Supplementary Information for

"Scaling dictates the decoder structure"

By Jingxiang Shen, Feng Liu, and Chao Tang*

Content.

Supplemental text.

S1. Shape of the maternal morphogen profiles.

S2. Modelling length fluctuation and noisy morphogen gradients.

S3. Fitting the linear classification planes.

S4. Making predictions with the set of linear classifiers.

S5. The Bayesian decoder.

S6. Discussion on some of the predictions in Fig. 3d.

S7. *bcd*$^-$ and *bcd6X* embryos with reduced lengths.

S8. The three maternal gradients in *Drosophila* function as two bi-gradient pairs.

S9. Shift of the *eve* stripes under Bcd dosage change.

S10. Optimal decoder for noisy gradients without embryo length variation.

S11. Implementing the scaling decoder with a dynamical gene regulation model.

S12. Nos/mHb as the posterior gradient.

S13. Discussion on the long-germband insect *Megaselia abdita*.

Supplemental figures.

Fig. S1. The mHb profile can be fitted by a sigmoidal curve.

Fig. S2. Simulating noise in the Bcd gradient by a Poisson noise term.

Fig. S3. Predicting mutant fate-map with linear classifiers.

Fig. S4. Effects of the Bcd β factor and noise amplitude on model performance.

Fig. S5. The Bayesian decoder.

Fig. S6. Outputs of the scaling decoder on mHb=0 and mHb=mHb$_0$ planes.

Fig. S7. Maternal morphogen mutant embryos with greatly changed length.

Fig. S8. The three maternal gradients function as two bi-gradient pairs.

Fig. S9. Shift of the *eve* stripes under Bcd dosage change.

Fig. S10. Optimal decoder for noisy gradients with or without length variation.

Fig. S11. The ODE-based gap gene regulation model.

Fig. S12. The scaling gene circuit model.

Fig. S13. Generating scaling *Drosophila* gap gene pattern with a more extended Bcd profile.

Supplemental tables.

Table S1. Peak and boundary positions extracted form maternal morphogen mutants.

Table S2. Parameters for the scaling ODE model.

Table S3. Parameters for the scaling gene circuit model.

42 **S1. Shape of the maternal morphogen profiles.**

43 **Bicoid (Bcd).** The Bcd concentration gradient is generated by diffusion form a
44 localized source. At steady state, its profile should be exponential, which is fully
45 consistent with experiments (therefore, we do not consider the possibility raised in
46 Ref. [1]):

47 $$Bcd(x) = e^{-\frac{x}{\lambda_B}}.$$

48 Its absolute length constant λ_B is fully determined by the diffusion constant D and
49 decay rate γ ($\lambda_B = \sqrt{D/\gamma}$), independent of embryo length [2-5]. Reformulating the Bcd
50 profile using the "relative" coordinate y, which is normalized by embryo length $y \equiv$
51 x/L, yields:

52 $$Bcd(y) = e^{-\frac{yL}{\lambda_B}} = e^{-\frac{y}{\lambda_B/L}}.$$

53 For larger embryos ($L>1$), the length constant appears to be shortened in the
54 normalized coordinate.

55

56 Throughout this paper, the length unit is chosen to be the length of a "standard size"
57 embryo L_0 (~490 μm). Therefore, the position x, embryo length L, and λ_B, are
58 dimensionless (normalized by L_0). Measured in this way, the Bcd length constant
59 λ_B=0.165, according to a very carefully performed quantitative measurement [6].

60

61 Though the Bcd length constant λ_B cannot scale with embryo length, a positive
62 correlation between Bcd amplitude (absolute concentration at the anterior pole) and
63 embryo length L has been observed experimentally [5,7]. i.e., larger embryos tend to
64 have higher overall Bcd dosage. To reflect this fact, an amplitude factor L^β is
65 introduced in the Bcd term. Taken together,

66 $$\mathrm{Bcd}(y, L) = L^\beta e^{-\frac{y}{\lambda_B/L}} \qquad (S1.1)$$

67 The exponent β should lie between 2 and 3 according to a not very precise
68 measurement [8]. Being somewhat conservative about this "amplitude correction" effect,

69 we take β=2 throughout the main text. No matter if β=3 (Fig. S4d). Our model is not
70 quite sensitive to the exact value of β.
71
72 This effect produces a "neutral point" $y=\beta\lambda_B$ with invariant Bcd concentration and is
73 proposed by some authors that is may contribute directly to scaling of the gap genes.
74 As mentioned in the main text, we don't agree with this explanation in general.
75 However, in our framework, although the optimal scaling decoder can always be
76 well-defined with or without this L^β factor, the exact Bcd profile affects the exact
77 orientations of the optimal decision planes. Therefore, to accurately describe the real
78 situation in *Drosophila* (hence making correct predictions on mutants), this Bcd
79 amplitude effect should not be ignored (Fig. S4b, c).
80
81 **Maternal Hb (mHb).** Like Bcd, the posterior gradient Nos should have an
82 exponential profile. (No amplitude correction effects are reported experimentally, so
83 the L^β factor is not added).
84
$$Nos(y) = e^{-\frac{1-y}{\lambda_N/L}}$$
85 It is well known that Nos functions solely through repressing the maternal component
86 of the gap-gene protein Hb (mHb) in posterior half of the embryo [9-11]. Therefore, the
87 "immediate" posterior morphogen should be mHb instead of Nos. If we assume that
88 mHb level is dictated by Nos through an inhibitive Hill function,
89
$$mHb(y) = \frac{mHb_0}{1+\left(\frac{Nos(y)}{K}\right)^n}$$
90 then the mHb gradient takes a sigmoidal shape:
91
$$mHb(y) = \frac{mHb_0}{1+e^{\alpha L(y-1+(1-\lambda_H)/L)}} \qquad (S1.2)$$
92 Where $\alpha = n/\lambda_N$, $\lambda_H = 1+\lambda_N \ln K$. Although the raw parameters (λ_N, K, and n) are
93 unknown, measured mHb profiles (Hb protein profile in n.c.12 embryos, from the
94 *FlyEX* database [12,13]) can be well fitted with this sigmoidal curve (with α=15 and
95 λ_H=0.425). See Fig. S1 for the fitting. Note that in this paper we normalize the value

96 of Hb (no matter maternal or zygotic) according to its maxima at n.c.14, so mHb has
97 an amplitude coefficient mHb_0=0.1.
98
99 **Torso (Tor).** The activity of Tor is induced at both terminus of the embryo by its
100 ligand in the perivitelline space [14,15]. Tor transduces the activation signal into the
101 syncytial embryo by phosphorylating ERK. Phosphorylated ERK (dpERK) diffuses in
102 the cytoplasm, trapped by the nucleus, and dephosphorylated (i.e. "degraded") inside
103 the nucleus [16]. This is a similar picture of the "localized synthesis, diffusion, and
104 decay" model of Bcd and Nos. So, it is reasonable to assume that the activity of Tor
105 has an exponential profile as well.

106 $$Tor(y) = e^{-\frac{y}{\lambda_T/L}} + e^{-\frac{1-y}{\lambda_T/L}} \qquad (S1.3)$$

107 Quantitative measurements on dpERK indeed show double-exponential profiles
108 (when projected onto the one-dimensional anterior posterior axis) [16,17]. In this paper
109 we use an estimated value λ_T=0.07 for its length constant.
110
111 These equations give Eqn. 5a-c in the main text.
112
113 Precise measurements on the scaling property of mHb and Tor is still lacking and very
114 difficult to perform. Our assumption that they should both be unscaled with embryo
115 length is the minimal assumption. This minimal assumption is consistent with the
116 known mechanisms through which the gradients are established, and is also supported
117 by the case studied in Fig. 5a. See SI-7 for detail.
118

S2. Modelling length fluctuation and noisy morphogen gradients

The WT point cloud used for fitting the linear classification planes are defined as follows. First, the A-P axis is discretized into 101 points y=0% to 100%. For each of the y position, we sample 400 embryo length values from the normal distribution $L \sim N(1, 0.1)$ and calculate the corresponding noise-free (Bcd, mHb, Tor) levels using Eqn. 5. This three noise-free values are noted by $\boldsymbol{m}=(m_1, m_2, m_3)$ for convenience. Obviously, $0<m_{1,3}<1$ and $0<m_2<0.1$. This give the 2-d "WT manifold" in Fig. 2c.

The fact that embryo length L only fluctuates within a limited range is important. Outside certain L range, the decoder behavior should not be subjected to selection pressure since embryo size hardly fluctuate that much under natural conditions. Thus, in the strictest sense, only within the region covered by *realistic* WT embryos, the effective input-output relation of the decoder should follow that dictated by scaling.

Secondly, a Poisson noise is added to each morphogen value m_i by assuming the actual number of molecules is a Poisson variable n_i with $\langle n_i \rangle = N*m_i$, and the final (normalized) morphogen level with noise is $m_i=n_i/N$. The (hypothetical) maximum molecule number N controls the noise magnitude. We set N=1000 throughout the main text. From a theoretical perspective, the noise terms turn the 2-dinemsional WT manifold (Fig. 2c) into a 3-dimensional WT point cloud (Fig. 2d).

Note that N=1000 does not correspond to the number of molecules per nucleus (which result in the intrinsic noise). Instead, N=1000 is chosen to make the positional error of modeled Bcd gradient close to that measured by Ref.[18], which included both intrinsic and extrinsic noises.

