boettiger and Levine, 2013). Allele by allele repression has been observed in the *Drosophila* embryo, potentially because repressors are better able to act in the refractory period following a burst (Esposito et al., 2016). Therefore, the more intermingled appearance of monoallelic *sna* nuclei that we observe can be explained by Sna autorepression silencing one allele at a time, as repression occurs in the refractory period between bursts that are not entirely synchronous between the two alleles. Similarly, allele by allele repression can also explain why monallelic nuclei are observed at the ventral borders of the *brk* and *sog* expression domains, where levels of the Dorsal activator are high.

The number of *ush* and *tup* mRNAs per cell is relatively constant in cells within the first 8 rows centred on the dorsal midline, but then sharply declines. As a result, for *ush* and *tup,* >60% of the total transcripts in the expression domain are synthesised by the dorsal most 8 cells, despite these cells only constituting around one third of the expression pattern. This mRNA distribution reflects the spatial BMP gradient as the peak of pMad is 8-10 cells initially then refines to 6 cells wide (Dorfman and Shilo, 2001; Mizutani et al., 2005). Moreover, modelling suggests that the concentration of BMP bound receptor complexes at the dorsal midline doubles between 20 min and 30 min into nc14 (Mizutani et al., 2005; Umulis et al., 2006). These times correspond to the onset times of *ush* and *hnt*, respectively, suggesting that *ush* transcription can respond to the initial low levels of signalling, whereas the peak threshold *hnt* requires more activated receptors. Furthermore, BMP-receptor levels peak at ~40 min into nc14 (Umulis et al., 2006), which coincides with the observed maximum fluorescence output we detect for *ush* and *hnt* (means of 41 and 46 min, respectively).

Based on our data, we suggest a threshold model of cell fate whereby cells on the edge of the expression domain synthesise sufficient mRNAs to adopt a particular cell fate, whereas cells in the centre would have a surplus of transcripts. In this model, the difference in mRNA numbers in cells across the expression domain can explain the lack of robustness when shadow enhancers are deleted (Antosova et al., 2016; Frankel et al., 2010; Perry et al., 2010). Perturbation of the system, such as removal of a shadow enhancer, would lead to a further reduction in mRNA number per cell so that those on the edge would only just exceed the threshold level. Another challenge, such as high temperature or reduced activator level, would further decrease the transcriptional output such that there are insufficient mRNAs to specify the correct cell fate. It will be interesting in the future to test how the different numbers of mRNAs per cell from key BMP target genes impact on the robustness of dorsal ectoderm cell fate decisions.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ACKNOWLEDGEMENTS

We thank Lauren Forbes-Beadle for helpful discussions, the Bloomington *Drosophila* Stock Center for flies, the Vincent lab for plasmids, the Ronshaugen lab for reagents, the Cambridge Fly Facility for microinjections, Peter March and Egor Zindy for image analysis advice and the University of Manchester Bioimaging Facility for support. This project was supported by a Wellcome Trust Investigator Award to H.L.A. and M.R. (204832/Z/16/Z) and Wellcome Trust PhD studentships to C.H. (205975/Z/17/Z), J.B. (215187/Z/19/Z) and T.G.M. (110566/Z/15/Z).

AUTHOR CONTRIBUTIONS

Conceptualization: C.H., J.B., M.R. and H.L.A.; Investigation: C.H, C.S. and P.U.; Software: J.B. and T.G.M.; Writing-original draft: C.H. and H.L.A.; Writing – Reviews & Editing: J.B, T.G.M, C.S., P.U. and M.R.; Supervision: M.R. and H.L.A.; Funding Acquisition: M.R. and H.L.A.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Antosova, B., Smolikova, J., Klimova, L., Lachova, J., Bendova, M., Kozmikova, I., Machon, O., and Kozmik, Z. (2016). The Gene Regulatory Network of Lens Induction Is Wired through Meis-Dependent Shadow Enhancers of Pax6. PLoS Genet *12*, e1006441.

Ashe, H.L., Mannervik, M., and Levine, M. (2000). Dpp signaling thresholds in the dorsal ectoderm of the Drosophila embryo. Development *127*, 3305-3312.

Bier, E., and De Robertis, E.M. (2015). EMBRYO DEVELOPMENT. BMP gradients: A paradigm for morphogen-mediated developmental patterning. Science *348*, aaa5838.

Boettiger, A.N., and Levine, M. (2013). Rapid transcription fosters coordinate snail expression in the Drosophila embryo. Cell Rep *3*, 8-15.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., *et al.* (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855 e1816.

Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, II (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412-415.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., *et al.* (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science *361*.

Chubb, J.R., Trcek, T., Shenoy, S.M., and Singer, R.H. (2006). Transcriptional pulsing of a developmental gene. Curr Biol *16*, 1018-1025.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Corrigan, A.M., Tunnacliffe, E., Cannon, D., and Chubb, J.R. (2016). A continuum model of transcriptional bursting. Elife *5*.

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

Esposito, E., Lim, B., Guessous, G., Falahati, H., and Levine, M. (2016). Mitosis-associated repression in development. Genes Dev *30*, 1503-1508.

Falo-Sanjuan, J., Lammers, N.C., Garcia, H.G., and Bray, S.J. (2019). Enhancer Priming Enables Fast and Sustained Transcriptional Responses to Notch Signaling. Dev Cell *50*, 411-425.

Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature *466*, 490-493.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358-368.

Garcia, H.G., Tikhonov, M., Lin, A., and Gregor, T. (2013). Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr Biol *23*, 2140-2145.

Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. Cell *123*, 1025-1036.

Hamaratoglu, F., Affolter, M., and Pyrowolakis, G. (2014). Dpp/BMP signaling in flies: from molecules to biology. Semin Cell Dev Biol *32*, 128-136.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13-23.

Khamlichi, A.A., and Feil, R. (2018). Parallels between Mammalian Mechanisms of Monoallelic Gene Expression. Trends Genet *34*, 954-971.

Lammers, N.C., Galstyan, V., Reimer, A., Medin, S.A., Wiggins, C.H., and Garcia, H.G. (2019). Multimodal transcriptional control of pattern formation in embryonic development. biorxiv.

Larson, D.R., Fritzsch, C., Sun, L., Meng, X., Lawrence, D.S., and Singer, R.H. (2013). Direct observation of frequency modulated transcription in single cells using light activation. Elife *2*, e00750.

Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, A., Rivera, C.M., Ren, B., and Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. Nature *565*, 251-254.

Lee, C., Shin, H., and Kimble, J. (2019). Dynamics of Notch-Dependent Transcriptional Bursting in Its Native Context. Dev Cell *50*, 426-435.

Li, C., Cesbron, F., Oehler, M., Brunner, M., and Hofer, T. (2018). Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation. Cell Syst *6*, 409-423 e411.

Lucas, T., Ferraro, T., Roelens, B., De Las Heras Chanes, J., Walczak, A.M., Coppey, M., and Dostatni, N. (2013). Live imaging of bicoid-dependent transcription in Drosophila embryos. Curr Biol *23*, 2135-2139.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mizutani, C.M., Nie, Q., Wan, F.Y., Zhang, Y.T., Vilmos, P., Sousa-Neves, R., Bier, E., Marsh, J.L., and Lander, A.D. (2005). Formation of the BMP activity gradient in the Drosophila embryo. Dev Cell *8*, 915-924.

Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. Curr Biol *20*, 1562-1567.

Pichon, X., Lagha, M., Mueller, F., and Bertrand, E. (2018). A Growing Toolbox to Image Gene Expression in Single Cells: Sensitive Approaches for Demanding Challenges. Mol Cell *71*, 468-480.

Porcher, A., Abu-Arish, A., Huart, S., Roelens, B., Fradin, C., and Dostatni, N. (2010). The time to measure positional information: maternal hunchback is required for the synchrony of the Bicoid transcriptional response at the onset of zygotic transcription. Development *137*, 2795-2804.

Reeves, G.T., and Stathopoulos, A. (2009). Graded dorsal and differential gene regulation in the Drosophila embryo. Cold Spring Harb Perspect Biol *1*, a000836.

Reinius, B., and Sandberg, R. (2015). Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. Nat Rev Genet *16*, 653-664.

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.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C*., et al.* (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*.

Saunders, A., Core, L.J., Sutcliffe, C., Lis, J.T., and Ashe, H.L. (2013). Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. Genes Dev *27*, 1146-1158.

Senecal, A., Munsky, B., Proux, F., Ly, N., Braye, F.E., Zimmer, C., Mueller, F., and Darzacq, X. (2014). Transcription factors modulate c-Fos transcriptional bursts. Cell Rep *8*, 75-83.

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

Sutherland, D.J., Li, M., 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.

Symmons, O., Chang, M., Mellis, I.A., Kalish, J.M., Park, J., Susztak, K., Bartolomei, M.S., and Raj, A. (2019). Allele-specific RNA imaging shows that allelic imbalances can arise in tissues through transcriptional bursting. PLoS Genet *15*, e1007874.

Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J.M., Robert, M.C., Capozi, S., Chouaib, R., Backer, V., Mateos-Langerak, J.*, et al.* (2016). A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. Nat Commun *7*, 12248.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

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.

Waltzer, L., and Bienz, M. (1999). A function of CBP as a transcriptional co-activator during Dpp signalling. EMBO J *18*, 1630-1641.

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

Wong, V.C., Bass, V.L., Bullock, M.E., Chavali, A.K., Lee, R.E.C., Mothes, W., Gaudet, S., and Miller-Jensen, K. (2018). NF-kappaB-Chromatin Interactions Drive Diverse Phenotypes by Modulating Transcriptional Noise. Cell Rep *22*, 585-599.

Zamudio, A.V., Dall'Agnese, A., Henninger, J.E., Manteiga, J.C., Afeyan, L.K., Hannett, N.M., Coffey, E.L., Li, C.H., Oksuz, O., Sabari, B.R.*, et al.* (2019). Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. Mol Cell.

Zhao, M., Yang, X., Fu, Y., Wang, H., Ning, Y., Yan, J., Chen, Y.G., and Wang, G. (2013). Mediator MED15 modulates transforming growth factor beta (TGFbeta)/Smad signaling and breast cancer cell metastasis. J Mol Cell Biol *5*, 57-60.

Zoller, B., Little, S.C., and Gregor, T. (2018). Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting. Cell *175*, 835-847 e825.

MAIN FIGURE LEGENDS

Figure 1: BMP target genes show monoallelic transcription and graded mRNA outputs

(A) Expression patterns of BMP target genes *Race*, *hnt*, *ush* and *tup* in *Drosophila* embryos. Nuclei are false coloured according to the number of active transcription foci based on FISH images (Figure S1A). A line of best fit shows the midline of the expression domain (i).

Enlarged regions from FISH images (full images in Figure S1A) show monoallelic (yellow outline) and biallelic (pink outline) transcription of BMP target genes (ii).

(B) Proportion of nuclei displaying monoallelic or biallelic transcription within the expression domain (n = 3 embryos).

(C) The median distance of monoallelic or biallelic nuclei to the middle of the expression domain (n = 3 embryos).

(D) Detection of nascent transcription sites (smiFISH, magenta) and single *ush* mRNAs (smFISH, green).

(E) Number of mRNAs per cell in single embryos of increasing age in nc14 (i) with quantitation

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

based on the number of active transcription foci shown for the oldest embryo (ii). Spearman correlation coefficient is shown for each pair of variables.

(F) The number of *ush* mRNAs per cell plotted according to the nuclear distance from the dorsal midline are shown at three time points (data from embryos 1, 4 and 6 in Ei). Embryos are ordered by increasing age and points are coloured based on the number of active transcription foci.

(G) smFISH image of the *ush* transcription domain in a *st2-dpp* embryo. The white rectangle highlights the analysis region; the red rectangles show the wt expression domain edge positions.

(H) Nuclei in the regions highlighted in (G) are coloured based on the number of active transcription foci (i) and quantitated (ii). Data for *st2-dpp* edges from Hi (n = 3) and for wt edges from data in F (n = 3).

All embryos are oriented dorsally with anterior to the left. Scale bar, 3 µm (Aii), 5 µm (D) or 20 µm (G). $^{*}p < 0.05$, $^{***}p < 0.001$, $^{****}p < 0.0001$. Mean ± SD (B, Hii), mean (C) or median ± 95% confidence intervals (Eii). Student's t-test (C), a Kruskal-Wallis test with a Dunn's multiple comparisons test (Eii) or Student's t-test between the proportions of monoallelic nuclei (Hii). **See also Figure S1.**

Figure 2: Monoallelic transcription is a general feature of gradient interpretation

(A) Representation of the expression patterns of DI target genes *sog*, *brk* and *sna* in *Drosophila* embryos, with nuclei false coloured according to the number of active transcription foci based on FISH images (i). A line of best fit shows the midline of the expression domain. Enlarged regions from the associated FISH images showing monoallelic (yellow outline) or biallelic (pink outline) transcribing nuclei (ii).

(B) Proportion of nuclei displaying monoallelic or biallelic transcription within the expression domain.

(C) As in (B) but distance to the middle of the expression domain is plotted.

(D) Proportion of monoallelic transcription in the dorsal (purple) versus the ventral (pink) half of the expression domain for each embryo analysed in (C).

(E) Analysis of monoallelic transcription of BMP target genes and *sna* between each half (green or blue) of the expression domain half. Data for BMP target genes are from Figure 1C.

Embryos in (Ai) are oriented anterior to the left and positioned laterally (*sog*, *brk*) or ventrally (*sna*). Biological replicates = 5 (*sog*), 4 (*brk*) and 3 (*sna*) embryos. Scale bar, 3 µm (Aii). $^{*}p < 0.05$, $^{***}p < 0.001$, $^{****}p < 0.0001$, ns = not significant. Mean ± SD (B) or mean (C). Student's t-test (C) or a paired Student's t-test (D, E).