To be specific, positional noise (standard deviation σ of position y) of the Bcd gradient in standard-sized (L=1) embryos can be expressed as

147 $$\sigma_y \left| \frac{d\bar{N}_{Bcd}}{dy} \right| = \sigma_{N_{Bcd}} \qquad (S2.1)$$

148 For Poisson distribution: $\left(\sigma_{N_{Bcd}}\right)^2 = \bar{N}_{Bcd} = Ne^{-y/\lambda}$. Thus:

149 $$\sigma_y = \frac{\lambda}{\sqrt{N}} e^{-\frac{y}{2\lambda}} \qquad (S2.2)$$

150 Substituting λ=0.165 and N=1000 into this equation, gives the modeled positional
151 noise as the black curve in Fig. S2, overlapped on the experiment results of (Gregor,
152 et. al. 2007).

153

154 The measured results on Bcd noise should be reliable in the region 0.2<y<0.6, since it
155 is far enough from Bcd mRNA is distribution, also, Bcd protein level here is high
156 enough to be safe from experimental detection limit. In this region 0.2<y<0.6,
157 strengths of Bcd noise introduced by our Poisson term are close to the measured
158 values. (Our model is not quite sensitive to the exact value of *N*, see Fig. S4 E-F for
159 the results when *N*=500 or 2000.)

160

161 Note that "Bcd noise" here should stand for the measured embryo-to-embryo
162 fluctuation in an ensemble of standard-sized (*L*=1) embryos. i.e., it accounts for both
163 intrinsic noise (finite number of Bcd protein molecules per nucleus), and extrinsic
164 noise (e.g., embryo-to-embryo variability in overall Bcd amplitude) except for length
165 variation. Theoretically, intrinsic and extrinsic noises are different, in that extrinsic
166 noise is correlated for nucleus belonging to the same embryo. However, since our
167 decoder works in a *spatially decoupled* manner, it cannot distinguish whether two
168 different (Bcd, mHb, Tor) points come from the same embryo or not. In other words,
169 only the overall strength of fluctuation matters, no matter the fluctuation comes from
170 intrinsic molecular noise or embryo-to-embryo variation. The decoder deals with all
171 nucleus in all embryos of the *Drosophila* species simultaneously. Therefore, intrinsic
172 and extrinsic noises are not treated separately here.

173 **S3. Fitting the linear classification planes.**

174 The entire point cloud in Fig. 2d consists of 101 subsets $\boldsymbol{m}|_y$, each of them have 400
175 points. The 100 classification planes locate at *y*=0.5%, 1.5%, …, 99.5%, numbered as
176 classifiers #1, …, #100 (and we only consider #6 to #95 in the main text). Each of the
177 planes should perform the *local* classification task of distinguishing ***m*** points belongs
178 to adjacent *y*'s.

179

180 For example, the plane #3 locating at *y*=2.5% should first go through the ***m*** point
181 representing the *noise-free* morphogen levels at *y*=2.5% in *standard-sized* WT by
182 itself. Secondly, the plane orientation is defined by that can best distinguishing the
183 point classes $\bigcup_{y=\{0\%,1\%,2\%\}} \boldsymbol{m}|_y$ against $\bigcup_{y=\{3\%,4\%,5\%\}} \boldsymbol{m}|_y$. To find the best-fit
184 plane orientation numerically, we simply enumerate the Euler angles θ and φ of its
185 normal vector at the resolution of 1° and find the one with the highest classification
186 accuracy.

187

188 The noise due to finite sampling and discretizing θ and φ are eliminated by averaging
189 the classification plane orientations for 25 repeats of the above sampling and fitting
190 steps.

191

192 **S4. Making predictions with the set of linear classifiers.**

193 The portion of (Bcd, mHb, Tor) space where the decoder output is directly dictated by
194 scaling (that is, the region covered by the WT point cloud) does not include all the
195 situations in the morphogen mutant embryos. Extrapolations are therefore needed for
196 making predictions in general. Fortunately, for the *Drosophila* case most mutants of
197 interest lie not far away from the WT point cloud. Thus extrapolations could make
198 sense here.

199

200 We think the most simple and natural assumption is to extrapolate linearly with the
201 classification planes defined above.

202

203 Firstly, we know from Fig. 6c that the decision boundaries realized by gene
204 interaction network can only follow the scaling requirements to the linear order in
205 general, thus starting only from scaling we simply have no information about possible
206 higher-order features.

207

208 Secondly, the unstable manifold of a bi-stable diagram tends not to have large
209 curvature except in the neighborhood of a critical point (Fig. 6b). Although the real
210 gap gene network is not a simple bi-stable system, we think this intuition should still
211 hold.

212

213 Finally, in Fig. 6d-f we presented a differential-equation-based gene regulation model
214 using the known gap gene interaction network. This kind of model indeed extrapolate
215 in a very much linear way.

216

217 We next describe how the linear extrapolations are carried out precisely. Obviously,
218 100 well-separated planes should divide the morphogen space (the cube with $0<m_1<1$,
219 $0<m_2<0.1$, $0<m_3<1$) into 101 slices, corresponding to $\tilde{y}$=0% to 100%. If a query $\boldsymbol{m}$

220 point falls into the slice $\tilde{y}$=n%, it locates on the posterior side of classification planes

221 #1 through #n, and on the anterior side of planes #n+1 to #100. Therefore, the

222 corresponding cell fate $\tilde{y}$ can be read out from the classification results of all the

223 linear classification planes. Consider the y=0.4 point in an L=1 *bcd⁻tor⁻* embryo,

224 whose $\boldsymbol{m}$=(0,0.059,0) (according to Eqn. 5, shown as black cross in Fig. S3a). This

225 point lies in the $\tilde{y}$=0.56 slice, i.e., between the classification planes #55 and #56.

226

227 In some other situations, however, the 100 linear classifiers could have contradictory

228 outputs. Say, the point (0.004, 0, 0.24) are classified to the posterior side by classifier

229 #70 but to the anterior side by #30. We introduce a "posterior dominance rule" to

230 tackle this difficulty. Anytime when this happens, output of the anterior classifier (#30

231 here) is always ignored. The reason for us to introduce the posterior dominance rule is

232 simple – some anterior classification planes may intersect with the much more

233 posterior region of the point cloud, vary far from where they were fitted (Fig. S3b).

234 This posterior dominance rule works well, yielding the results in Figs. 2-5.

235

236 To be more precise how the fate map curves in Fig. 3 are obtained, we explain here

237 that which "fate slice" a query point should fall is determined by analyzing the

238 classification outputs by all the 100 classification planes. In the ideal case, $\tilde{y}$=n%

239 means that classifiers #1 through #n output "posterior", and classifiers #n+1 to #100

240 output "anterior". Graphically, the decoder outputs are recorded by a column of pixels,

241 with grey stands for "anterior" and white stands for "posterior", and the fate map in

242 Fig. 3 is then given by extracting the grey-white boundary of Fig. S3c.

243

244 Also note that those classifier outputs ignored by the posterior dominance rule are

245 shown in lighter color in Fig. S3c. Graphically, this rule is equivalent to wiping out all

246 the grey pixels once there exists a white pixel above them. The "posterior dominance

247 rule" is a quite rough rule after all. Sometimes it leads to artifacts near the embryo

248 terminus. For example, the fate map of *bcd1X* (and *vas$^-$exu$^-$bcd6X*) in Fig. 3 shows
249 abrupt jumps near the anterior (and posterior) end. However, this kind of error do not
250 affect prediction in most cases.

251 **S5. The Bayesian decoder.**

252 With our settings, given position y and embryo length L, the morphogen levels $\boldsymbol{M}$
253 satisfy the Poisson distribution:

254
$$p(M_i|y,L) = \frac{\left(Nm_i(y,L)\right)^{M_i}}{M_i!} e^{-Nm_i(y,L)} \quad (i = 1,2,3),$$

255 where m is the *noise-free* morphogen profile defined in Eqn. 5 in main text, and N is
256 the "effective maximal molecule number" as described above. Also, noise on different
257 morphogens are assumed to be independent:

258
$$p(\boldsymbol{M}|y) = \int dL\, \rho(L) \prod_{i=1}^{3} p(M_i|y,L),$$

259 where the embryo length follows the Gaussian distribution $L \sim N(\mu=1, \sigma=0.1)$

260
$$\rho(L) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(L-1)^2}{2\sigma^2}}$$

261 Inferring y from the morphogen levels $\boldsymbol{M}$ can be performed using the Bayes formula

262
$$p(y|\boldsymbol{M}) = \frac{p(\boldsymbol{M}|y)\, p(y)}{p(\boldsymbol{M})}$$

263 Since y is uniformly sampled from 0 to 1, the Bayesian decoder becomes a maximal
264 likelihood decoder.

265
$$\underset{y}{\operatorname{argmax}}\, p(y|\boldsymbol{M}) = \underset{y}{\operatorname{argmax}}\, p(\boldsymbol{M}|y)$$

266 Decoding results for the WT point cloud of Fig. 2d by the Bayesian decoder is shown
267 in Fig. S5a. Although arbitrary decision boundary geometry is allowed by the
268 Bayesian decoder, its Root Mean Squared Error (RMSE) is even larger than the linear
269 decoder presented in the main text (Fig. S5b). This is conceivable, as maximization of
270 posterior likelihood could lead to minimal regression error only when the decoding
271 error is Gaussian, which not satisfied here. Therefore, we claimed in the main text that
272 the remaining classification errors of the linear decoder should due to the morphogen
273 noise rather than nonlinearity in classifier geometry. This can be visualized by
274 choosing a tangential view of the classification planes in Fig. 2D-E (Fig. S5d).