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 3: Temporal dynamics of BMP target gene transcription

(A) Cartoon summarising the endogenous genomic imaging locus for *ush* (orange) and *hnt* (blue) with 24xMS2 loops inserted into the 5'UTR.

(B) Maximum projected still from Video S1 showing *ush* transcription false coloured based on fluorescence intensity. Enlarged region (bottom) shows the active transcription site in each nucleus.

(C) Mean fluorescence of *ush* and *hnt* transcription, with time of maximum fluorescence shown, for one representative embryo (209 (*ush*) and 192 (*hnt*) active nuclei). See Figure S2C for biological replicates.

(D) Schematics of the *ush* (anterior) and *hnt* (central) analysis domains (red boxes). Cumulative expression domains of representative embryos are coloured depending on the sum fluorescence produced by nuclei throughout nc14 (note different scales).

(E) Clustering analysis data shown for one representative embryo. (Left) Nuclei are coloured by cluster. (Middle) Heatmaps of single-cell traces, sorted according to transcription onset (scale as indicated, grey indicates periods where nuclei were not tracked). Time of transcriptional onset was traced to visualise onset fronts of different clusters and the position at which half the nuclei in a cluster initiated transcription is indicated (T_{50}). (Right) Mean fluorescence values of nuclei in each cluster. Number of data points given next to heatmaps.

(F) As in (E) showing single cell traces for *hnt* transcription profiles.

Embryos are oriented dorsally with anterior to the left. Scale bar, 10 µm (B, top) and 3 µm (B, bottom). Mean ± SEM (C) or median ± 95% confidence intervals (E, F right). **See also Figure S2.**

Figure 4: Different BMP signalling levels alter transcriptional burst kinetics

(A) Schematic of the transcription burst parameters and key parameter definitions.

(B, C) Representative fluorescence trace from a central cluster nucleus showing *ush* (B) and *hnt* (C) transcription and inferred promoter states.

(D, E) Global analysis of burst parameters for *ush* (D) and *hnt* (E) transcription in different spatial domains. Data points are coloured according to clusters.

(F) Bursting simulation of *ush* (orange) and *hnt* (blue) transcription based on mean burst parameter values from the central expression domains and shown for typical transcription periods of 30 min for *ush* and 20 min for *hnt*.

Mean ± SD (D, E) for n = 3 biological replicates. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$, ns = not significant. One-way ANOVA with a Dunnett's multiple comparisons test shows the difference to the central cluster (D, E). **See also Figure S3**.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 5: BMP signaling level is decoded by modulating promoter occupancy

(A) Mean fluorescence output of nuclei transcribing *ush* plotted according to the nuclear position across the dorsal midline for one representative embryo.

(B-F) Burst kinetic parameters of single *ush* nuclei plotted across the midline and against mean fluorescence intensity for promoter occupancy (B), k_{on} (C), k_{off} (D), burst frequency (E) and amplitude (F). Pearson correlation coefficient is shown for each pair of variables between burst parameter and mean fluorescence. Linear regression is shown ± 95% confidence intervals. **See also Figure S4.**

Figure 6: Increased BMP signalling levels increase promoter occupancy

(A) Schematic shows the imaging region (dotted line) and the analysis domain (red line). Cumulative expression domain of *ush* shows all active nuclei in one representative *st2-dpp* embryo (i) and coloured depending on the sum fluorescence produced by nuclei throughout nc14 (ii).

(B) Transcription onset times of *ush* shown for biological replicates with n = 3 in MS2 wt embryos (209, 186, 223 nuclei) and n = 3 in *st2-dpp* embryos (121, 89, 107 nuclei).

(C) Mean fluorescence of *ush* transcription over time in *st2-dpp* and wt MS2 embryos, with the indicated time point of maximum fluorescence. Data are from one representative wt MS2 embryo with 209 nuclei and biological replicates for *st2-dpp* embryos with 121, 89 and 107 active nuclei, respectively.

(D) Representative fluorescence trace from a central cluster nucleus showing *ush* transcription in a *st2-dpp* background and inferred promoter states.

(E) Global analysis of burst parameters for *ush* transcription in different spatial domains. Data points are coloured according to genotypes.

Embryos are oriented dorsally with anterior to the left. Mean ± SD (B, E) or mean ± SEM (C) for n = 3 biological replicates. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns = not significant. Student's t-test (B) and a two-way ANOVA with a Sidak's multiple comparisons test shows the difference between means of the two genotypes (E). Data for *ush* transcription in *ush* MS2 wt embryos were analysed in previous Figures and are shown here for comparison.

Figure 7: BMP responsive promoters integrate signalling levels through modulation of burst size.

(A) Summary of the endogenous genomic imaging loci for embryos with a *ush* promoter (orange) or the *hnt* promoter (green) and 24xMS2 loops present in the *ush* 5'UTR.

(B) Transcription onset times of *ush* transcription in *ush* and *hnt>ush* embryos. Median onset time of biological replicates are plotted with n = 3 for *ush* (209, 186, 223 nuclei) and *hnt>ush* (159, 187, 202 nuclei).

(C) Mean fluorescence of *ush* transcription and time point of maximum fluorescence in

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

representative embryos for n = 1 biological replicates with 209 (*ush*) and 186 (*hnt*>*ush*) active nuclei. See Figure S5B for biological replicates.

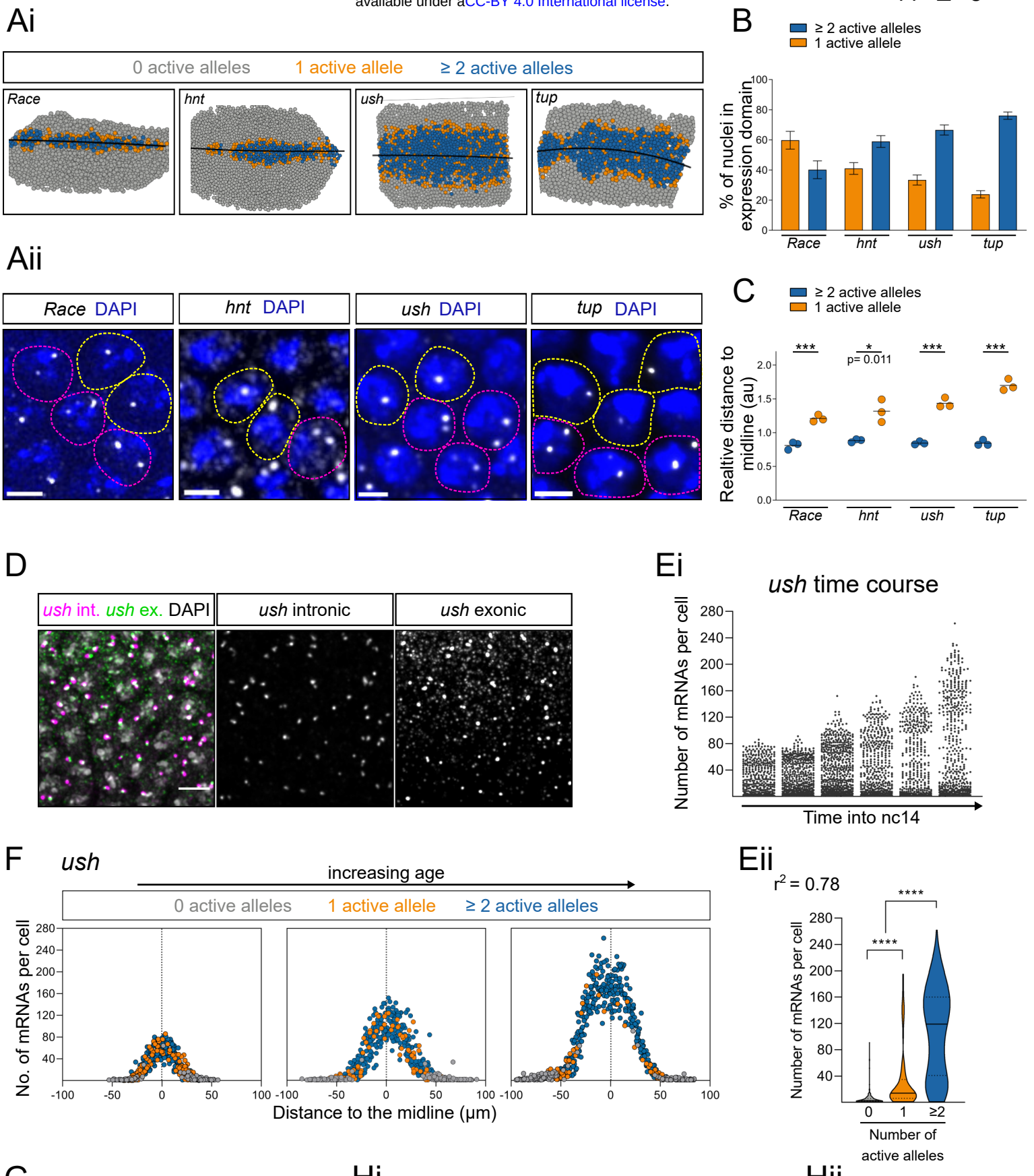
(D) Representative fluorescence trace from a central cluster nucleus showing *ush* transcription in a *hnt*>*ush* embryo and the inferred promoter states.

(E) Global analysis comparing burst parameters for *ush* transcription spatially and between genotypes. Data points are coloured according to genotype.

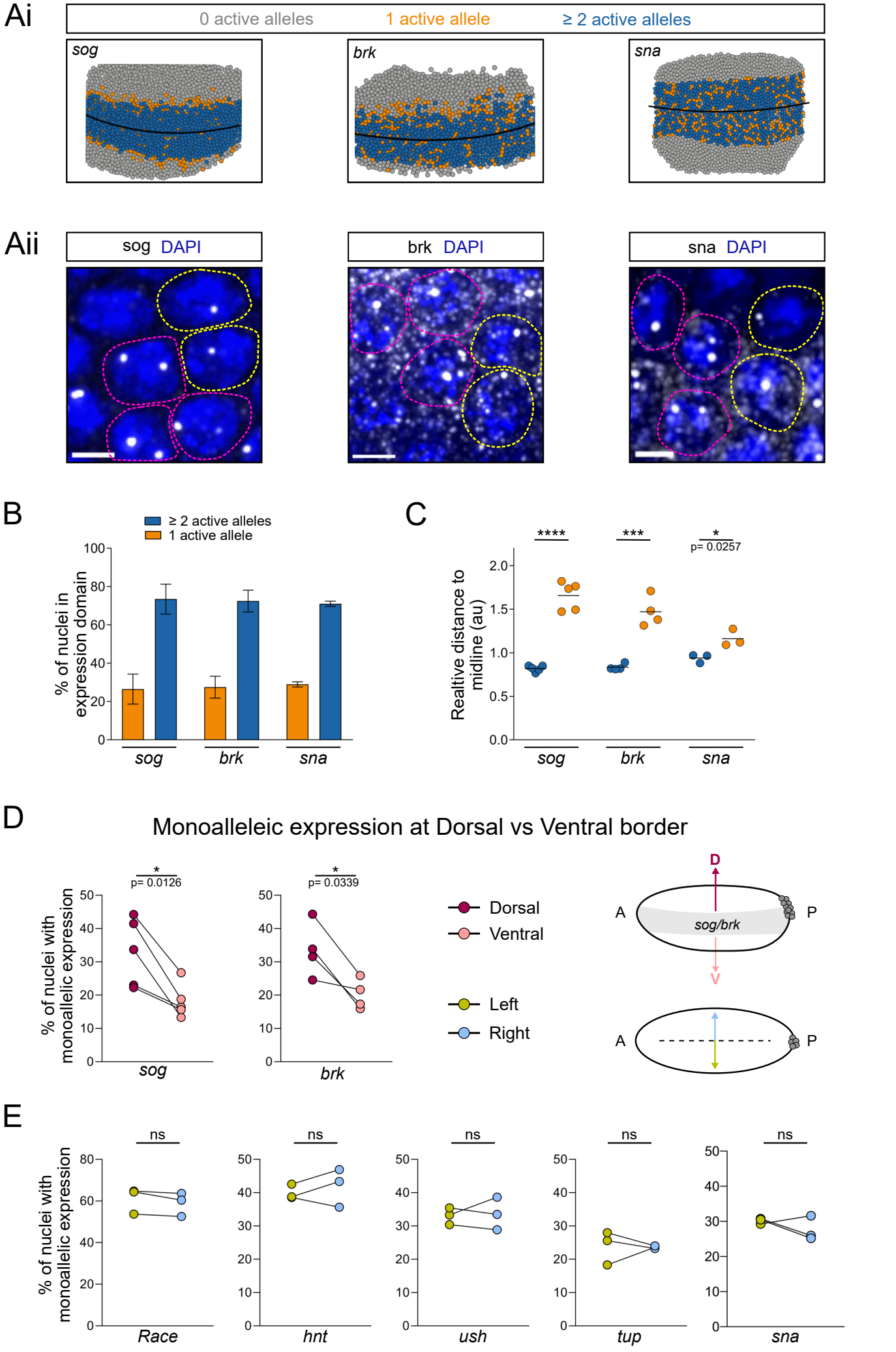
(F) Bursting simulation of *ush* (orange), *hnt* (blue) and *hnt*>*ush* (green) transcription based on mean burst parameter values from the central expression domains and typical transcription onset times.