275

276 In comparison with the Bayesian decoder, there are some additional arguments on our
277 linear extrapolation hypothesis. As expected, decision boundaries of the Bayesian
278 decoder are effectively linear within the WT point cloud (Fig. S5c, on the alpha plane).
279 However, situations outside the WT point cloud are quite different: There, outputs of
280 the Bayesian decoder are determined by extremely improbable cases, say very large L
281 variation or very large fluctuation in morphogen level, which are basically irrelevant
282 to realistic embryos. While the linear classifier emphasizes more on simplicity of the
283 decision boundary geometry, and turned out to be a better way of extrapolation.
284

285 **S6. Discussion on some of the predictions in Fig. 3d.**

286 (***mhb***$^{-}$). Among the mutant cases in Fig.3d, we do not have quantitative data for *mhb*$^{-}$,
287 but we are quite confident that this prediction is correct. It has long been known that
288 although *nos*$^{-}$ embryos (where mHb is uniformly high) lack all abdominal segments, it
289 can be largely rescued by further eliminating mHb completely. The *nos*$^{-}$*mhb*$^{-}$ double
290 mutation embryo is viable and has basically normal morphology [9,10]. This observation
291 led people to discover that for WT the only function of Nos is to inhibit mHb, but it
292 also shows that even if the morphogen mHb does not exist at all, the embryo should
293 not miss any segment. Our prediction in *mhb*$^{-}$ matches this observation. The predicted
294 fate map is almost along the diagonal albeit slight distortions. Fig. S6 provides a
295 graphical visualization on the case of *nos*$^{-}$ and *mhb*$^{-}$.

296

297 (***vas***$^{-}$***exu***$^{-}$***bcd6X***). The *vas*$^{-}$*exu*$^{-}$*bcd6X* embryo lacks *nos* and has flattened and overall
298 increased Bcd gradient [19], and the measured *bcd* profile is used for this prediction.
299 This mutant display mirrored head structures near the posterior pole. The yellow
300 squares marked in Fig. 3d *vas*$^{-}$*exu*$^{-}$*bcd6X* panel represent the expression peaks and
301 boundaries of the head gap genes *otd*, *btd*, and *ems*, measured by Ref.[19].

302

303 (***bcd1X*** **and** ***bcd4X***). *bcd1X* or *bcd4X* here means the situations where Bcd dosage is
304 halved or doubled exactly, not the actual *bcd* copy number. Hb boundary in *bcd1X* or
305 *bcd4X* is predicted to shift by -7.3% or +9.1% by our scaling decoder, and the
306 experimentally measured shifts are -6.4% and +9.4% according to Ref. [6].

307

308

309

310 **S7.** ***bcd***

338 and viii) are kept at fixed distances from the anterior pole by reading the anterior Tor
339 gradient. In between is the *Kr* domain, which shrinks with *L*.
340
341 Using the *fat2RNAi* technique, Ref. [7] also studied the gap gene patterns in *bcd6X*
342 embryos under length change. Note that with 6 copies of *bcd*, the resulting Bcd
343 protein dosage was actually only be approximately doubled [20], rather than multiplied
344 by 3. This knowledge is consistent with the position of cephalic furrow reported in [7],
345 that CF locates at y=0.42 in the *bcd6X* embryos, very close to those Bcd dosage ≈ 2.2
346 embryos reported by Ref. [6]. So, Bcd dosage is set to 2.2 in our model to simulate
347 these *bcd6X* embryos. Scaling is not preserved in these embryos – that the gap gene
348 domain boundaries shift significantly when *L* changes (Fig. S7d). This is the expected
349 result, as according to our basic assumption, scaling stems from cancelation of the
350 first-order effects of morphogen level difference due to a change of *L*. Those
351 first-order derivatives are different for the altered zeroth-order profiles, thereby failed
352 to cancel each other out.
353
354 In Fig. S7d, the measured boundary positions by [7] are shown as dots, while our
355 predictions are the solid lines (no "shrinking ratio" is assumed here). Errors between
356 predictions and the experiments are generally acceptable, except for some of the
357 boundaries – namely, both boundaries of the *kni* domain, and the anterior boundaries
358 of the posterior *gt* and *hb* domains. Note that these errors could be corrected by more
359 careful data processing. And the observation that the *kni* and posterior *gt* domains
360 may disappear in short enough embryos [7] does not seem to be captured by our
361 extrapolation based predictions.
362
363 Especially, shift of the mid-embryo *hb* domain can be studied analytically, because
364 the effect of Tor is nearly negligible in the central region. Profiles of Bcd and mHb
365 when Bcd dosage is 2.2 are as follows.

366 $$\begin{cases} \mathrm{Bcd}(y', L') = 2.2\, L'^{\beta} e^{-y'L'/\lambda_B} \\ \mathrm{mHb}(y', L') = \mathrm{mHb}_0\left(1 + e^{\alpha L'(y'-1+(1-\lambda_H)/L')}\right)^{-1} \end{cases} \quad (S7.1)$$

367 Assume that there exists a WT embryo of length L with perfectly scaling gap gene
368 pattern, in which the y=0.473 position (*hb* boundary) has identical Bcd and mHb
369 values as the above equation. i.e.,

370 $$\begin{cases} L^{\beta} e^{-yL/\lambda_B} = 2.2\, L'^{\beta} e^{-y'L'/\lambda_B} \\ \mathrm{mHb}_0\left(1 + e^{\alpha L(y-1+(1-\lambda_H)/L)}\right)^{-1} = \mathrm{mHb}_0\left(1 + e^{\alpha L'(y'-1+(1-\lambda_H)/L')}\right)^{-1} \end{cases}$$

371 Relationship between *y'* and *L'* can be easily solved by eliminating the unknown *L*:

372 $$L' = \lambda_B \frac{\ln 2.2 - \beta \ln \frac{1-y'}{1-y}}{1 - \frac{1-y'}{1-y}} \quad (S7.2)$$

373 This analytical result is compared with experiments in Fig. S7e. Note that this
374 prediction not even depends on the precise profile of mHb.
375

376 Similar to Fig. 5a-d, we present predicted gap gene domain positions of many other
377 maternal morphogen mutants with greatly changed embryo lengths in Fig. S7f. Some
378 of them may be tested by experiments in the future.
379

380 **S8. The three maternal gradients in *Drosophila* function as two bi-gradient pairs.**
381 The “tri-gradient system” of Bcd, mHb, and Tor can largely be decomposed into two
382 parallel bi-gradient systems: Bcd&mHb in the middle part, and Bcd&Tor near both
383 termini.
384
385 The contribution of each morphogen to scaling can be characterized as follows. From
386 the results shown in Fig. 5a, c, d, we can calculate the derivative of the position y
387 where a certain cell fate $\tilde{y}$ appears with respect to embryo length L (evaluated at
388 L=1.0), defined as the size sensitivity S_L.
389
$$S_L \equiv \lim_{\Delta L \to 0} \left| \frac{\Delta y}{\Delta L} \right| \qquad (S8.1)$$
390 With the fitted decoder classification planes, S_L can be computed as follows. Let
391 (Bcd(y), mHb(y), Tor(y)) be the morphogen levels at y in an embryo (WT or maternal
392 morphogen mutant), the corresponding cell fate is $\tilde{y}(y)$. Since we would like to track
393 the position where $\tilde{y}$ is fixed, the total shift in the (Bcd, mHb, Tor) space caused by
394 ΔL and Δy should be perpendicular to the normal vector $\boldsymbol{K}$ of the local decision plane:
395
$$\left(\Delta L\left(\frac{\partial Bcd}{\partial L}, \frac{\partial mHb}{\partial L}, \frac{\partial Tor}{\partial L}\right) + \Delta y\left(\frac{\partial Bcd}{\partial y}, \frac{\partial mHb}{\partial y}, \frac{\partial Tor}{\partial y}\right)\right) \cdot \boldsymbol{K}(\tilde{y}) = 0 \qquad (S8.2)$$
396 $S_L=\Delta y/\Delta L$ can thus be obtained.
397
398 S_L values for WT are always below 0.1 (black curve in Fig. S8a-c); so, even standard
399 deviation of L is 10%, positional error introduced by imperfect scaling should be less
400 than 1% (consistent with Fig. 2f). In *bcd*$^-$, patterns follow the fixed absolute distances
401 to both termini (Fig. S8a, the dashed grey lines). When the Tor gradient is absent (Fig.
402 S8c), S_L increases significantly near both termini compared with WT, following that
403 dictated by Bcd alone (dashed grey line); while in the middle part (*x/L* between 0.4 to
404 0.6) S_L is still close to zero due to the remaining mHb gradient. The situation is
405 similar when the mHb gradient is lost *(mhb*$^-$ or *nos*$^-$, Fig. S8b).
406

407 Taken together, these model predictions point to an intuitive interpretation to our
408 phenomenological decoder. First, Bcd plays a central role throughout the entire
409 embryo. Second, the gradient of mHb works together with Bcd as a pair of
410 "bi-gradient" morphogens in the middle part (y = 0.25 to 0.75), while Bcd and Tor
411 works together near both ends (0 to 0.35, and 0.65 to 1.0).