(G) Cartoon shows the different promoter state profiles for *ush* and *hnt*, how these respond to altered BMP signalling levels, and the resulting effects on allele activity and mRNA number. Mean ± SD (B, E) or mean ± SEM (C) for n = 3 biological replicates. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns = not significant. Student's t-test (B) and a two-way ANOVA with a Sidak's multiple comparisons test shows the difference between means of the two genotypes (E). Data for *ush* transcription in *ush* MS2 wt embryos were analysed in previous Figures and are shown here for comparison. **See also Figure S5.**

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

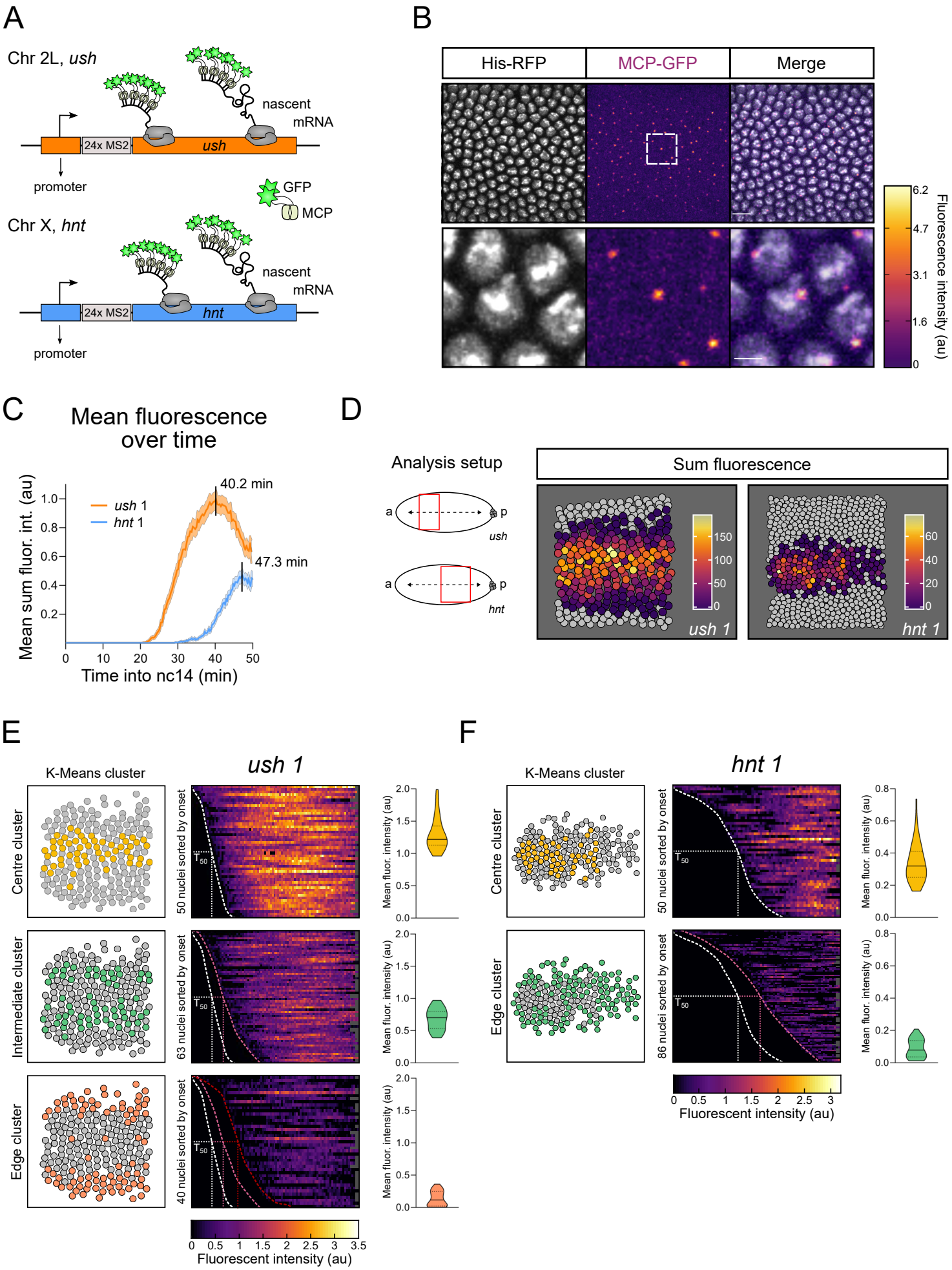


bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

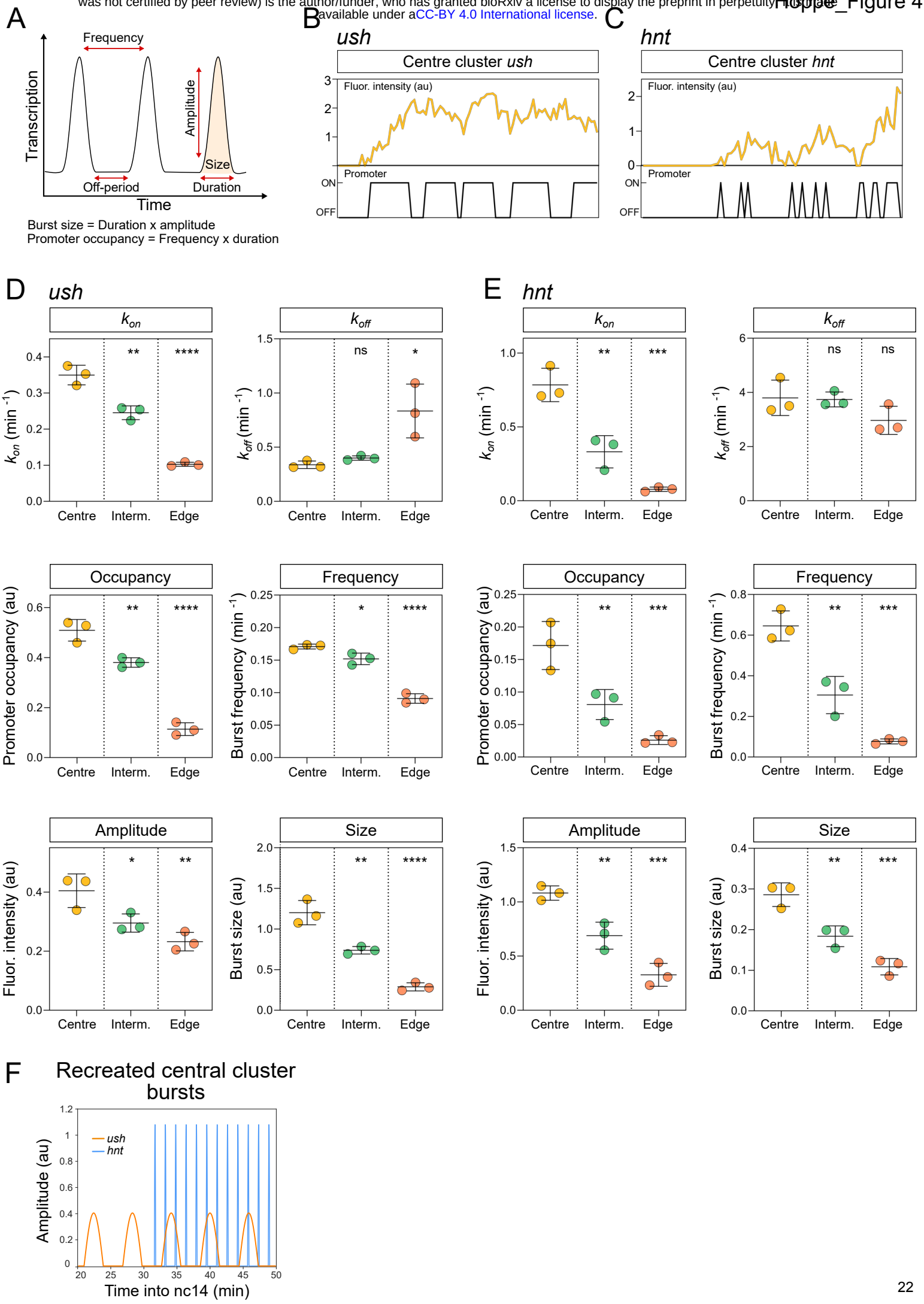


bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

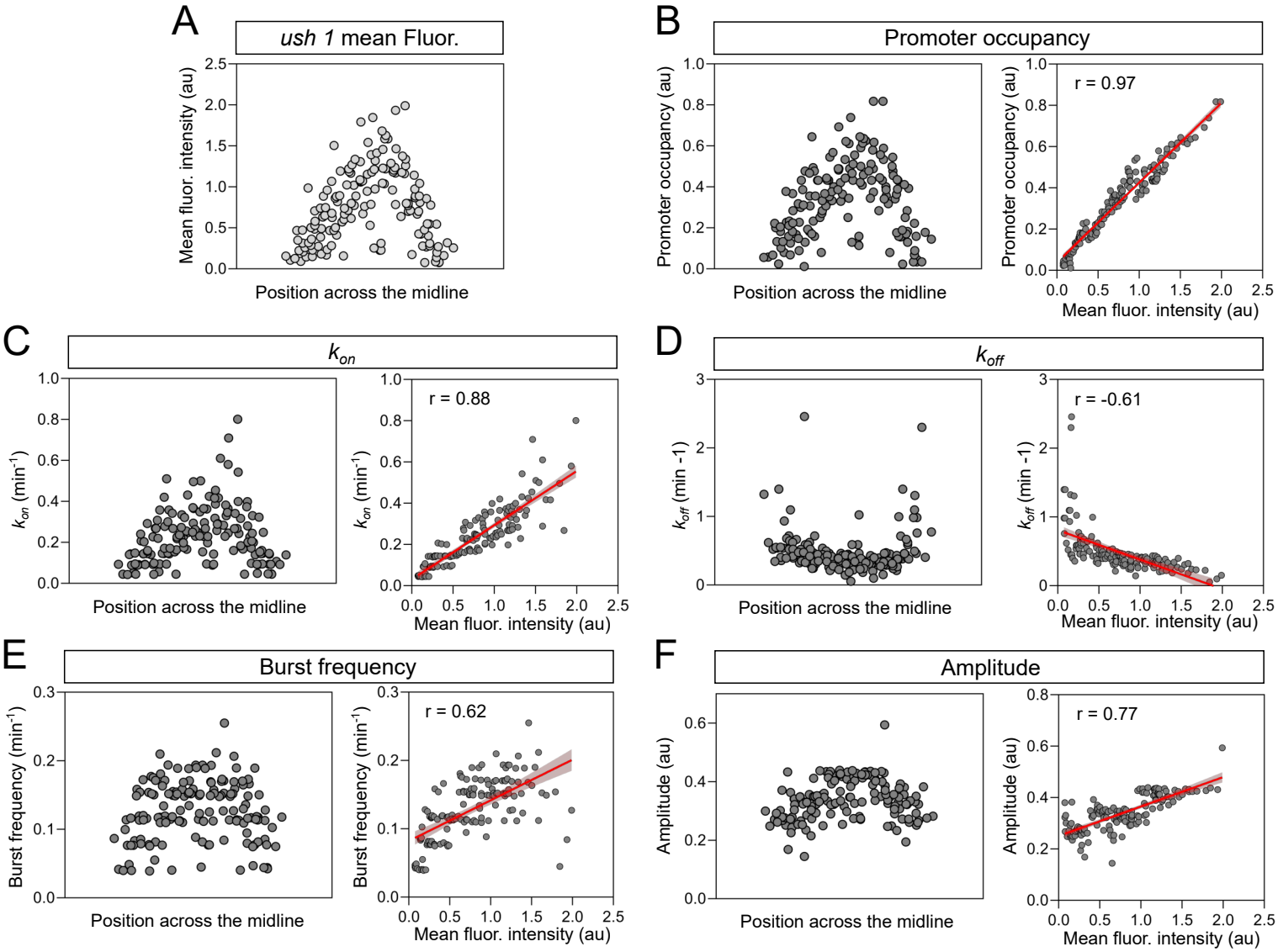
Hoppe_Figure 3



bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

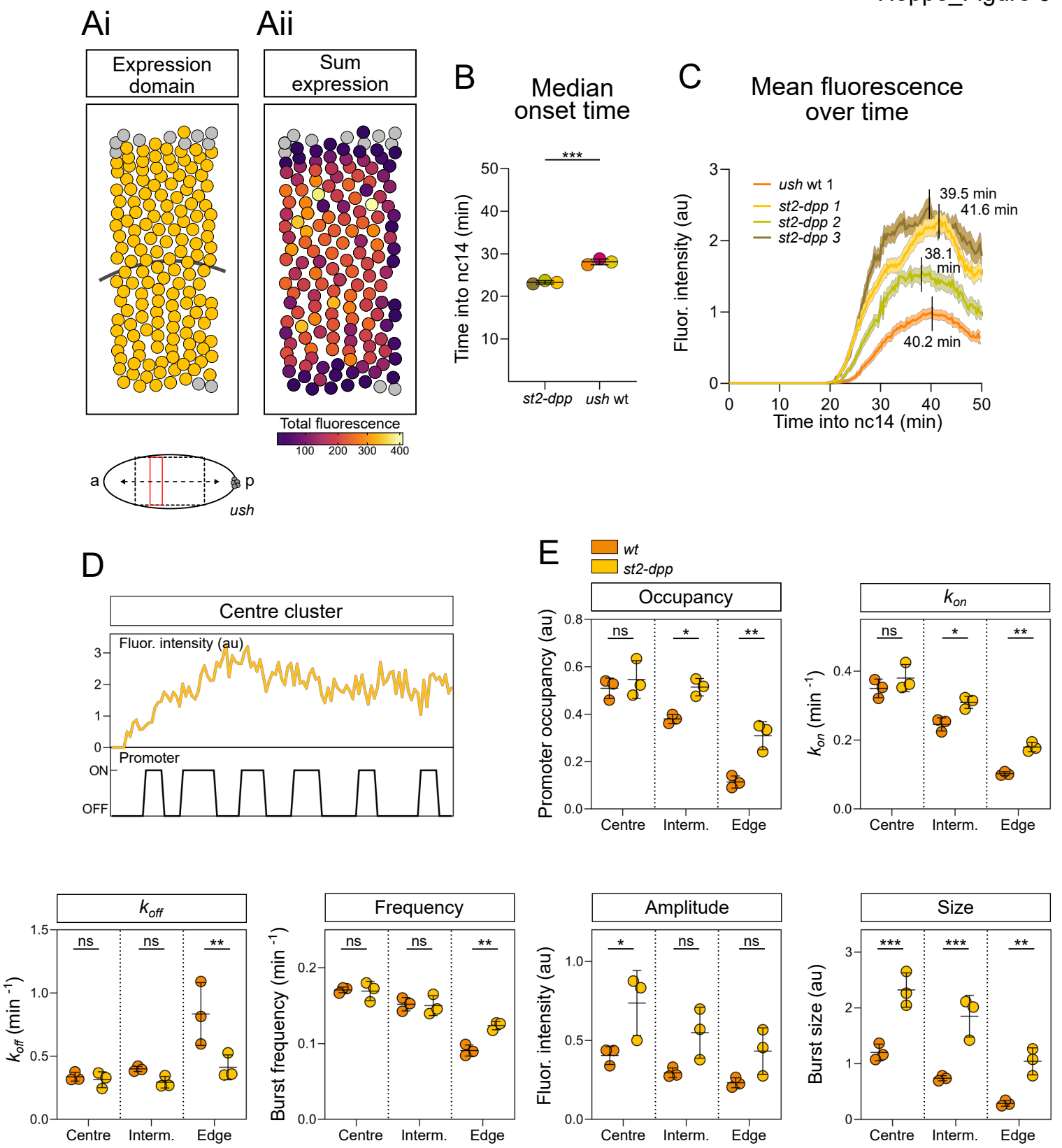


bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

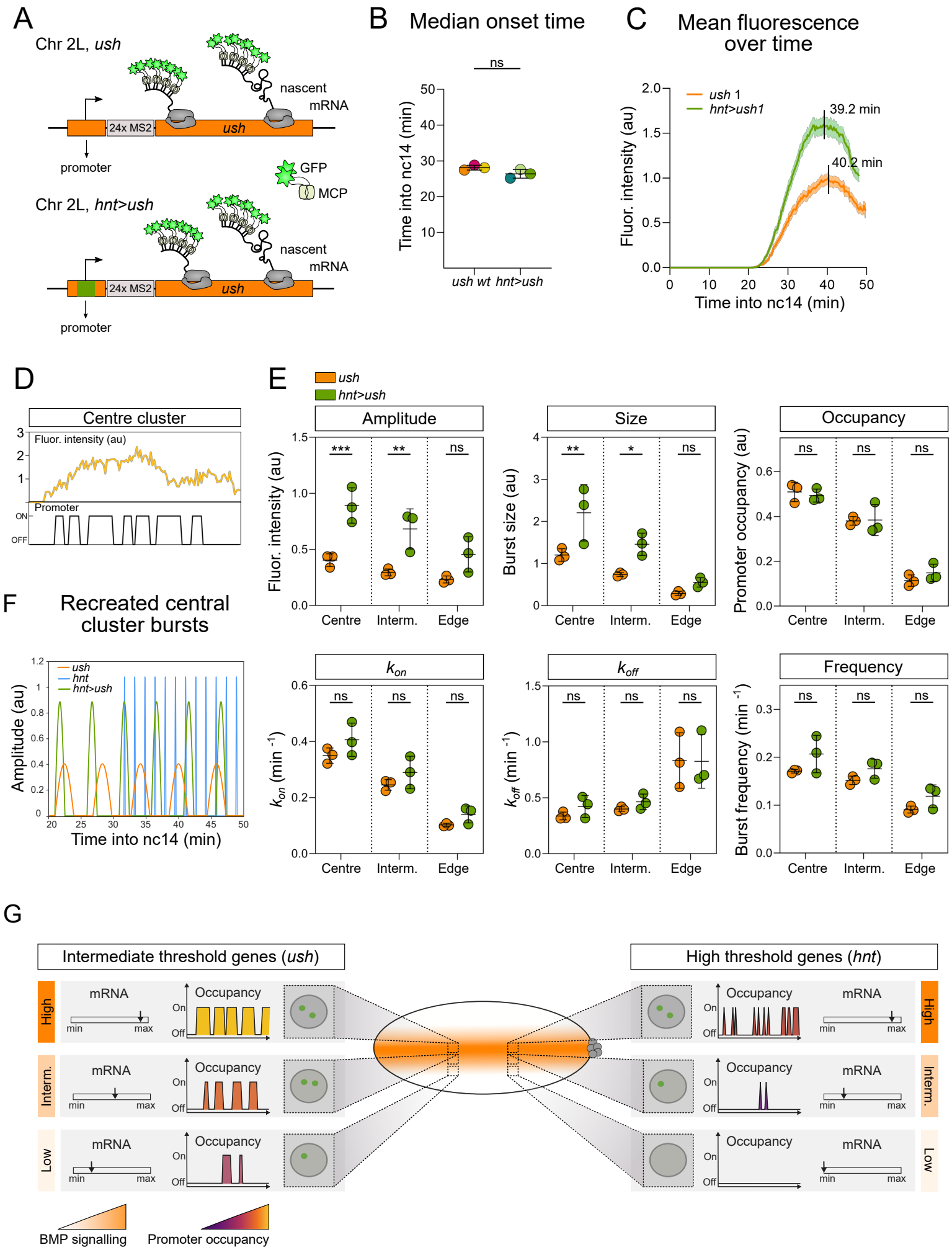


bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hoppe_Figure 6



bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

METHODS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experimental animals

Drosophila melanogaster flies were grown and maintained at 18°C while fly crosses for imaging were raised and maintained at 25°C. All flies were raised on standard fly food (yeast 50g/L, glucose 78g/L, maize flour 72g/L, agar 8g/L, nipagen 27ml/L, and propionic acid 3ml/L). Embryos were collected on apple juice agar plates that contained yeast paste.

The following fly lines were used for experiments in this study; st2-*dpp* (Ashe et al., 2000), $y^{1}w$*;*P{His2Av-mRFP1}II.2; P{nos- MCP.EGFP}2* (BDSC Cat# 60340, RRID:BDSC_60340), $y^{67c23}w^{118}$; *24xMS2-ush* (this study), $y^{67c23}w^{118}$; *hnt>24xMS2-ush* (this study), y^{1}*M{vas-Cas9}ZH-2A*w^{118}, 24xMS2-*hnt* (this study), and $y^{67c23}w^{118}$ which we used as wildtype.

Generation of endogenous MS2 lines

Live imaging fly lines were generated through a two-step method of CRISPR/Cas9 genome editing with homologous recombination and ϕC31 integrase-mediated site-specific transgenesis.

First deletions in the 5'UTR regions of *ush* isoform RC (456bp; Chr 2L: 523446-523902, dm6 genome) and *hnt* isoforms RA and RB (705bp; ChrX: 4617319-4618023, dm6 genome) were generated. Two PAM sites (flyCRISPR Optimal Target Finder tool: http://flycrispr.molbio.wisc.edu/tools) were used to create double strand breaks.

The plasmid pTV^{cherry} (gift from the Vincent lab; DGRC #1338) was used as a donor plasmid containing an attP reintegration site flanked on either side by homology arm sequences. Homology arms were inserted using KpnI and SpeI restriction sites, respectively.

<u>*ush* HA1:</u> forward primer GGTACCgtgcatagccacgacgttagg,
reverse primer GGTACCccggggacgagacgagacctctta
<u>*ush* HA2:</u> forward primer ACTAGTggaagtgacaacataattgcc,
reverse primer ACTAGTtccaagccttcactccactc

<u>*hnt* HA1:</u> forward primer GCTAGCgaagggttgctggtcacc,
reverse primer GCTAGCcattgggtgcgtgtgtgtg

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

hnt HA2: forward primer ACTAGTcaactgttgaacacaatttcac,
reverse primer ACTAGTcacacatgcatacatccagtc

The pU6-BbsI-chiRNA plasmid (RRID:Addgene_45946) was used to deliver guide RNAs (gRNA). 5' phosphorylated oligonucleotides were annealed and ligated into the BbsI restriction site. Together, gRNA plasmids and the donor plasmid were injected into Cas9 expressing flies (BDSC Cat# 51323, RRID:BDSC_51323) by the Cambridge University injection service.
ush gRNA1: forward primer cttcgtctcgtctcgtccccgctc,
reverse primer aaacgagcggggacgagacgagac
ush gRNA2: forward primer cttcgattatgttgtcacttcccgt,
reverse primer aaacacgggaagtgacaacataatc
hnt gRNA1: forward primer cttcgcgcaaataggattacacat,
reverse primer aaacatgtgtaatcctatttgcgc
hnt gRNA2: forward primer cttcgattgtgttcaacagttgcga,
reverse primer aaactcgcaactgttgaacacaatc