412

413 This feature can also be illustrated from another perspective. We can directly
414 calculate the contribution of each morphogen in discriminating adjacent points $y-$
415 $\delta y/2$ and $y+\delta y/2$ in the WT embryo (for WT $y\equiv\tilde{y}$), i.e. specifying the cell fates
416 in WT. The linear classifier sitting at position y works by computing the sign of the
417 inner product:

418 $$Z(y')\equiv\big(Bcd(y')-Bcd(y),mHb(y')-mHb(y),Tor(y')-Tor(y)\big)\cdot\boldsymbol{K}(y)$$

419 Z values for adjacent y' points should differ by

420 $$Z(y+\delta y/2)-Z(y-\delta y/2)=\left(\frac{\partial Bcd}{\partial y},\frac{\partial mHb}{\partial y},\frac{\partial Tor}{\partial y}\right)\cdot\boldsymbol{K}(y)\delta y.$$

421 The contribution of Bcd, for example, is simply defined as the contribution of the Bcd
422 term in this inner product:

423 $$c_{Bcd}=\frac{\frac{\partial Bcd}{\partial y}\cdot K_1(y)}{\left(\frac{\partial Bcd}{\partial y},\frac{\partial mHb}{\partial y},\frac{\partial Tor}{\partial y}\right)\cdot\boldsymbol{K}(y)} \quad \text{(S8.3)}$$

424 Values of c_{Bcd}, c_{mHb} and c_{Tor} along the A-P axis are shown in Fig. 5e. Obviously, these
425 three c terms should add up to 1. The regions in which mHb and Tor play a role is
426 clearly shown.

427

428 Note that we can define the Bcd dosage sensitivity S_{Bcd} in a similar manner. S_{Bcd}
429 describes the shift Δy of certain cell fate $\tilde{y}$ upon an infinite small change of Bcd
430 dosage (by a factor of 1+ε, thus the Bcd exponential profile is effectively shifted by
431 $\lambda_B\varepsilon$). $S_{Bcd}\equiv\Delta y/\lambda_B\varepsilon$. Δy here is determined by

432 $$\left(\varepsilon(Bcd,0,0)+\Delta y\left(\frac{\partial Bcd}{\partial y},\frac{\partial mHb}{\partial y},\frac{\partial Tor}{\partial y}\right)\right)\cdot \boldsymbol{K}(\tilde{y})=0\,.$$

433 Since Bcd has an exponential profile, working out the formula yields that S_{Bcd} is in
434 fact the same quantity as c_{Bcd} defined above.
435
436 An interesting point is that morphogen contributions are different for WT embryos of
437 different lengths – larger embryos seem to rely more on mHb (Fig. S8d). This may be
438 a testable prediction for future experiments – the same perturbation in mHb (or Nos)
439 should introduce larger pattern shift (or more sever segment defects) for larger
440 embryos (Fig. S8e).
441
442 Also note that Bcd functions throughout the entire embryo is also feasible
443 biochemically. There should be still around 100 Bcd molecules per nucleus
444 (concentration on the order of nM) even at the most posterior nucleus (estimated by
445 the measurements of [18], with GPF maturation effect corrected following Ref.[6]).
446

S9. Shift of the *even-skipped* (*eve*) stripes under Bcd dosage change.

Shift of the cephalic furrow (CF, corresponds to the fate $\tilde{y}$=0.344) in response to Bcd dosage change (but embryo length is fixed to L=1.0) is discussed in the main text. This prediction can obviously be generalized to other "marks" on the fate map. For example, the seven *eve* stripes at $\tilde{y}$=(0.353, 0.435, 0.505, 0.56, 0.62, 0.675, 0.75) according to the measurements of [21]. Fig. S9a presents the predictions of the *eve* stripe positions under Bcd dosage change with or without the presence of the other maternal gradients (mHb and/or Tor). The intuitive explanation raised in Section S8 is again reflected in these results – that Bcd dosage robustness in the middle/terminal part depends on mHb/Tor (arrow heads). Note that since the 5th strip of *eve* locates at the posterior "boulder" between the region mainly governed by Bcd-mHb and that by Bcd-Tor, it should have exactly the same behaviors as CF discussed in Fig. 5f-h (which locates at the anterior "boulder" of this kind). This may be tested by future experimental studies.

Our predictions on *bcdnX* without further mutating mHb or Tor can be compared to the measurements of [22]. Their measurements on the *eve* stripes were carried out at some different developmental timepoint (and the embryo orientations are not well-controlled). So, we use *their* measured WT positions $\tilde{y}$=(0.33, 0.415, 0.5, 0.565, 0.635, 0.7, 0.78) to predict their results on *BcdnX* (Fig. S9b). The Bcd dosage of their *bcd4X* embryos is assumed to be 1.7 times of the WT since no exact values were provided.

470 **S10. Optimal decoder for noisy gradients without embryo length variation.**

471 Given the distribution of input signal, the decoder structure can be properly shaped to
472 best utilize the available information, hence minimizing output error. This optimal
473 decoding idea is commonly used in biological scenarios, ranging from neural sciences
474 to developmental biology [23]. However, to get meaningful results via this approach,
475 there should be much insight about the exact form of input fluctuations that the
476 decoder really cares about – although minimizing output noise is always "a good
477 thing", in many real-world situations, the decoder structure is shaped mainly by the
478 requirements of other (more important) functions, not just by the simplest form of
479 input noise alone.

480

481 In the case of *Drosophila* A-P patterning, we have considered two types of input
482 fluctuations: noisy morphogen gradients and fluctuation in embryo length. We reason
483 that the real-world *Drosophila* decoder must deal with both kinds of noise (especially
484 the latter, i.e., scaling), since the decoder should function in all cells in all embryos of
485 different lengths. Therefore, a decoder optimally designed only for noise attenuation
486 in L=1 embryos is actually *not* optimal, by definition, for the *Drosophila* species in
487 natural environment. In this section, we discuss the difference between the full scaling
488 decoder (that in the main text, which takes into consideration both types of noise) and
489 the optimal noise attenuation decoder where fluctuation in L is not considered. By
490 these analyses, we emphasize the importance of having the *correct* form of input
491 noise in obtaining the *correct* optimal decoder structure.

492

493 The first situation studied here is called "noise-only decoder", where the Poisson
494 noise term on morphogen gradients is the same as in Fig. 2d, but L is now fixed to be
495 1 instead of being a random variable. The second model is a more drastic
496 simplification, that morphogen noise is assumed to be vanishingly small. In this limit,
497 each $\boldsymbol{m}(y)$ point is blurred into a small 3-d sphere, therefore, the optimal decision

498 plane separating adjacent “spheres” is just the normal plane of $\boldsymbol{m}(y)$. We call it the
499 “null model”, as it represents the most naïve form of “optimal” decoder one may
500 imagine. Systematic investigation of the “noise-only decoder” and the “null model”,
501 in comparison with the full “scaling decoder” in the main text, is summarized in Fig.
502 S10.
503
504 First, the noise-only decoder and the null model *are* different from the optimal scaling
505 decoder. This can be seen by direct visualization of the decision planes (Fig. S10a, e,
506 i), which is further highlighted by the different extrapolations on the Bcd-mHb plane
507 (Fig. S10b, f, j). Some of the differences are marked by arrowheads. Also, the
508 morphogen contribution plots are different in three models (Fig. S10c, g, k). In Fig.
509 S10d, h, the decoding result of these two other models on the full WT point cloud
510 (that of Fig. 2d in the main text) are shown, scaling errors are in general larger
511 compared to the scaling decoder (Fig. S10L). Therefore, the noise-only decoder and
512 the null model are no longer optimal if fluctuations in embryo length are considered.
513
514 These structural differences are significant, since they lead to observable differences
515 in mutant predictions (Fig. S10m, n, o). Predictions of our scaling decoder (panel o)
516 match experiments the best, while the noise-only decoder or the null model has
517 prediction error typically on the order of >5% embryo length (marked by
518 arrowheads).
519
520 As an example, consider the vi-vii boundary (*gt*-*hb* boundary). In the scaling decoder,
521 Bcd plays a non-negligible role near the posterior pole to allow for scaling (Fig. S10k).
522 Graphically, the vi-vii boundary is inclined up (Fig. S10i, arrow head). This feature is
523 supported by experimental measurements, that the posterior *hb* domain expands
524 anteriorly in *bcd*$^-$ (Fig. 4b, Fig. S10r). However, this is not captured by the null model
525 (Fig. S10q), nor the noise-only decoder (Fig. S10p), where this boundary is nearly

526 parallel to the Bcd-mHb plane (Fig. S10a, e, arrowheads). In this example, the scaling
527 decoder *is* the best description of *Drosophila* gap gene system among the three, as it
528 predicts the mutant situation the most precisely.
529
530 Fig. S10s&t is in parallel with Fig. 5a of the main text, albeit for the noise-only and
531 null models. Especially, the null model predicts that *Kr* and *kni* domains should
532 disappear at nearly the same *L*, which contradicts experiments [7] qualitatively. Fig. S10
533 u&v is about the issue of cephalic furrow, in parallel with the main text Fig. 5f-h.
534 Admittedly, the noise-only decoder predicts the CF position as satisfactorily as the
535 scaling decoder. But the null model failed to predict the robustness of X_{CF} with
536 respect to Bcd dosage change (Fig. S10v), which is the most important feature here.
537
538 In summary, the analyses above demonstrate that scaling has indeed introduced some
539 additional (and correct) constraints on the decoder structure, beyond merely decoding
540 noisy gradients at fixed embryo length. The gap gene system seems to be "optimized"
541 for both the kinds of fluctuations: noisy morphogen gradients and that in embryo
542 length.
543

544 **S11. Implementing the scaling decoder with a dynamical gene regulation model.**

545 In this section, we present an ordinary differential equation (ODE) model which
546 realizes the phenomenological scaling decoder for the *Drosophila* case.