Next, the attB-attP system was used for site-specific reintegration. Reintegration fragments were inserted into the RIV^{cherry} plasmid (gift from the Vincent lab; DGRC #1331). Wildtype sequences of promoter and 5'UTR regions, previously removed in the CRISPR process, were inserted into RIV^{cherry} using the NotI site to reconstitute wildtype loci. The 24xMS2-loop cassette (pCR-24xMS2L-stable, RRID:Addgene_31865) was inserted using the BglII site. The RIV^{cherry} plasmid was co-injected with a ϕC31 integrase plasmid (Injection service) into the balanced CRISPR fly lines. Successful transformants were balanced and the marker region was removed by crossing to a cre-recombinase expressing fly line (BDSC Cat# 1501, RRID:BDSC_1501).

Promoter swap fly line *hnt>ush*

The core promoter sequence of *hnt* was inserted into the previously generated fly line carrying the *ush* 5'UTR deletion and an attP site. The core *hnt* promoter sequence and annotation (200 bp; Chr X: 4,617,464 - 4,617,663 dm6 genome) was determined based on peaks from Global Run-On Sequencing (GRO-Seq) data (Saunders et al., 2013). After co-injection with the ϕC31 plasmid, successful transformants were crossed to a Cre-recombinase expressing fly line (BDSC Cat# 1501, RRID:BDSC_1501). Full cloning details available upon request.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

All MS2 tagged lines generated for this study are homozygous viable and fertile.

METHOD DETAILS

Fluorescence in situ hybridisation

Embryo collections (2-4h), RNA

probe synthesis and in situ hybridisation with digoxygenin-UTP-labelled (Sigma, 11277073910) or biotin-UTP-labelled probes (Sigma, 11685597910) were performed as described (Kosman et al., 2004). Antisense probes were approximately 1kb in length (Primer sequences in Table S1). The following primary and secondary antibodies were used: Sheep Anti-Digoxigenin Antibody (1:250 Roche Cat# 11333089001, RRID:AB_514496), mouse anti-biotin (1:250 Roche, 1297597), donkey anti-Sheep IgG Secondary Antibody, Alexa Flue or 555 (1:500; Thermo Fisher Scientific Cat# A-21436, RRID:AB_2535857) and donkey anti-Mouse IgG Secondary Antibody, Alexa Fluor 647 (1:500; Thermo Fisher Scientific Cat# A-31571, RRID:AB_162542). Samples were incubated with DAPI (1:500; NEB, 4083) and mounted in ProLong Diamond Antifade Mountant (Thermo Fisher, P36965).

DNA oligonucleotides

Exonic probe sets for smFISH (Biosearch Technologies) and intronic probe sets for smiFISH (Sigma) can be found in Table S2. Probes for smFISH were conjugated to Quasar 570 fluorophores and smiFISH probes were hybridised to Z-flaps conjugated to Quasar 647 fluorophores (*hnt* and *ush*) or to X-flaps which were conjugated to Quasar 647 fluorophores (*tup*) (gift from the Ronshaugen lab, 2B Scientific).

smFISH/smiFISH

smiFISH probes were hybridised to Flaps as described (Tsanov et al., 2016) and mixed with smFISH probes. Fixed embryos, staged to be 2-4h old, were transferred into Wheaton vials (Z188700-1PAK, Sigma), washed 5 min in 50% methanol/50% phosphate-buffered saline with 0.1% Tween-20 (9005-64-5, Sigma) (PBT), followed by four 10 min washes in PBT, a 10 min wash in 50% PBT/5% wash buffer (10% formamide in 2X SSC; 300mM NaCl and 30mM trisodium citrate adjusted to pH 7) and two 5 min washes in 100% wash buffer. Next, embryos were rinsed once and incubated 2h at 37ºC in smFISH hybridisation buffer (2.5mM dextran sulphate, 10% formamide in 2X SSC). During that time the hybridisation buffer was exchanged twice. Probes were diluted in hybridisation buffer to a final concentration of 1.25mM for smFISH Stellaris probes, and 4mM probe/FLAP duplex for smiFISH probes. Embryos were incubated in probe solution for 14h at 37ºC, washed min in pre-warmed hybridisation buffer at 37ºC, followed by three 15 min washes in pre-warmed wash buffer at 37ºC. At room temperature, embryos were 15 min in wash buffer and three times 15 min in

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

PBT in the dark. One of the PBT washes included DAPI (1:500). Embryos were then mounted in ProLong Diamond Antifade Mountant. All washes were performed with agitation. The number of cytoplasmic mRNA molecules was quantified based on signal in the exonic smFISH probe channel. The number of nascent transcription sites was determined based on signal in the intronic smiFISH channel.

FISH/smFISH microscopy

Images were acquired with a Leica TCS SP8 AOBS inverted microscope using a 40x/ 1.3 HC Pl Apo CS2 or 63x/ 1.4 Plan APO objective with 2x line averaging. The confocal settings were as follows, pinhole 1 airy unit, scan speed 400Hz and format 2048 x 2048 pixels. Images were collected with either Photon Multiplying Tube Detectors or Hybrid Detectors and illuminated using a white laser. The following detection mirror settings were used: Photon Multiplying Tube Detector at 405nm (4.66%); Hybrid Detectors: 490nm (10%, 0.3 to 6us gating), 548nm (26.1%, 0.3 to 6us gating) and 647nm (17%, 0.3 to 6us gating). All images were collected sequentially and optical stacks were acquired at 300nm spacing. Raw images were then deconvolved using Huygens Professional software (SVI, RRID:SCR_014237) and maximum intensity projections are shown in the figures.

Live Imaging microscopy

Female flies of the genotype His2av-RFP; MCP-GFP (BDSC Cat# 60340, RRID:BDSC_60340) were crossed to wildtype or *st2-dpp* (Ashe et al., 2000) expressing males. Female offspring from this cross were mated with males homozygous for the 24*xMS2* tagged target gene locus to supply a maternal source of His-RFP; MCP-GFP.

Embryos were dechorionated in bleach and positioned dorsally on top of a coverslip (Nr. 1, 18x 18 mm; Deltalab, D101818), thinly coated with heptane glue. A drop of halocarbon oil mix (4:1, halocarbon oil 700: halocarbon oil 27; Sigma H8898 and H8773) was placed in the middle of a Lumox imaging dish (Sarstedt, 94.6077.305) and two coverslips (Nr. 0, 18x 18mm; Scientific Laboratory Supplies, PK200) were placed on either side of the oil drop, creating a bridge. The coverslip with the embryos glued to it was then inverted into the oil, sandwiching the embryos between the imaging dish membrane and the coverslip.

Embryos were imaged on a Leica TCS SP8 AOBS inverted confocal microscope with a resonant scan head, using a 40x/ 1.3 HC PL apochromatic oil objective. Images were obtained with the following confocal settings, pinhole 1.3 airy units, scan speed 8000Hz bidirectional, format 1024 x 700 pixels at 8 bit. Images were collected using the white laser with 488nm (8%) and 574nm (2%) at 8x line averaging and detected with hybrid detectors. Three-dimensional optical sections were acquired at 1 µm distance, a final depth of 55 µm

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and a final temporal resolution of 20 seconds per time frame. Images were processed with the Leica lightning deconvolution software. The mounting medium refractive index was estimated to be 1.41. Maximum intensity projections of 3D stacks are shown in the result sections. Embryos were imaged for 70-90 min and included the cleavage cycle of nc14 and the onset of gastrulation. During analysis all datasets were adjusted in time to account for slight temperature differences during imaging that can alter the speed of development. Therefore, nc14 was defined as the time between telophase of cleavage cycle 14 and the beginning of cephalic furrow formation. For the purpose of this study, nc14 was defined to last for 50 min similar to (Berrocal et al., 2018).

QUANTIFICATION AND STATISTICAL ANALYSIS

Image Analysis of static FISH and smFISH images in Imaris

Nuclei and RNA puncta were initially detected using the Imaris software 9.2 (Bitplane, Oxford Instruments, Concord MA, RRID:SCR_007370). RNA puncta were then assigned to nuclei in a proximity based method using custom python scripts.

Nuclei were identified and segmented using the Imaris "surface" function. Nascent transcription foci were identified using Imaris "spots" function and estimated to be 0.6 µm in diameter with a z-axis point spread function of 1 µm. Single mRNA puncta were identified with spot volumes of 0.3 µm across and 0.6 µm in the z direction. Customised Python scripts were used to analyse the data extracted from Imaris and are described below.

Quantification of cell width bins for expression domain edge comparison

Bins of one cell width were defined to be 5 µm wide. The wt expression domain of *ush* was determined to be approximately 100 µm in width and the wt edge region was defined as the outermost 15% of the expression domain. The wt edge domain in *st2-dpp* and wt embryos was defined as 15 µm wide area located approximately 30-45 µm away from the dorsal midline.

Nuclear tracking and spot identification in live imaging data sets in Imaris

Nuclei were first smoothed and blurred using a wavelet filter (Imaris X-tension by Egor Zindy) and then segmented using the Imaris "surface" function based on the His-RFP fluorescent channel. Nuclei were tracked through time in 3D using the inbuilt autoregressive motion with a maximum frame gap size of 5 and a maximum travel distance of 5 µm. Active transcription sites were detected using the Imaris "spots" function in three-dimensions. Transcription foci were estimated to be 1.8 µm across with a z-axis point spread function estimation of 7.8 µm. To determine the background fluorescence of the data set, a set of "spots" were generated for background correction. Here, four spots were inserted every third

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

time frame, avoiding nascent transcription sites. The background correction spots have the identical volume to the transcription site spots.

Custom python scripts for live imaging data analysis

1. Spot assignment to nuclei

For both static and live imaging spots were assigned to nuclei using the long axis of the nucleus as a reference for the midline of each nucleus. The long axis for each nucleus was calculated, using the Imaris 9.2 ellipsoid axis C and spots were then assigned to the nearest nuclei axis within the 3D space. The number of spots assigned to each nuclei was recorded.

2. Nuclei distance to midline

The midline for the expression domain was calculated by fitting a polynomial (2-dimensions) using the coordinates of the mRNA spots as detected by Imaris 9.2. The distance of each nucleus was then calculated back to the midline and reported in µm.

3. Mitotic Wave correction

To correct for time differences in transcriptional onset due to the mitotic wave, the temporal profile of cell areas was synchronised. The microscopy time frame at which of telophase was noted for each cell area along the AP axis. These data were then used to set the zero time point for each position along the long axis of the embryo were adjusted relative to this time point.

4. Background subtraction

Background was recorded from the first time point where fluorescent foci were identified in the MS2 data. Background was then recorded every 3 frames until the end of the video. The background was then fit as a linear polynomial (1 dimension). The equation of the line was then used to calculate the background level at every time point. The raw value was then corrected for background as in: $\frac{\text{raw value - background}}{\text{background}}$.

Modelling Changes in Kinetic Parameters of Transcription

We used a memory-adjusted hidden Markov model (mHMM) to infer the promoter state activity given MS2 flourescence data (Lammers et al., 2019). The model parameters are the transition rates between on and off states of the promoter and the mean/variance of the signal in the on and off states. In order to investigate the spatial regulation of transcriptional parameters, K-means clustering (using sklearn.cluster.KMeans) of the MS2 fluorescence traces was used to partition each *ush* and *hnt>ush* embryo into three clusters of cells with similar dynamics. The MS2 fluorescence dataset for each *hnt* embryo was instead divided into three approximately equally-sized groups of cells based on expression level, due to the

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

inability of the K-means algorithm to subdivide the narrower *hnt* expression domain into three distinct clusters. The mHMM was trained separately on each of these three cell cluster datasets per embryo in order to generate the graphs showing global transcriptional parameters per cluster or expression group. Inferred global transcriptional parameters included promoter switching on rate (*kon*), promoter switching off rate (*koff*), Pol II initiation rate (kini, expressed in terms of A.U.), promoter mean occupancy (<n>), burst size (*kini* / *koff*) and burst frequency ((*kon* * *koff*) / (*kon* + *koff*)) (Zoller et al., 2018). The global parameters for each embryo were then used to generate a set of inferred posterior promoter traces for each individual cell within the embryo (using the Forward-Backward algorithm) allowing for estimation of cell-specific promoter switching rates, mean occupancy, burst frequency and amplitude.

The model state-space for the mHMM is the sequence of promoter on-off states within a window of length K which is determined by the elongation time (determined by length of the gene and estimated transcription speed, see Lammers et al., 2019). The state-space of the mHMM is therefore 2^K in size. This state-space is too large for us to use the original matlab implementation of the model here because of computational space and time limitations, and therefore we reimplemented the model in python using a truncated state-space approximation. We used the Forward algorithm to rank states dynamically by probability given the current and previous observations in the sequence and we removed states below M in rank at each time, where M is a user-defined number of stored states that determined the accuracy of the approximation. Full details of this scalable implementation of the memory adjusted HMM are described in a forthcoming publication (Bowles et al., in preparation).

Statistical Analysis

Statistical comparisons were performed using two-tailed Student's t tests, Mann-Whitney test, Kruskal-Wallis test with multiple comparison, one-way ANOVA with multiple comparison, two-way ANOVA with multiple comparison and paired Student's t tests using GraphPad Prism (RRID: SCR 002798) and R. Statistical test and sample sizes can be found in Figure legends. Statistical significance was assumed by $p<0.05$. Individual p values are indicated in Figure legends.

METHOD REFERENCES

Ashe, H.L., Mannervik, M., and Levine, M. (2000). Dpp signaling thresholds in the dorsal ectoderm of the Drosophila embryo. Development *127*, 3305–3312.

Berrocal, A., Lammers, N., Garcia, H.G., and Eisen, M.B. (2018). Kinetic sculpting of the seven stripes of the Drosophila even-skipped gene. BioRxiv 335901.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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*, 846–846.

Lammers, N.C., Galstyan, V., Reimer, A., Medin, S.A., Wiggins, C.H., and Garcia, H.G. (2019). Multimodal transcriptional control of pattern formation in embryonic development. BioRxiv 335919.

Saunders, A., Core, L.J., Sutcliffe, C., Lis, J.T., and Ashe, H.L. (2013). Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. Genes Dev. *27*, 1146–1158.

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.

Zoller, B., Little, S.C., and Gregor, T. (2018). Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting. Cell *175*, 835-847.e825.

SUPPLEMENTARY INFORMATION

Figure S1: BMP target gene transcription during nc14. Refers to Figure 1.

(A) Expression patterns of BMP target genes *Race*, *hnt*, *ush* and *tup* in *Drosophila* embryos. Insets are shown in Figure 1 Aii.

(B) Percentages of monoallelic and biallelic *ush* transcription in each time course embryo.

(C) Number of *hnt* mRNAs per cell in single embryos of increasing age in nc14 (i), proportion of monoallelic and biallelic transcription (ii), with quantitation based on the number of active transcription foci shown for the oldest embryo (iii). Spearman correlation coefficient is shown for each pair of variables.

(D) The number of *hnt* mRNAs per cell plotted according to the nuclear distance from the dorsal midline is shown for embryos from (Ci). Embryos are ordered by increasing age and points are coloured based on the number of active transcription foci.

(E, F) As in (C, D), but data are for *tup*.

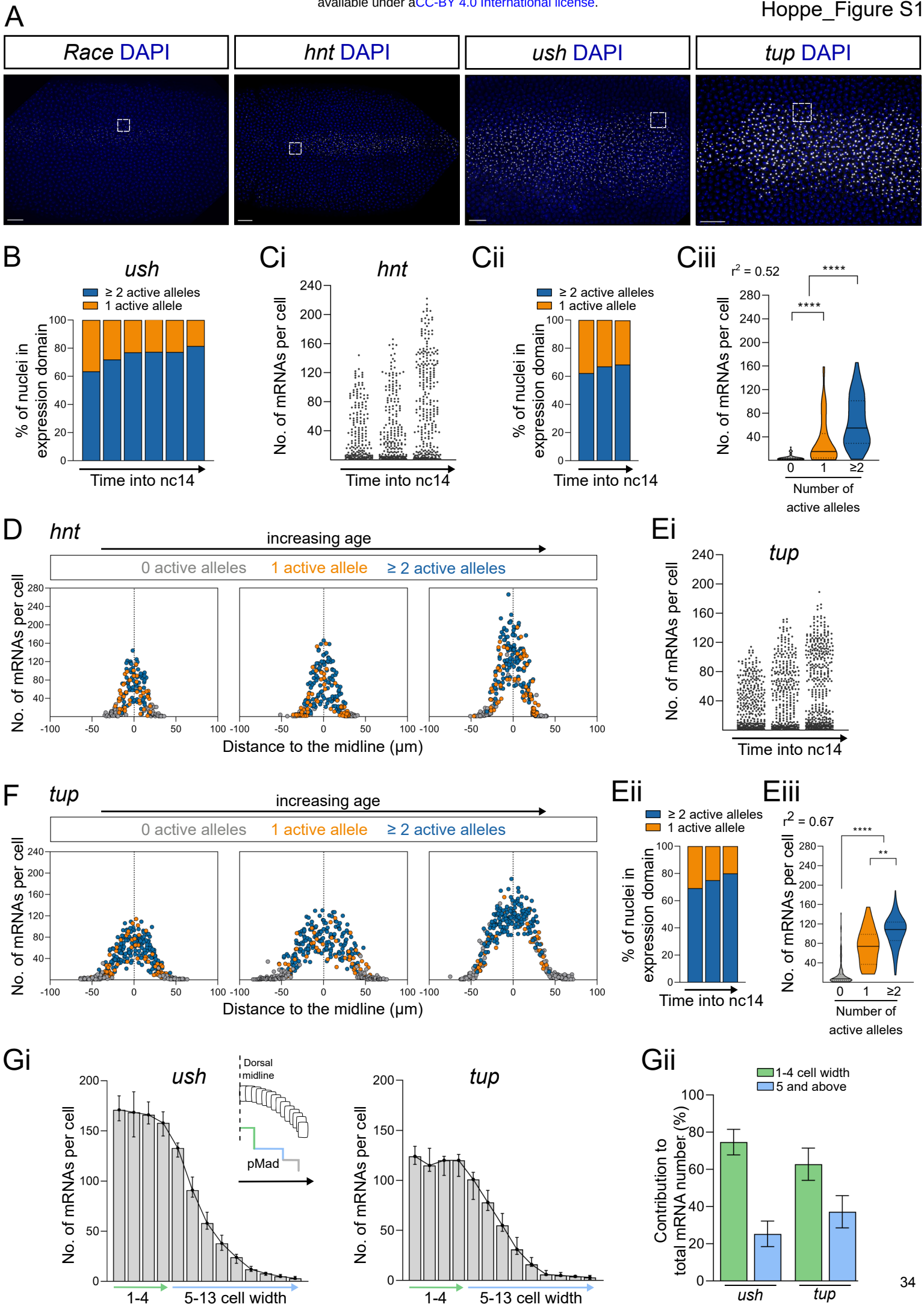
(Gi) Number of *ush* and *tup* transcripts produced by cells within 1 cell wide bins mirrored at the midline and shown for the oldest time course embryos containing n = 453 (*ush*) and 364 (*tup*) cells in total. Inset cartoon shows pMad step-gradient during nc14. Bin width = 5µm from the midline.