547

548 The idea is fully illustrated by the toy model in Fig. 6a-c, although the gap gene
549 network has more degrees of freedom and more scaling boundaries, and does not
550 necessarily have to reach a dynamical attracting point. As we show below, after being
551 properly fitted, the ODE gap gene regulation model turned out to have basically the
552 same input-output relations as the previous phenomenological decoder. It also has
553 satisfactory predictions on mutant patterns.

554

555 Under the local-readout hypothesis, diffusion of gap gene products is ignored. At each
556 spatial point, the ODE model calculates the synthesis rates out of the current gap gene
557 product ($\boldsymbol{P}$) and morphogen ($\boldsymbol{M}$) levels. (Degradation coefficient is simply set to a
558 fixed value γ=0.05 min^{-1}.)

559 $$\frac{d}{dt}\boldsymbol{g} = \boldsymbol{f}(\boldsymbol{P}, \boldsymbol{M}) - \gamma \boldsymbol{g}$$

560 Here we model the regulation logics of anterior and posterior *hb* domains separately,
561 since they are known to be generated by different regulation rules even at a course
562 grained level [24,25]. Similar situation applies to the two *gt* domains [26,27]. Therefore, a
563 total of 7 “gap genes” are considered.

564 $$\boldsymbol{g} = (g_1, \ldots, g_7) = (hb_a, hb_p, Kr, kni, gt_a, gt_p, tll)$$

565 Since hb_a and hb_p encode the same Hb protein, $\boldsymbol{P}$ has only 5 components.

566 $$\boldsymbol{P} = (g_1+g_2, g_3, g_4, g_5+g_6, g_7)$$

567 And in the input term $\boldsymbol{M}$, we consider Bcd, Tor and Caudal (Cad, which is shaped by
568 translational repression from Bcd directly, and its profile is fitted to the measurements
569 of Ref.[13,28]. $\text{Cad} = 1/(1 + 1250 * \text{Bcd}^{2.47})$).

570

571 The mHb profile is used as the initial condition of hb_a, and the rest $\boldsymbol{g}$ dimensions are
572 initialized at zero.
573
574 The $\boldsymbol{f}$ term implements the known gap gene interactions reviewed in Ref.[24] (Fig. 6d
575 and Table S2) with a set of Hill-function-like formulas defined as follows. For
576 example, if P_j plays an activating role in f_i, the corresponding term is:

577 $$f_i^j = v_{ij} \cdot \sigma\left(P_j, K_{ij}, b_{ij}\right).$$

578 Instead, if the regulation role is inhibitory,

579 $$f_i^j = 1 - \sigma\left(P_j, K_{ij}, b_{ij}\right).$$

580 Here, σ is an S-shaped function with K and b as its parameters.

581 $$\sigma(x, K, b) = \frac{s\left(K(x-b)\right) - s(-Kb)}{1 - s(-Kb)}, \quad \text{where } s(z) = \frac{1}{1 + exp(-z)}$$

582 This function mimics the Hill function, with K and b together defines the steepness
583 and half-maxima position of the S-shaped curve. These are free parameters to be fitted.
584 Our reason for choosing this function form is purely technical – it performs better in
585 parameters optimization by gradient descent than the original Hill function. Taken
586 together, when a gap gene is regulated by multiple factors, the activation terms are
587 summed and multiplied by the repression terms.

588 $$f_i = \left(\sum\nolimits_{Act.} f_i^j\right)\left(\prod\nolimits_{Inh.} f_i^j\right) \quad i = 1 \cdots 6$$

589 An exception is *tailless* (*tll*). As *tll* is known to act upstream of the gap gene network,
590 and is shaped mainly by Tor and Bcd [29], here we explicitly write down the equation
591 for *tll* using Hill functions.

592 $$f_7 = \frac{1}{1 + \left(\frac{Bcd}{0.011}\right)^4}\left(1 - \frac{1}{1 + \left(\frac{Tor}{0.07}\right)^2}\right)$$

593

594 With the above dynamical equations and initial condition, the model is integrated
595 numerically at each embryonic position (101 discrete points along the A-P axis)

596 independently, as we ignore diffusion of the gap gene products. The model trajectory
597 is compared with the measured WT profiles [30] in the following two aspects to define
598 the Loss function for parameter fitting:

599 (1) Averaged temporal trajectory. Those protein profiles in [30] are measured in the 14th
600 nucleus cleavage cycle (n.c.14) at the blastoderm stage. We smoothed those
601 profiles both spatially and temporally, and extracted 7 equally-spaced frames
602 between 8 and 41 minutes into n.c.14. (The t=41min frame is the one shown in
603 Fig. 2b of main text). The model trajectory for a L=1 embryo is compared with
604 these frames at the corresponding time steps, defining the first term in Loss
605 function.

606 (2) Scaling. We simulate a batch of 16 embryos with L's sampled from the normal
607 distribution $N(0, 0.15)$. Obviously, these embryos may have different gap gene
608 patterns in the initial simulation steps due to their different initial profiles of *hb*.
609 Therefore, to impose scaling, we add a second Loss term penalizing pattern
610 differences *only* at the final frame (t=41 min into n.c.14).

611 Parameter fitting is then carried out using the Adam optimizer implemented in
612 Tensorflow. Note that no mutant information is used for fitting. Table S2 lists a
613 typical fitted parameter set.

614

615 To evaluate the fitted model, we generate a set of WT morphogen profiles with L
616 ranging from 0.85 to 1.15 according to Eqn. 5 (Fig. S11a). The ODE model
617 successfully evolves to yield final patterns that scale with embryo lengths (Fig. S11b),
618 and it indeed shows the overall structure quite similar to the phenomenological
619 decoder (Fig.6e, f, main text). This is expected, as we have shown previously that any
620 scaling local decoding scheme should follow the structure of our phenomenological
621 decoder. Fig.6e shows the Tor=0 section (corresponding to the α plane of Fig.3a).
622 Here, colors stand for the high-expressing gap gene given by the gene-circuit
623 dynamics at the “final” timepoint. Dashed black lines represents the same linear

624 classification planes as in Fig. 3a. Also note that in Fig. 6e-f, outside the WT region,
625 domain boundaries of this ODE model indeed extend in an almost linear manner,
626 supporting our linear extrapolation hypothesis.
627
628 As expected, having the correct decoder geometry for scaling should naturally lead to
629 correct predictions on mutants (Fig. S11c, even gap gene mutants (Fig. S11e) which is
630 not covered by the phenomenological decoder framework). To further evaluate these
631 predictions in a quantitative way, we compared the predicted and measured positions
632 of each gap gene expression domain (peaks and boundaries) in 9 different maternal
633 morphogen mutants (*bcd*$^-$, *nos*$^-$, *tor*$^-$, *bcd1X, bcd4X, bcd*$^-$*nos*$^-$, *bcd*$^-$*tor*$^-$, *nos*$^-$*tor*$^-$,
634 *mhb*$^-$*tor*$^-$), and the root mean square error is only 3.3% embryo length (EL). This is an
635 impressive result. To our knowledge, it has not been reported in previous literatures
636 that a gap gene regulation model (if fitted with WT data only) can make correct
637 predictions in these knockout mutants.
638
639 As a negative control (Fig. S11g) to emphasize the importance of the quantitative
640 constraints imposed by scaling, we also fit this model with the scaling requirement is
641 removed from fitting. The resulting model shows no scaling property and hence
642 incorrect mutant predictions (Fig. S11 g, h, i)
643
644 We also implement the scaling decoder using another widely used gene circuit
645 modeling scheme [31-33]. The only difference is in the form of the ***f*** term (with $\gamma = 0.035$
646 min^{-1}), and the regulation network itself emerged through data fitting (as the sign of
647 the regulation matrixes ***W*** and ***V***).

648 $$\frac{d}{dt}g_i = \sigma\left(\sum_j W_{ij}P_j + \sum_k V_{ik}M_k + b_i\right) - \gamma g_i \;; \qquad \sigma(x) = \frac{1}{1+e^{-x+3}}$$

649 The resulting effective decoder structure and predictions on mutants are also similarly
650 satisfactory (Fig. S12). See Table S3 for model parameters.