(Gii) Contribution of cells (1-4 cell widths and >5 cell widths from the midline) to the overall number of transcripts for n = 6 (*ush*) and 3 (*tup)* embryos.

Embryos are oriented dorsally with anterior to the left. Scale bar, 20 µm (A). $^{**}p < 0.01$, $^{****}p < 0.0001$. Median ± 95% confidence intervals (Ciii, Eiii, Gi) or mean ± SD (Gii). Kruskal-Wallis test with a Dunn's multiple comparisons test (Ciii, Eiii).

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hoppe_Figure S1



bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S2: Characterisation of *ush* and *hnt* transcription dynamics using live imaging. Refers to Figure 3.

(A) ISH of *ush* (top) or *hnt* (bottom) in wt embryos and embryos homozygous for the MS2-loop tagged gene locus.

(Bi) Schematics of the *ush* and *hnt* imaging domains (yellow boxes). Stills from time-lapse data sets (Video S1, S2) false coloured for active transcription show the expression pattern at 37.5 min into nc14.

(Bii) Schematics of the *ush* and *hnt* analysis domains (red boxes). The cumulative expression patterns for representative embryos are shown.

(C) Mean fluorescence of *ush* and *hnt* transcription, with the time of maximum fluorescence shown for all biological replicates.

(D) Transcription onset times of *ush* and *hnt* transcription. Median onset times of biological replicates are plotted (n = 3 for *ush* (209, 186, 223 nuclei) and *hnt* (192, 183, 188 nuclei)).

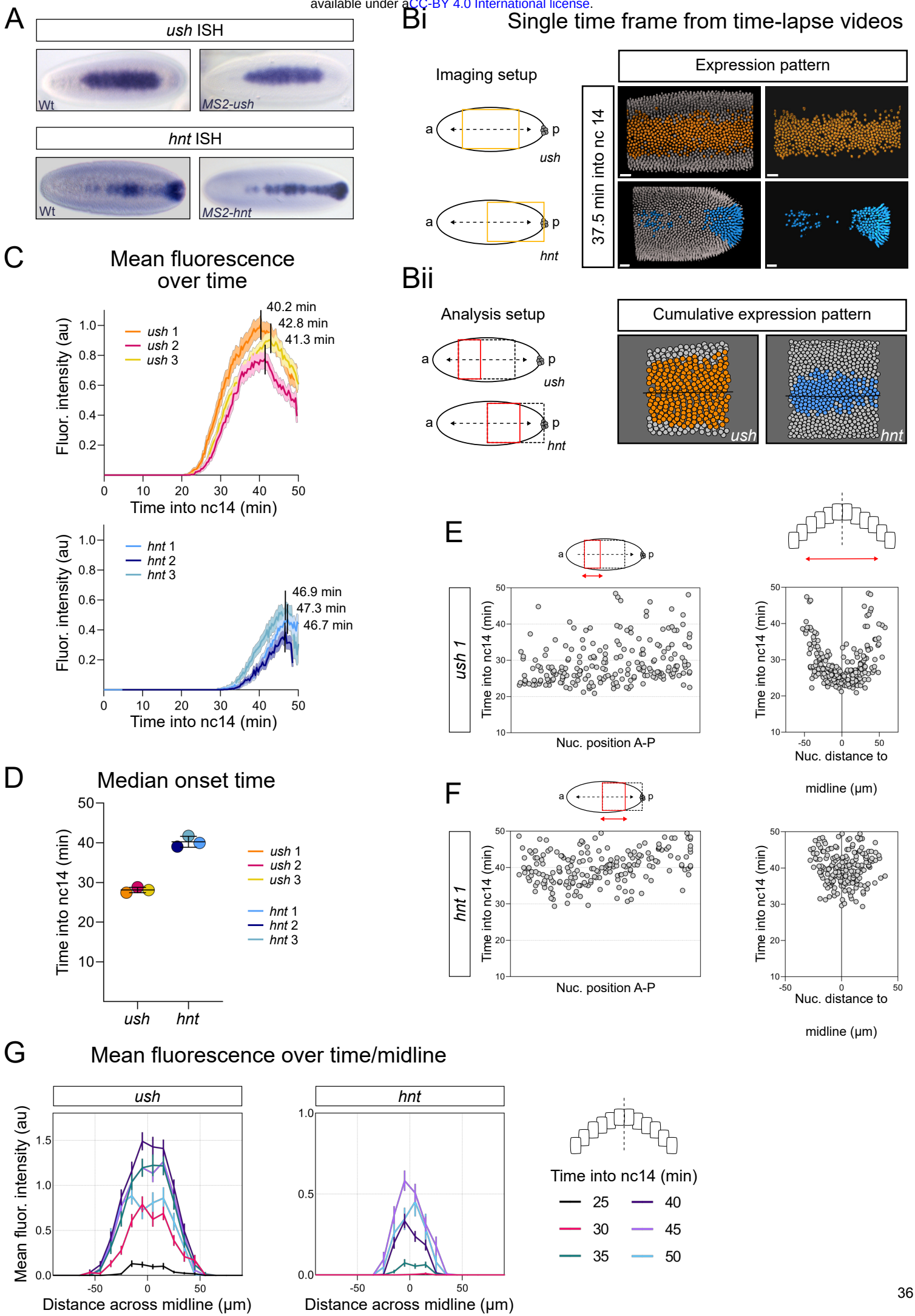
(E) The transcription onset time of *ush* is plotted according to the nuclear position along the AP and DV axes for one representative biological replicate. Each dot represents one nucleus.

(F) As in (E) except transcription onset times of *hnt* are plotted.

(G) Mean fluorescence values of *ush* and *hnt* transcriptionally active nuclei divided into positional bins along the embryo cross section. Bin width = 10 µm, time resolution = 5 min. Data were pooled from all biological replicates. Note different y-axis.

Embryos and schematics are oriented anterior to the left and positioned dorsally. Scale bar, 20µm (Bi). Mean ± SEM (C), mean ± SD (D) or mean ± 95% confidence intervals (G).

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S3: Changes in burst kinetics during *ush* and *hnt* transcription in response to different BMP signalling levels. Refers to Figure 4.

(A) Burst parameter definitions (Zoller et al. 2018) that are used to investigate changes in burst kinetics (i) and schematic of a two-state promoter model where the promoter switches between an active ON and an inactive OFF state. When in the ON state mRNA is produced with the rate of k_{ini}. The probability of switching between the two states is contained in the rates k_{on} and k_{off} (ii).

(B, C) Representative fluorescence trace from an intermediate and edge cluster nucleus showing *ush* (B) and *hnt* (C) transcription and inferred promoter states.

(D, E) Global analysis of mean burst duration ($1/k_{off}$) and promoter off period ($1/k_{on}$) for *ush* (D) and *hnt* (E) transcription in different spatial domains. Data points are coloured according to spatial regions.

Mean ± SD (D, E) for n = 3 biological replicates. $^{*}p < 0.05$, $^{***}p < 0.001$, $^{****}p < 0.0001$, ns = not significant. One-way ANOVA with a Dunnett's multiple comparisons test shows the difference to the central cluster (D, E).

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hoppe_Figure S3

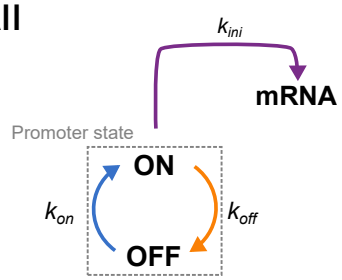
Ai

Parameter definitions:

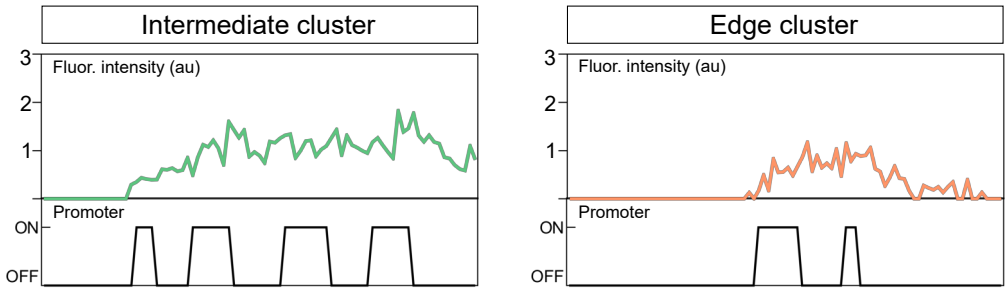
Burst frequency:	$\frac{k_{on} \times k_{off}}{(k_{on} + k_{off})}$
Burst size:	Duration x amplitude
Pol II loading rate:	k_{ini}
Burst duration:	$\frac{1}{k_{off}}$
Promoter off period:	$\frac{1}{k_{on}}$
Promoter occupancy:	$\frac{k_{on}}{(k_{on} + k_{off})}$

(Zoller et al., 2018)

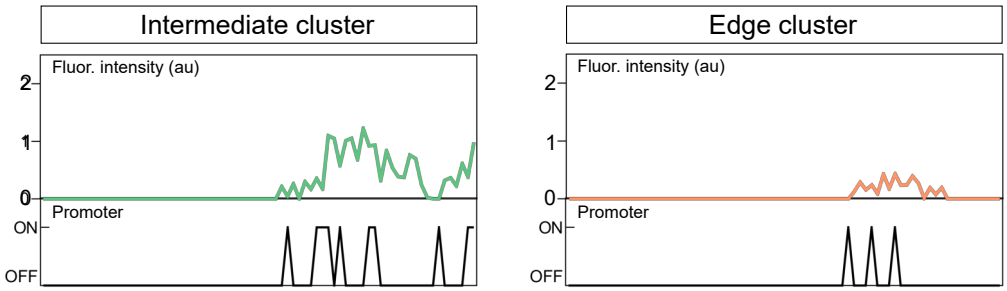
Aii



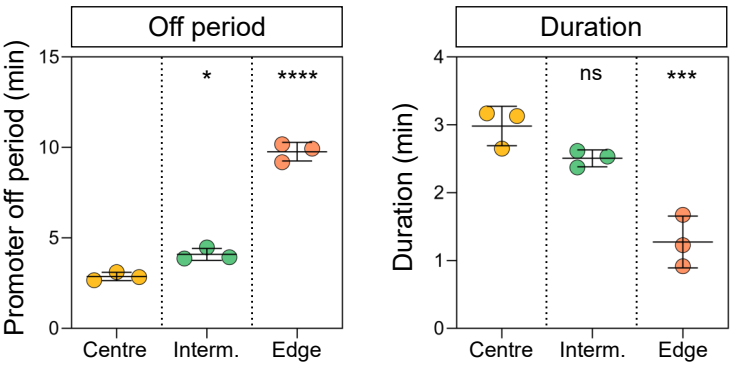
B *ush*



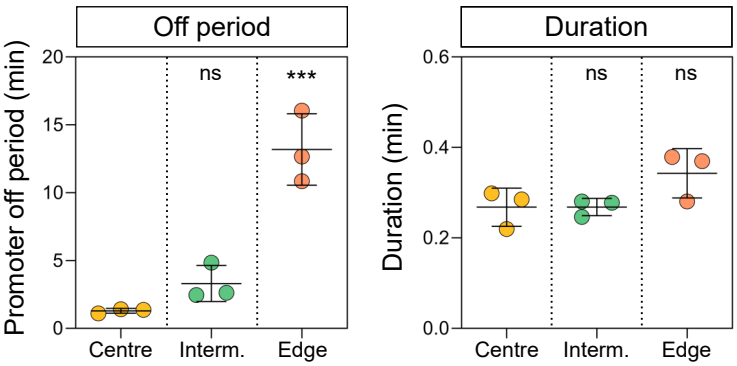
C *hnt*



D *ush*



E *hnt*



bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

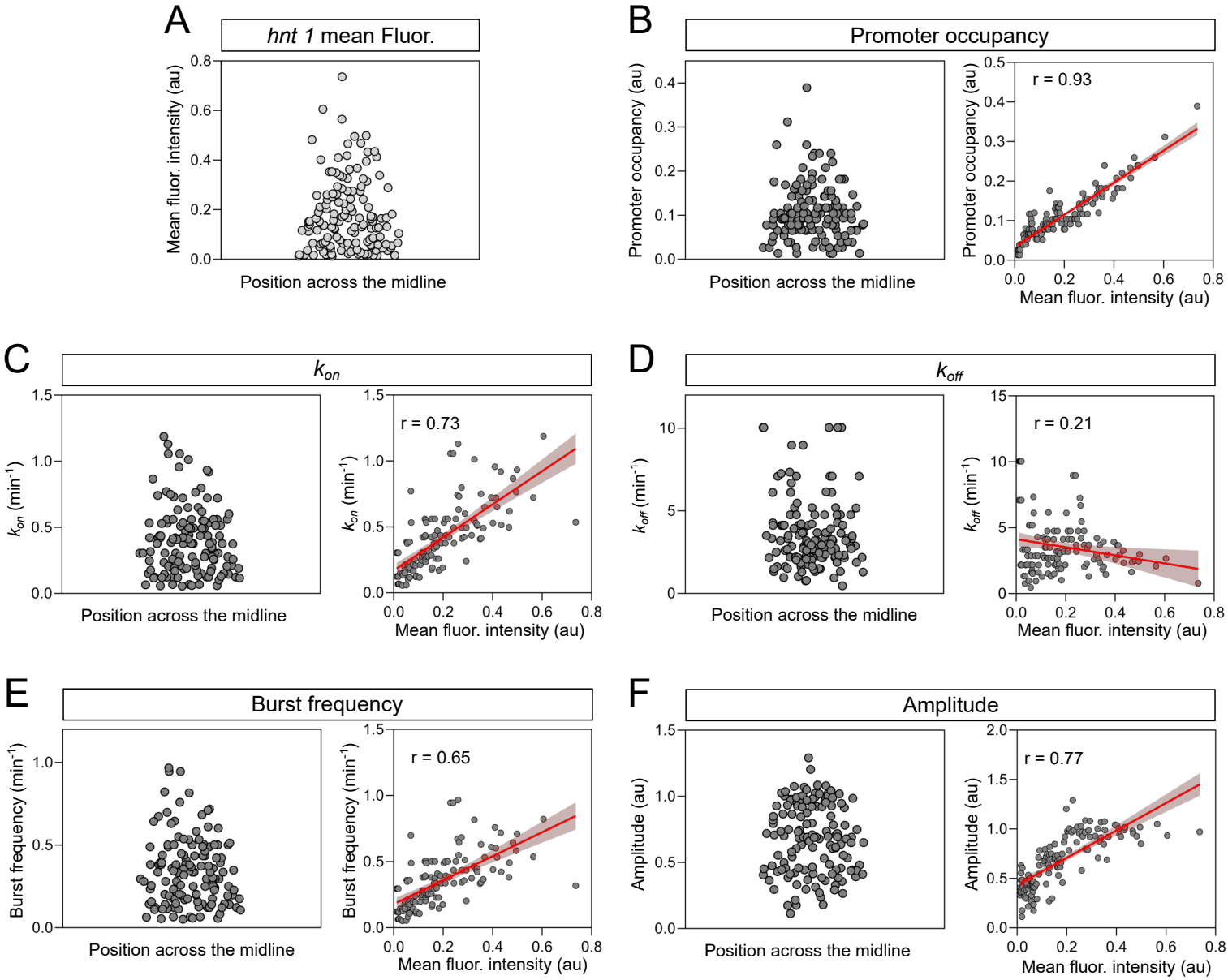
Figure S4: Contribution of individual burst parameters to mean expression level. Refers to Figure 5.

(A) Mean fluorescence output of nuclei transcribing *hnt* plotted according to the nuclear position across the dorsal midline for one representative embryo.

(B-F) Burst kinetic parameters of single *hnt* nuclei plotted across the midline and against mean fluorescence intensity for promoter occupancy (B), k_{on} (C), k_{off} (D), burst frequency (E) and amplitude (F). Pearson correlation coefficient is shown for each pair of variables between burst parameter and mean fluorescence. Linear regression is shown ± 95% confidence intervals.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hoppe_Figure S4



bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S5: Replacing the *ush* promoter sequence changes burst amplitude. Refers to Figure 7.

(A) Cumulative expression domains of *ush* transcription in *ush* (orange) and *hnt>ush* (green) embryos.

(B) Mean fluorescence of *ush* transcription in *ush* and *hnt>ush* embryos, with time of maximum fluorescence listed for all replicates.

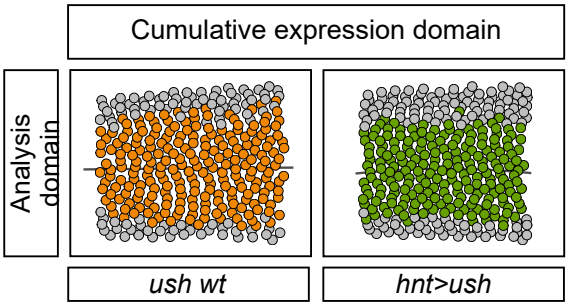
(C) Clustering analysis data shown for one representative embryo for each genotype. (Left) Nuclei are coloured by cluster. (Middle) Heatmaps of single-cell traces, sorted according to transcription onset (scale as indicated, grey indicates periods where nuclei were not tracked). Time of transcriptional onset was traced to visualise onset fronts of different clusters and the position at which half the nuclei in a cluster initiated transcription is indicated (T_{50}). (Right) Mean fluorescence values of nuclei in each cluster. Number of data points given next to heatmaps.

(D) Representative fluorescence trace from an intermediate and edge cluster nucleus showing *ush* transcription and inferred promoter states in a *hnt>ush* embryo.

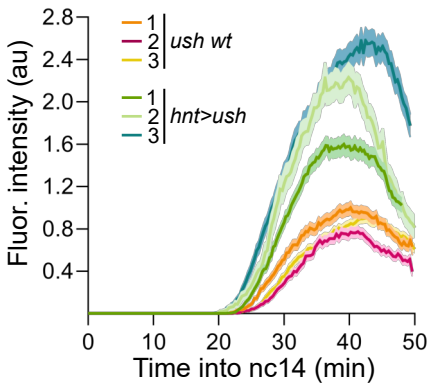
Embryos are oriented dorsally with anterior to the left. Mean ± SEM (B), median ± 95% confidence intervals (C, right). ****$p < 0.0001$. Mann-Whitney test (C). Data for *ush* transcription in *ush* wt embryos were analysed in previous Figures and are shown here for comparison.

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A

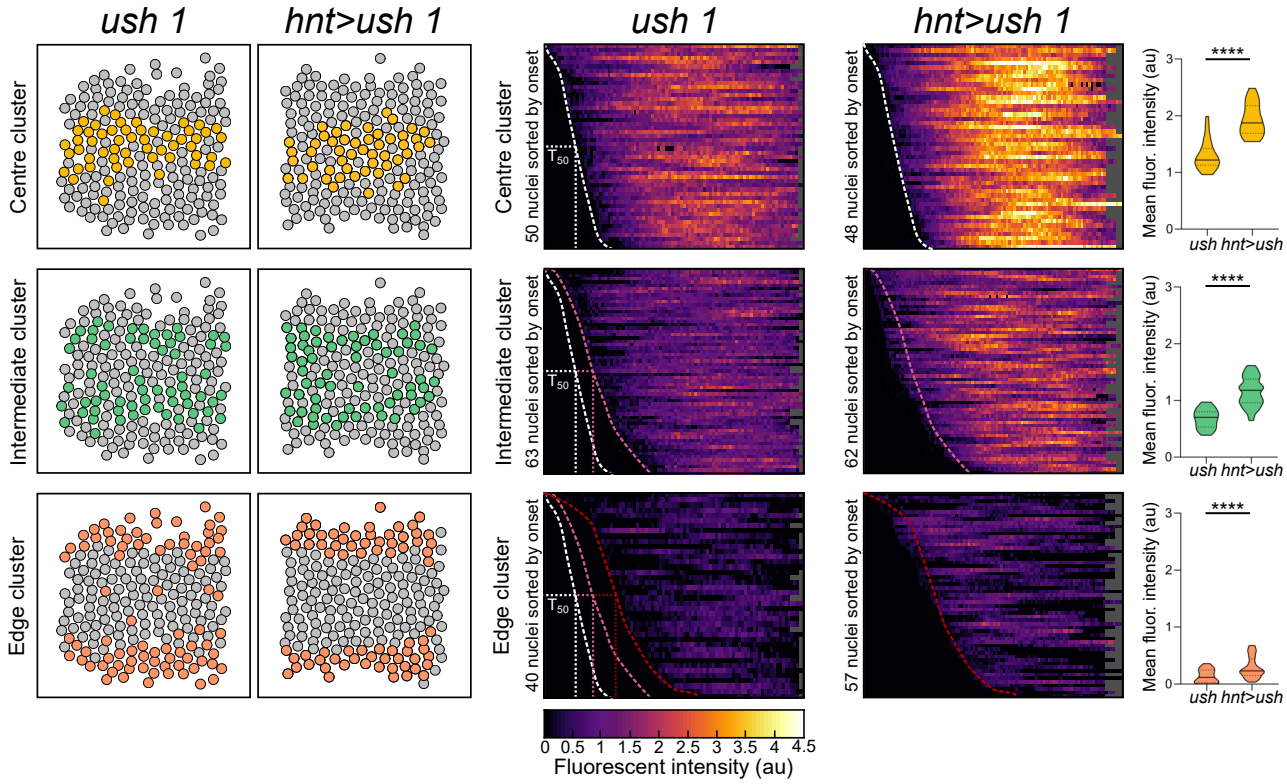


B

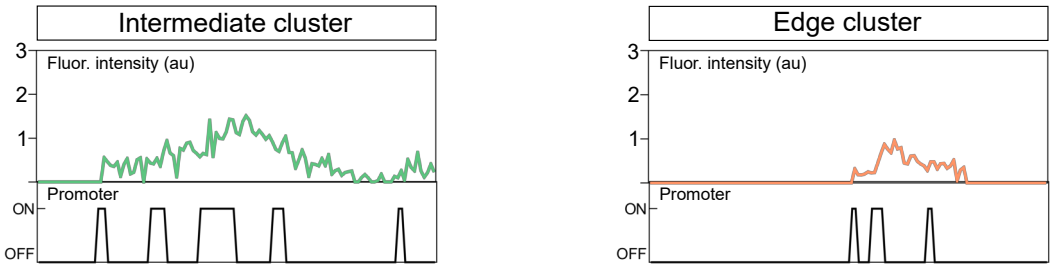


Max. fluorescence intensity	
ush wt 1	40.2 min
ush wt 2	41.3 min
ush wt 3	42.8 min
hnt>ush 1	36.8 min
hnt>ush 2	39.2 min
hnt>ush 3	42.5 min

C



D



bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Video S1: Maximum intensity projection of a representative embryo showing endogenous *24xMS2-ush* transcription (grey) and Histone-RFP (red) imaged with a 40x objective and 20 sec time resolution during nc14.

Video S2: As in Video S1 but showing *hnt* transcription.

Video S3: Maximum intensity projection of a representative embryo showing endogenous *24xMS2-ush* transcription (grey) and Histone-RFP (red) imaged with a 40x objective and 20 sec time resolution during nc14. The expression domain is broadened by a single copy of the st2-*dpp* transgene.

Video S4: As in Video 1 but showing *ush* transcription in a *hnt>ush* embryo.

Table1: Primer pairs used to generate intronic RNA probes used for in situ hybridisation, related to Methods.

Gene	Primer	Primer sequence (T3 or T7 promoter in lower case)
Race	sense	attaaccctcactaaagggaAGTAGAAACATTATTGCAAT
Race	antisense	gaattaatacgactcactatagggaAAGCAAAAAATTACGTTTTT
hnt	sense	attaaccctcactaaagggaATTCCCAAAACCCCTCCCTT
hnt	antisense	gaattaatacgactcactatagggaCCAGTCTTCGATTGTCGCG
ush	sense	attaaccctcactaaagggaGTGAGAATTATTCATAC
ush	antisense	gaattaatacgactcactatagggaATTAAACTACAGT
tup	sense	attaaccctcactaaagggaTAATTACAAACAAATTAA
tup	antisense	gaattaatacgactcactatagggaATTAAATATTTACC
sog	sense	attaaccctcactaaagggaAATTTTATTTTCAATCTATT
sog	antisense	gaattaatacgactcactatagggaAAAAAACGAGAAAATA
brk	sense	attaaccctcactaaagggaGAACAGTTGAACGGATCGGGAGCTT
brk	antisense	gaattaatacgactcactatagggaCGATTCTCCAAATAGCCATGCAG
sna	sense	attaaccctcactaaagggaACACCGGAAAGGAACTCCAG
sna	antisense	gaattaatacgactcactatagggaTCTGTTTGTTTGGTCTTCGCC

Table 2: smFISH and smiFISH probes used for FISH and complementary to intronic and exonic sequences in *hnt*, *ush* and *tup,* related to Methods.