651

652 The last thing needs to be noted here is diffusion of the gap gene products, through
653 which decoders in adjacent nuclei may communicate with each other, going beyond
654 the "local-decoding paradigm". However, remind that "local-decoding" is in fact a
655 more stringent requirement. Although diffusion is a physical effect and can never be
656 shut down in reality, it has been pointed out by the authors of "gene circuit model" [32]
657 that their model does not actually rely on diffusion of the gap proteins – shutting
658 down diffusion *in silico* does not affect any dynamical process of their model albeit
659 making the resulting profile less smooth. Similarly, our ODE model here does not
660 include the effect of diffusion, but adding a diffusion term,

661 $$\frac{d}{dt}\boldsymbol{g}(x) = D\nabla^2\boldsymbol{P} + \boldsymbol{f}\big(\boldsymbol{P}(x), \boldsymbol{M}(x)\big) - \gamma\boldsymbol{g}(x)\,,$$

662 appears only to smooth the established spatial pattern, not affecting scaling (Fig.
663 S11k). Here, diffusion constant is set to 0.3 $\mu m^2/s$, twice as that estimated using the
664 exponential tail of the measured gap gene protein pattern of the *FlyEX* database [12,13]
665 (Fig. S11i, j). In Fig. S11k, the parameter sets listed in Table S2 was used, which is
666 fitted in the absence of diffusion. The model would be compatible to even stronger
667 diffusion if such diffusion effect is considered during parameter fitting.

668

669

S12. Nos/mHb as the posterior gradient.

Our work is largely inspired by the bi-gradient model. The original bi-gradient models [20,34], however, are not widely accepted partly because they introduced a hypothetical (and non-exist) posterior gradient. This hypothetical gradient is needed because Nos was ruled out from the very beginning, based on the observations by (and only by) Ref.[4] that positional variability of Hb boundary do not seem to be seriously affected in *nos*$^-$ or *mhb*$^-$ embryos. By contrast, in our model mHb/Nos *is* the second gradient. Below, we demonstrate that the observations in [4] can in fact be well explained with the updated understanding of Bcd gradient since 2002, namely, it has very low noise and its amplitude partially scales with the embryo length. Therefore, those observations cannot rule out mHb/Nos; a hypothetical posterior gradient is not actually needed.

There are two arguments relevant to this topic in the 2002 paper by Houchmandzadeh et al. [4]. First, if Nos/mHb is the factor that helps setting Hb-Kr boundary (Tor is effectively zero here) together with Bcd, then in *nos*$^-$ or *mhb*$^-$ embryos, positional noise of the Hb boundary should follow that of Bcd. The observed positional error of Hb boundary increases from 1%EL (WT) to 1.6%EL (*nos*$^-$), while the positional noise of Bcd measured by [4] is almost 30%EL. There seem to be a big gap. However, as has been correctly pointed out later by Gregor et al., [18], this 30% positional error of Bcd profile is an artifact due to inappropriate normalization; and the true Bcd positional error should be around 1~2% EL if measured with GFP-tagged Bcd and normalized properly [18,35] (and Fig. S2). Therefore, this first argument is invalid from the current point of view. Hb boundary noise *is* at the same level of Bcd in *nos*$^-$ or *mhb*$^-$ background.

A second related observation of Ref.[4] is that, scaling of the Hb boundary seems not to be completely destroyed in *nos*$^-$ or *mhb*$^-$ embryos: its absolute position (measured

698 from the anterior pole, in μm, denoted as x_{hb}) remains to depend on embryo length
699 (with a linear correlation coefficient about 0.7). We think this evidence is also not
700 sufficient for excluding mHb/Nos. This correlation can be explained by the Bcd
701 amplitude effect – that larger embryo tends to have higher absolute level of Bcd [5] (and
702 SI-1). This amplitude effect (which is even slightly underestimated if setting β=2 as in
703 Eqn. 5a) makes the *hb* boundary appear to be partially scaling, even if it is determined
704 solely by Bcd threshold:

705 $$\left.\frac{\partial x_{hb}}{\partial L}\right|_{L=1} = \beta\lambda_B \approx 0.33 \qquad (S12.1)$$

706 The *hb* boundary locates near the center of embryo. Thereby this quantity should be
707 approximately 0.5 for perfect scaling, and equals to 0 for non-scaling as naively
708 expected. The seemly partially preserved scaling in nos^- or mhb^- can be explained in
709 this way.
710

S13. Discussion on the long-germband insect *Megaselia abdita*.

Geometrically speaking, in our framework scaling stems from quantitative matching of the decoder decision boundaries with the *y*-constant curves. On the other hand, from the regulatory perspective, it is equivalent to say that all three morphogen levels change as *L* varies, but their effects on the entire gap gene network should cancel out (at the linear order) to give unchanging outputs. This kind of precise cancellation relies on quantitative tuning of regulatory link strengths in the gap gene network. Such "fine-tuning" may not seem to be a reliable mechanism at a first glance. However, it is consistent with the current understanding of the *evolution* of the gap gene network. The maternal morphogen system is very diverse among different long germband insects, but gap gene cross regulation network is much more conserved [24]. We suggest that by tuning its link strengths quantitatively, a long-germband insect species can easily make the "ancient" gap gene network adapt to its specific maternal morphogen system and achieve scaling patterning.

We would like to briefly discuss, with our scaling framework, another long-germband insects with different sets of maternal morphogen gradients – *Megaselia abdita*. As experimental data (especially on mutants) are very limited, we cannot discuss this point in depth. Like *Drosophila*, *Megaselia* also have Bcd and mHb. The mHb profile is very similar to that of *Drosophila*, but Bcd extends more to the posterior [36,37]. Different morphogen shapes lead to different geometry of the WT point clouds (Fig. S13a), hence different predictions on mutants. In the case of more extended Bcd, when the decision boundary is linearly extrapolated to predict *bcd*$^-$, the *kni* domain expands and the *Kr* domain disappears completely (Fig. S13 illustrates this point with λ_{Bcd}=0.225), which is qualitatively the situation observed in *bcd*$^-$ *Megaselia* [38].

737 **Supplemental Figure legends.**

738

739 **Fig. S1.** The mHb profile can be fitted by a sigmoidal curve. The Hb protein profile in
740 n.c.12 embryos are regarded as maternal Hb here. Data cited from the *FlyEX*
741 database.

742

743 **Fig. S2.** Simulating noise in the Bcd gradient by a Poisson noise term. Black curve:
744 positional error (standard deviation σ_y) calculated according to Eqn. S2.2 (with
745 N_0=1000, λ_B=0.165, and β=2). Blue/green data points: the measured Bcd positional
746 error [18], without/with the known measurement error being subtracted. The Poisson
747 noise term simulates Bcd noise correctly in the region $0.2<y<0.6$.

748

749 **Fig. S3.** Predicting mutant fate-map with linear classifiers. (**a**) A graphical illustration
750 of predicting the fate of y_{mut}=0.4 point in *bcd⁻tor⁻* mutant. Grey and white pixels stand
751 for different classifier outputs on this point. (**b**) A typical case where the "posterior
752 dominant rule" should be employed. A linear classification plane fitted at the anterior
753 (#20 here) intersected with much more posterior parts of the WT point cloud, far
754 away from where the #20 plane was fitted. Hence its classification on those posterior
755 points (yellow rectangle) should be ignored. (**c**) By listing the outputs of all
756 classification planes on all points form a mutant embryo, the predicted fate map is the
757 grey-white boundary. Those outputs ignored by the "posterior dominant rule" are
758 shown in lighter color.

759

760 **Fig. S4.** Effects of the Bcd β factor and noise amplitude on model performance. (**a**)
761 By re-adjusting the classification orientations, a scaling phenomenological decoder is
762 also obtained without the Bcd amplitude factor. It finds the correct $\tilde{y}$ values with
763 relatively small error (RMSE~1%) for WT embryos with length variations. (**b**-**c**)
764 Though scaling is not affected by dropping the amplitude factor, the geometry of the

765 WT point cloud hence the corresponding mutant phenotypes do change (marked by
766 the arrowhead in panel **c**). (**d**) Setting β to 3 does not have much influence on the
767 mutant predictions, compared with Fig. 3d of the main text. (**e-f**) With the settings in
768 the main text (β=2), changing the Poisson noise strength do not affect our main results.
769 The predictions are still satisfactorily consistent with experiments.
770
771 **Fig. S5.** The Bayesian decoder. (**a**) Decoding error for the WT ensemble in Fig.2d of
772 main text. (**b**) The Root-Mean-Squared decoding error is on the same order as our
773 linear decoder. Though arbitrary nonlinear boundary geometry is allowed by the
774 Bayesian decoder, its error is even larger than the linear one. (**c**) Bayesian decoder
775 outputs. Within the region covered by the WT point cloud, they are quite similar to
776 the linear decoder outputs. While there is much difference away from the WT point
777 cloud. Dashed black lines are the same as Fig.3a in the main text. (**d**) Tangential
778 views of the iv-v and v-vi *linear* classification planes of the decoder used in main text.
779 The remaining classification error is due to the Poisson noise rather than nonlinearity
780 in the domain interface geometry.
781
782 **Fig. S6.** Outputs of the scaling decoder on mHb=0 and mHb=mHb_0 planes. The 3-d
783 orientations of the decision planes naturally explans that while *nos*$^-$ embryos have lost
784 the abdominal fates (domains v & vi), *nos*$^-$*mhb*$^-$ embryos still have these two domains
785 just as WT.
786
787 **Fig. S7.** Maternal morphogen mutant embryos with greatly changed length. (**a**) A
788 graphical illustration of the prediction in Fig. 4a of the main text. (**b**) Measured gap
789 gene domain positions in normal-length and greatly shortened *bcd*$^-$ embryos[7]. The
790 narrow anterior *gt* domain is not shown here. (**c**) Overlapping panels A and B for
791 comparison. (**d**) Predicted and measured gap gene domain boundaries in greatly
792 shortened *bcd6X* embryos. Solid lines, model prediction; dots, measurements by Ref.[7].