Gene	Probe	sequence	Gene	Probe	sequence
smFISH probe sets					
hnt	1	aatggcgaattttgcgcttg	*hnt*	25	cggaatagctgctgcatata
hnt	2	tagtccatcacaatggatgc	*hnt*	26	gtctgggactggaacatgag
hnt	3	gatagcaccttggagcaaat	*hnt*	27	ggtggtgccataaagggaaa
hnt	4	gcaacttgatgctgtttgtg	*hnt*	28	gattgatacggtttggtgga
hnt	5	agattgttgttagccttgtg	*hnt*	29	aggactccattcttgatgac
hnt	6	atgttgttgttgatggtggt	*hnt*	30	agcagtgttcacaggcaaag
hnt	7	ccaaaagagtggcgacatcg	*hnt*	31	cgaacgcagcgtaaatctcg
hnt	8	gacaacttaggagcaaggca	*hnt*	32	ttgtgcatagaactgcggat
hnt	9	tggtatgtctgagtgcattg	*hnt*	33	cgaaatgggagcatgaccat

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Gene	Probe	sequence	Gene	Probe	sequence
hnt	10	gacgatgggtggcaaagatc	*hnt*	34	caacagatcggtgttcttgt
hnt	11	gtaagcgacaacagtgcgac	*hnt*	35	ccaaagtgcaggagaggatt
hnt	12	tgtcttgtgcttcagataca	*hnt*	36	ctcatgatcatgctcgtttt
hnt	13	ttcaggatgctctggatatc	*hnt*	37	ggctgctcatctagaatcaa
hnt	14	aaactggaactggagctggt	*hnt*	38	cttgagcgatcatcgttttc
hnt	15	ttcacggactgctcaaagtt	*hnt*	39	acttctttgagggattttcc
hnt	16	tgtactgactcgaattgggc	*hnt*	40	atccttgaacggtgaactgg
hnt	17	ggcacgaaggttcgggaaaa	*hnt*	41	tggcaatacggacaggagac
hnt	18	aaatgtacccgattgtgacc	*hnt*	42	tcttggtggtgaacaacagc
hnt	19	aaacggcgtagggacacatg	*hnt*	43	cagcgactcagtttcaatgg
hnt	20	tgaaggcatagttgcacacg	*hnt*	44	gcaagcagaaggcacagata
hnt	21	ctcagatgacgttcacagtt	*hnt*	45	tgtgcttcagtgtgaacttc
hnt	22	gtgggactgatcgaacggaa	*hnt*	46	tcactgatgcccaatagatt
hnt	23	ctgggatttgctctcattta	*hnt*	47	tactccagtatttatggctc
hnt	24	gctgagatccaaaacatcca	*hnt*	48	ctttggtatacgtaaggcgg
ush	1	tctgcggtatcggaacaatc	*ush*	25	gtccagggaagtgtacttaa
ush	2	atctttggaatctctgctgt	*ush*	26	ggggcaatagtagttctgat
ush	3	ggaatcgttcaactgatcct	*ush*	27	ggacaaggcagaccaatttc
ush	4	cgaactcagcatcttcatca	*ush*	28	acaggacacttgttcagact
ush	5	tgaagatcgtgttcctgttc	*ush*	29	aatgggttcgaatgtgggtg
ush	6	acagggcaggcacataaatc	*ush*	30	agatgcaggtcatgtatagc
ush	7	ttgagggcgaactgaatgca	*ush*	31	agggatgatccatggaatcg
ush	8	tccgtgtccttaattctatg	*ush*	32	aggtggagacatagttgcag
ush	9	catgtggcgatttagggata	*ush*	33	agtcacatcttgatctctgg
ush	10	ctattccgttctggagaaga	*ush*	34	accgttggatacatcggaat
ush	11	cggtttgttgactagagcta	*ush*	35	gagcagtacttcttcatcac
ush	12	cttgatgttgttgaaccgga	*ush*	36	caggtaggtcttcacgtaat
ush	13	cgaactgcagtagtgttgtt	*ush*	37	tgcagtagaactgcttgtga
ush	14	aaaggtggatgcgaatgcgg	*ush*	38	cgtgtaataccactcaagct
ush	15	acaattatgatcgggtgggt	*ush*	39	ggaggctaggattcgattag
ush	16	tgggcctggaataaagctag	*ush*	40	ttgcggtttggatagtgtac
ush	17	cagattctctagttaaccct	*ush*	41	ggagattttccgggaatact
ush	18	aagctcagtgaattcagggt	*ush*	42	tcggtaagtcgaaggagtca
ush	19	cttccacatctaagagagca	*ush*	43	tgtgttattcctttaggtgt
ush	20	gagggagcgcaaacgatttg	*ush*	44	gtaagctttggcatgcatta
ush	21	actactagtttgggcaggaa	*ush*	45	gccctcaatttaatttctgt
ush	22	cccttttcacatagatctg	*ush*	46	cgcagaccattgcaaacttg
ush	23	acaatattgcactccatgca	*ush*	47	agaattgctcgctttatggg
ush	24	tgccaagtagttctcatact	*ush*	48	cttttattgtcgcacacact
tup	1	caatctctgccattaccat	*tup*	24	gggtcgcttgtccctaata
tup	2	tagctggtgagccaaatgg	*tup*	25	cattgagcacggtccgaac
tup	3	ttgtggttgtgattgtgca	*tup*	26	tgagcgtatgcagctgttt

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Gene	Probe	sequence	Gene	Probe	sequence
tup	4	cgatggctgtaatccagtt	*tup*	27	gcggattagcattgtagca
tup	5	ccaaatccaggtgatggtt	*tup*	28	tccttcatgagcgcatcag
tup	6	tatcacatcgtgtccatgg	*tup*	29	gatgactcgcggcgacagg
tup	7	catccgacgcagtgagata	*tup*	30	cgcttgttctggaaccata
tup	8	tactgatcgtggatctggc	*tup*	31	aatggtcttcttcttgtcc
tup	9	actccagatcgggggcaac	*tup*	32	ctgcatttgcagcttcatc
tup	10	tcctggcacttcagacacg	*tup*	33	aactggcgatcatggggat
tup	11	tacagctttcgtccaggaa	*tup*	34	cctgaagattcagtgggga
tup	12	tgccatcgcgcacaaaaca	*tup*	35	cggttgatatgtctgcaca
tup	13	aatcacgcttgcagtaggt	*tup*	36	aagtcgcttaaggctttcc
tup	14	ccgcatttatcacattttg	*tup*	37	gtcgagatcggcgtgaagg
tup	15	tcatttttgctgaaggagt	*tup*	38	ggtattgatcgctccattg
tup	16	aagattttcgttttggccc	*tup*	39	aactgctgaaatgcgggcg
tup	17	ggagcatcgaaagcactcg	*tup*	40	catcccgttcaggtcgtag
tup	18	cagcaattgtcgcgcacac	*tup*	41	cggatgcggcggcagaatg
tup	19	gcatcgcgtaacgcgaatt	*tup*	42	ctgctggttctgatggttg
tup	20	tcctccttgcagtacaaag	*tup*	43	gagtccaaactgctgcctc
tup	21	acgatttctccagcacatc	*tup*	44	atggtggtgcgaggtgatc
tup	22	ccgacgaagaagtgaggct	*tup*	45	gacgtaggagtcggtgctg
tup	23	tttgtgtgagcccgattcg	*tup*	46	gtcatcgctctccaggtag
smiFISH probe sets					
hnt	1	gttcacgcacaaatcacaga	*hnt*	25	ttcattttggttggctgatt
hnt	2	gcgctcgattatttatccaa	*hnt*	26	tttgttgtccattggctttc
hnt	3	acttttccaatttcatcctt	*hnt*	27	atttctaccgataacgagcc
hnt	4	tgttagcttagctgaagacc	*hnt*	28	tgaaacctcagaggctgatc
hnt	5	acactgctgcacttcaattg	*hnt*	29	cccccgaaaggcaacaaaaa
hnt	6	ttccttttgatttaatcgga	*hnt*	30	gtggtagtggatcgatttga
hnt	7	gggttttgggaatggacttg	*hnt*	31	ctatgactatgagcggtgtc
hnt	8	ttgacccaaaacaccaaggg	*hnt*	32	cactcattagcatttaccgt
hnt	9	atttttgacccggctaatt	*hnt*	33	gtgtgggtgaaaaatgtggg
hnt	10	gttttagcatcggtcatggg	*hnt*	34	gctgctcttatcaagaattt
hnt	11	cgtggtgtgtcaaagttcaa	*hnt*	35	aattcaatacacatctggca
hnt	12	ttttcggactggactgtctg	*hnt*	36	ccgcagaaaggtgcgaagaa
hnt	13	cctgaccacggaaatttctg	*hnt*	37	aattggcgactgttgcaagt
hnt	14	ctataaacgtcgcctcatca	*hnt*	38	agttcctgtgaagagcgaat
hnt	15	cccctgcgatattaaacaga	*hnt*	39	aacagctgattcaggcaagt
hnt	16	aaccgtttgacatggagcac	*hnt*	40	gtcttcgattgtcgcgtaat
hnt	17	cggaatcaagccgagtgaag	*hnt*	41	aaaatcgatagtggggcacc
hnt	18	caaatttctaatcgccatct	*hnt*	42	aactgggtgtagggttcaat
hnt	19	gaacacaacttggtcagtga	*hnt*	43	atgatatccacactgacagc
hnt	20	atgacccttgggaattatca	*hnt*	44	gcttctttcgatattcct
hnt	21	aattgttgggcggttttcac	*hnt*	45	agtcctagcacatctgtc
hnt	22	atatatacttttggggcgg	*hnt*	46	aacctacggctagctctt

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Gene	Probe	sequence	Gene	Probe	sequence
hnt	23	tgagataaagattccctccc	*hnt*	47	aatcaggcgatcaggggt
hnt	24	cgttgtccatgagtcatgaa	*hnt*	48	ggatggtgggagaaacga
ush	1	attgctagtttgcttttctt	*ush*	25	ccatttcgtttaagctgcaa
ush	2	ttgtaattggcatcatcgcg	*ush*	26	cagataaattcccatccact
ush	3	agcaccactcataactttgc	*ush*	27	aggagggcatgccaaaagaa
ush	4	atcggctaaacgtggcttaa	*ush*	28	taacccaactagccgattag
ush	5	aggctctacgaaattgagcg	*ush*	29	acaaacatcaaggggcggac
ush	6	acatgacgtgcaggttgatt	*ush*	30	tctgcaagataagccgaagc
ush	7	attttcagtttttgcagtcc	*ush*	31	atgcttcccgcataaacaag
ush	8	actacagtgtcggacaagtc	*ush*	32	gtcgatgtttttgttgttgc
ush	9	cgaaatgcggacttggactt	*ush*	33	tgaaatcccgacaccacaat
ush	10	cgcagataggcgactgataa	*ush*	34	ccaatatttcatcgcaaccg
ush	11	aactgaggccgtggaatgaa	*ush*	35	gctattttagctatccacac
ush	12	atgcagatacatatgagccg	*ush*	36	gactttccttcgacttgatg
ush	13	gcacacacagatactcgtac	*ush*	37	gggcatagaattcatttcca
ush	14	tcaccgcagatcgtagatag	*ush*	38	ggcaacttggagatgatgga
ush	15	tccttcgttttaattcagct	*ush*	39	acacgaaggcagaggatgtg
ush	16	gctttgtgttgtttttctca	*ush*	40	ggagccgaaaggaggcaaaa
ush	17	aactacagtacccagagtac	*ush*	41	taggccccgaaaaaagtgtg
ush	18	gtttactacgcacagtagtt	*ush*	42	agacacatcattacctacca
ush	19	ccagtatctttgcatctttt	*ush*	43	ggcttgcactaagttgaagt
ush	20	cattgtttgcttttgtccac	*ush*	44	aaataaaggaccccacagca
ush	21	ttccagccaaatggttttac	*ush*	45	attcccaaaaagaggggagc
ush	22	tttatgcatagtccggacta	*ush*	46	taatctgccatcgactgttg
ush	23	gtagtttgaagcgaataccc	*ush*	47	atcaaggaaaagaaggcccc
ush	24	gaggtttgcttgttcatttc	*ush*	48	gaactactgtctccgaaacc
tup	1	gcatttgttaacacggcttt	*tup*	25	cgaattgagatgcggccg
tup	2	gttagacaactgtcagccaa	*tup*	26	gggcaatcatcaggggttta
tup	3	cgactacttgagaagttgcc	*tup*	27	ggaaattagataatgccccc
tup	4	agcgatctcggacttacata	*tup*	28	tttcgaatgagagttgggcc
tup	5	agcattcggtgaatttgctg	*tup*	29	tgcctagggctgagcaaata
tup	6	tttcactgacgtttcagcag	*tup*	30	caagcgcatgacaccgagtg
tup	7	acaaacctagcgctatgact	*tup*	31	tttgccgaacgctacaattg
tup	8	accactgtcgtattgttttt	*tup*	32	ctaacagaacagcggcctaa
tup	9	gtttttatgtggcgaacgga	*tup*	33	aattggctgggctaacaagc
tup	10	attttcaagcacccactatt	*tup*	34	ctttttgtattggcatcagc
tup	11	atatttttctacagccgagc	*tup*	35	ggaccagctgttgattagaa
tup	12	gaaatgcaaaacgctcgcca	*tup*	36	tgcacgtcatacatttagcg
tup	13	aacgacgccattaaagtggc	*tup*	37	cgatgcaaacacaccgattt
tup	14	actcataaagccagcgacta	*tup*	38	tgcatggcaaatagttgcga
tup	15	agtatttgaagtcctcatcc	*tup*	39	agctttgcacgtgttaatca
tup	16	atttttttggtcggattgtgc	*tup*	40	catcaatcatccgagcgaac
tup	17	agatgtttgagctggcttag	*tup*	41	atcgatgagtgagccgagtg

bioRxiv preprint doi: https://doi.org/10.1101/837179; this version posted November 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Gene	Probe	sequence	Gene	Probe	sequence
tup	18	ttaggcagctagaatgcttc	*tup*	42	aaattcttttgtcagtgccg
tup	19	gatctgcaaactcctcaact	*tup*	43	cagagcttgtttatggctat
tup	20	cgaaaaaaggcgggcagtcg	*tup*	44	acacaatttcaagggggtg
tup	21	caaagagccgaagagctctg	*tup*	45	acagccttgaatgagttcac
tup	22	acagtttctcagtttctgtc	*tup*	46	gaagatcccatgattggagc
tup	23	gtacctgattcagatccaac	*tup*	47	ggtagtatttcccaaatcga
tup	24	acatttttgcattgctgtcg	*tup*	48	accagaatcctctttgtttt