793 (**e**) Analytical calculation of the *hb* boundary position in *bcd6X* embryos of varying
794 lengths. The black curve follows Eqn. S7.2, with parameters λ_B=0.165, and β=2. (**f**)
795 Predicted gap gene domain positions of many other maternal morphogen mutants with
796 greatly changed L's. The horizontal dashed lines mark typical range of natural length
797 variation within a fly line.
798
799 **Fig. S8.** The three maternal gradients function as two bi-gradient pairs. (**a-c**) Size
800 sensitivity S_L evaluated at L=1. This quantity is equivalent to absolute slopes of the
801 domain boundaries in Fig. 5a, c, d in the main text. S_L is low everywhere in the WT
802 embryo, indicating good scaling (black, S_L <0.1). Missing Bcd (**a**) destroys scaling
803 completely (purple), while missing the posterior (**b**) or terminal (**c**) morphogen only
804 affects scaling in part of the embryo. The dashed grey lines show the non-scaling
805 baselines of S_L. (**d**) Contribution of each morphogen (defined in Section S8) in
806 discriminating adjacent positions in WT embryos. Bcd forms bi-gradient pair with
807 mHb in the middle part, and with Tor near both ends. Note that morphogen
808 contributions are different for different embryo lengths. Nots that mHb seems to play
809 a more important role in larger embryos. (**e**) Therefore, the same perturbation in Nos
810 should introduce more sever segmentation defects for larger embryos. Here, we
811 present predictions on larger and smaller embryos with reduced Nos dosage/activity.
812 Gap gene pattern should be basically normal-looking for L=0.7, while for L=1.3 the
813 *Kr* domain should expand greatly. This prediction may be checked experimentally in
814 the future.
815
816 **Fig. S9.** Shift of the *even-skipped* (*eve*) stripes under Bcd dosage change. (**a**)
817 Predictions with or without the other two maternal morphogens. (**b**) A
818 semi-quantitative comparison with the experimental measurements in Ref.[22].
819

Fig. S10. Optimal decoder for noisy gradients with or without length variation. Here, "scaling decoder" stands for the model studied in the main text. The "noise-only decoder" is defined similarly – *L* is fixed to 1 in generating the WT point cloud, while the Poisson noise term is the same as the scaling decoder. "Null model" is the most naïve form of linear decoder, that the decision planes are always defined to be normal to the (Bcd, mHb, Tor) curve. (**a-h**) Geometry, contribution of each morphogen, and performance in decoding the full-version WT point cloud (i.e., that in Fig.2d main text) of the noise-only decoder and the null model They differs from the scaling decoder (**i-l**) significantly. (**m-o**) The scaling decoder turns out to be a better description of the real *Drosophila* gap gene system; its predictions (grey-white boundary, as in Fig. S3) on mutant patterns match the best with experiments (black dotts). For the noise-only decoder and the null model, where their prediction deviates from the experiments are marked by red arrowheads. (**p-r**) According to the scaling decoder, Bcd should play a non-negligible role even near the posterior pole. This is supported by experiments (expansion of the posterior *hb* domain in bcd^{-}). Note that this feature is not captured by either the null model or the noise-only decoder. (**s-v**) Analysis of the noise-only decoder and the null model following Fig. 5a and f-h of the main text.

Fig. S11. The ODE-based gap gene regulation model. (**a**) The morphogen profiles for *L* ranging from 0.85 to 1.15 following Eqn. 5. These morphogens define the external inputs and initial condition of an ordinary differential equation model for the gap gene network. Note that mHb is multiplied by a factor 4 for better visualization. (**b**) This gene regulation model generates scaling gap gene pattern by reading the non-scaling morphogens in panel A. The requirement of scaling in included in parameter fitting. (**c**) Mutant Predictions (solid lines) and corresponding measured profiles (dashed lines, cited from Refs.[6] and[21]). (**d**) A systematic assessment of mutant predictions. Predicted vs. measured positions of peaks and boundaries of the gap gene domains are

848 shown. The predicted positions have a root-mean-square-error (RMSE) of around 3.3%
849 embryo length, which is quite impressive. (**e**) This model even has reasonable
850 predictions on gap gene mutants. Take the *Kr*$^{-}$ mutant as an example here.
851 Semi-quantitatively, without *Kr*, the posterior *gt* domain should expand anteriorly and
852 eliminate the *kni* domain by inhibition. This is exactly the situation observed in
853 experiments [26,39]. (**f-h**) Regulation network topology alone cannot ensure scaling. (**f**)
854 The ODE model failed to achieve scaling if scaling is not explicitly introduced in
855 parameter fitting. (**g-h**) When being viewed as a decoder of maternal morphogens, its
856 structure deviates from that of our phenomenological scaling decoder (dashed lines).
857 Having the "correct" regulation network topology (identical to Fig. 6d) is not enough
858 for scaling. Quantitative features matter. (**i-j**) Estimating diffusion constant for the gap
859 gene products using the "exponential tails" in the protein profiles. Half-life of the Hb
860 or Kr protein is assumed to be 14 min. Note that this is only a rough estimation. (**k**)
861 The scaling ODE model of panel B is robust to diffusion. The resulting pattern
862 remains normal and scaling, albeit smoothed by diffusion. Diffusion constant here is
863 relatively large, twice the estimated value.
864

865 **Fig. S12.** The scaling gene circuit model. (**a**) With its parameters being properly fitted,
866 using the Loss function defined in Section S11, the gene circuit model generates
867 scaling output pattern successfully. (**b**) The equivalent decoder structure also follows
868 the phenomenological linear decoder (dashed lines). (**c**) The gap gene cross regulation
869 network emerges from data fitting. (**d-e**) This gene circuit model also has satisfactory
870 predictions on maternal morphogen mutants, as expected. (**f**) It can also reproduce Fig.
871 5a-d of the main text, reflecting its structural similarity with the phenomenological
872 scaling decoder.
873

874 **Fig. S13.** Generating scaling *Drosophila* gap gene pattern with a more extended Bcd
875 profile. (**a**) Increasing the Bcd length constant (form 0.165 used in main text to 0.225

876 here) leads to “stretching” of the WT point cloud. The decision boundaries of a
877 scaling decoder are also shifted (dashed black lines). (**b**) As a result, by extrapolating
878 with these decision planes, the fate corresponding to the *Kr* expression domain (iv) no
879 longer presents on the Bcd=0 plane. Compare this panel with Fig. 3a-b to see this
880 difference. Therefore, in a semi-quantitative sense, *bcd*$^{-}$ embryo should no longer
881 have a *Kr* domain with this more extended Bcd. Note that the *Drosophila* gap gene
882 positions are used in this figure, making it comparable with Figs. 2 and 3 in the main
883 text.
884

885 **Table S1.** Peak and boundary positions extracted form maternal morphogen mutants.

wt		*bcd1X*		*bcd4X*		bcd^{-}		nos^{-}	
Pos (% EL)	name	pos	name	pos	name	pos	name	pos	name
12.8	hb1L	14.8	hb1L	15.6	hb1L	18.1	hb2P	18.2	hb1L
47.4	hb1R	40.8	hb1R	56.6	hb1R	11.5	hb2R	50.5	hb1R
80.8	hb2P	80.7	hb2P	84.8	hb2P	25.8	hb2L	82.8	hb2P
74.8	hb2L	74.2	hb2L	79.4	hb2L	77.7	hb2P	74.6	hb2L
87.7	hb2R	86.5	hb2R	89.3	hb2R	68	hb2L	89	hb2R
51.6	krP	46.5	krP	60.6	krP	86.2	hb2R	56.6	krP
45.8	krL	40.3	krL	55.4	krL	39.4	krP	50	krL
58.5	krR	54.3	krR	66.1	krR	32.2	krR	67.4	krP
61.6	kniP	58.5	kniP	68.7	kniP	47.1	krR	38.3	gt1P
56.5	kniL	52.5	kniL	64.4	kniL	47.4	kniP	23.8	gt1L
66.6	kniR	63.9	kniR	72.7	kniR	35.8	kniL	43.5	gt1R
35.4	gt1P	28.3	gt1P	44.4	gt1P	55.6	kniR		
20	gt1L	18.9	gt1L	26.6	gt1L	26.3	gt2P		
40.9	gt1R	33.5	gt1R	49.4	gt1R	21.9	gt2R		
69.7	gt2P	67.7	gt2P	74.8	gt2P	31.5	gt2L		
64.7	gt2L	62.6	gt2L	71	gt2L	58.6	gt2P		
74.8	gt2R	73.5	gt2R	79.9	gt2R	50.8	gt2L		
8	otdL					69	gt2R		
26	otdR								
22	emsL								
31	emsR								
25	btdL								
33	btdR								

886

887

888 Table S1 continued.

tor-		*bcd-tor-*		*nos-tor-*		*bcd-nos-*		*nos-tor-mhb-*	
pos	name	pos	name	pos	name	pos	name	pos	name
48.4	hb1R	10	krP	3.32	hb1L	17.2	hb2P	42.7	hb1R
53.6	krP	29	krP	49.6	hb1R	11.2	hb2R	46.7	krP
47.5	krL	47.1	krR	60	krP	25.3	hb2L	40	krL
60.9	krR	51.6	kniP	80	krP	82.8	hb2P	55	krR
65.6	kniP	43.1	kniL	49.3	krL	74.8	hb2L	60	kniP
59.2	kniL	62.6	kniR	33.3	gt1P	89.4	hb2R	53	kniL
72.5	kniR	55.5	gt2L	9.9	gt1L	50	krP	69.7	kniR
35.4	gt1P	70	gt2P	39.3	gt1R	33.2	krR	30	gt1P
11.2	gt1L					66.6	krR	11.1	gt1L
41.1	gt1R					25.2	gt2P	35.4	gt1R
78.8	gt2P					20.2	gt2R	76	gt2P
70.7	gt2L					29.2	gt2L	66	gt2L
						74.8	gt2P		
						70.5	gt2L		
						80	gt2R		

889

6Bvas-exu-	
pos	name
40	otdR
50	emsR
56	btdR
24	emsL
30	btdL
82	otdR
80	emsR
78	btdR
90	emsL
85	btdL

890

891

892

893 **Table S2.** Parameters for the scaling ODE model.

Regulation link	Sign	W	b	c
Hb to hb-a	Act.	12.86	8.1300	0.2665
Kr to hb-a	Inh.	9.674	-3.5490	/
Kni to hb-a	Inh.	33.98	10.6600	/
Bcd to hb-a	Act.	27.03	6.7250	0.8884
Tor to hb-a	Inh.	3.265	9.8870	/
Kr to hb-p	Inh.	16.35	8.4220	/
Kni to hb-p	Inh.	17.12	8.8830	/
Tll to hb-p	Act.	32.97	9.9490	2.2230
Tor to hb-p	Inh.	8.948	10.9600	/
Hb to Kr	Inh.	5.909	-0.3211	/
Kni to Kr	Inh.	11.4	-2.7410	/
Gt to Kr	Inh.	20.2	2.4620	/
Tll to Kr	Inh.	25.35	9.1960	/
Bcd to Kr	Act.	8.417	-0.6508	5.7930
Cad to Kr	Act.	20.05	-3.6540	1.5560
Tor to Kr	Inh.	24.06	12.2500	/
Hb to kni	Inh.	38.25	4.3050	/
Gt to kni	Inh.	24.56	-2.7790	/
Tll to kni	Inh	69.30	10.7100	/
Bcd to kni	Act.	/	/	0
Cad to kni	Act.	6.008	-0.6122	2.0760
Tor to kni	Inh.	3.993	6.9190	/
Kr to gt-a	Inh.	22.08	7.3770	/
Bcd to gt-a	Act.	27.1	2.0770	1.5040
Tor to gt-a	Inh.	27.27	1.8270	/

Hb to gt-p	Inh.	21.55	12.3600	/
Kr to gt-p	Inh.	14.45	2.8210	/
Tll to gt-p	Inh.	4.086	9.1700	/
Cad to gt-p	Act.	6.042	-3.1790	3.3710
Tor to gt-p	Inh.	7.685	1.8580	/

894

895

896

897 **Table S3.** Parameters for the scaling gene circuit model.

898 Weights: W&V

	hb-a	*hb-p*	*Kr*	*kni*	*gt-a*	*gt-p*
Hb	-0.855	4.031	-3.526	-15.49	1.084	-22.06
Kr	1.179	-22.44	4.152	-2.664	-14.11	-18.03
Kni	-64.61	-11.24	-3.721	6.641	-37.88	3.473
Gt	2.135	4.918	-10.06	-9.866	-0.609	2.392
Tll	-17.6	-0.248	-14.01	-44.98	1.735	-2.996
Bcd	-2.911	-38.12	-3.172	-8.143	-15.35	-31.18
Cad	-3.129	4.412	-1.654	2.084	-8.448	5.898
Tor	-2.928	-3.23	-51.2	-4.751	-20.07	-5.281

899

900 Bias:

	hb-a	*hb-p*	*Kr*	*kni*	*gt-a*	*gt-p*
Bias b	5.127	-2.953	4.732	1.728	7.834	-1.639

901

902

References

1. Bergmann, S., *et al.* Pre-steady-state decoding of the Bicoid morphogen gradient. *PLoS Biol* **5**, e46 (2007).
2. Driever, W. & Nusslein-Volhard, C. A gradient of bicoid protein in Drosophila embryos. *Cell* **54**, 83-93 (1988).
3. Grimm, O., Coppey, M. & Wieschaus, E. Modelling the Bicoid gradient. *Development* **137**, 2253-2264 (2010).
4. Houchmandzadeh, B., Wieschaus, E. & Leibler, S. Establishment of developmental precision and proportions in the early Drosophila embryo. *Nature* **415**, 798-802 (2002).
5. Cheung, D., Miles, C., Kreitman, M. & Ma, J. Scaling of the Bicoid morphogen gradient by a volume-dependent production rate. *Development* **138**, 2741-2749 (2011).
6. Liu, F., Morrison, A. H. & Gregor, T. Dynamic interpretation of maternal inputs by the Drosophila segmentation gene network. *Proc Natl Acad Sci U S A* **110**, 6724-6729 (2013).
7. Huang, A., Rupprecht, J. F. & Saunders, T. E. Embryonic geometry underlies phenotypic variation in decanalized conditions. *Elife* **9**, (2020).
8. Cheung, D., Miles, C., Kreitman, M. & Ma, J. Adaptation of the length scale and amplitude of the Bicoid gradient profile to achieve robust patterning in abnormally large Drosophila melanogaster embryos. *Development* **141**, 124-135 (2014).
9. Irish, V., Lehmannt, R. & Akam, M. The *Drosophila* posterior-group gene *nanos* functions by repressing *hunchback* activity. *Nature* **338**, 646-648 (1989).
10. Hulskamp, M., Schroder, C., Pfeifle, C., Jackle, H. & Tautz, D. Posterior segmentation of the Drosophila embryo in the absence of a maternal posterior organizer gene. *Nature* **338**, 629-632 (1989).
11. Struhl, G. Differing strategies for organizing anterior and posterior body pattern in Drosophila embryos. *Nature* **338**, 741-744 (1989).
12. Poustelnikova, E., Pisarev, A., Blagov, M., Samsonova, M. & Reinitz, J. A database for management of gene expression data in situ. *Bioinformatics* **20**, 2212-2221 (2004).
13. Pisarev, A., Poustelnikova, E., Samsonova, M. & Reinitz, J. FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucleic Acids Res* **37**, D560-D566 (2009).
14. Furriols, M., Casali, A. & Casanova, J. Dissecting the mechanism of torso receptor activation. *Mech Dev* **70**, 111-118 (1998).
15. Johnson, T. K., Henstridge, M. A., Herr, A., Moore, K. A., Whisstock, J. C. & Warr, C. G. Torso-like mediates extracellular accumulation of Furin-cleaved Trunk to pattern the Drosophila embryo termini. *Nat Commun* **6**, 8759 (2015).
16. Coppey, M., Boettiger, A. N., Berezhkovskii, A. M. & Shvartsman, S. Y. Nuclear trapping shapes the terminal gradient in the Drosophila embryo. *Curr Biol* **18**, 915-919 (2008).
17. Kim, Y., *et al.* MAPK substrate competition integrates patterning signals in the Drosophila embryo. *Curr Biol* **20**, 446-451 (2010).
18. Gregor, T., Tank, D. W., Wieschaus, E. F. & Bialek, W. Probing the limits to positional information. *Cell* **130**, 153-164 (2007).

945 19. Ochoa-Espinosa, A., Yu, D., Tsirigos, A., Struffi, P. & Small, S. Anterior-posterior positional
946 information in the absence of a strong Bicoid gradient. *Proc Natl Acad Sci U S A* **106**,
947 3823-3828 (2009).
948 20. Houchmandzadeh, B., Wieschaus, E. & Leibler, S. Precise domain specification in the
949 developing Drosophila embryo. *Phys Rev E Stat Nonlin Soft Matter Phys* **72**, 061920
950 (2005).
951 21. Petkova, M. D., Tkacik, G., Bialek, W., Wieschaus, E. F. & Gregor, T. Optimal Decoding of
952 Cellular Identities in a Genetic Network. *Cell* **176**, 844-855 e815 (2019).
953 22. de Lachapelle, A. M. & Bergmann, S. Precision and scaling in morphogen gradient
954 read-out. *Mol Syst Biol* **6**, 351 (2010).
955 23. Bialek, W. *Biophysics: Searching for Principles*. Princeton University Press (2012).
956 24. Jaeger, J. The gap gene network. *Cell Mol Life Sci* **68**, 243-274 (2011).
957 25. Margolis, J. S., Borowsky, M. L., Steingrímsson, E., Shim, C. W., Lengyel, J. A. & Posakony, J.
958 W. Posterior stripe expression of *hunchback* is driven from two promoters by a common
959 enhancer element. *Development* **121**, 3067-3077 (1995).
960 26. Eldon, E. D. & Pirrotta, V. Interactions of the Drosophila gap gene *giant* with maternal
961 and zygotic pattern-forming genes. *Development* **111**, 367-378 (1991).
962 27. Kraut, R. & Levine, M. Spatial regulation of the gap gene *giant* during Drosophila
963 development. *Development* **111**, 601-609 (1991).
964 28. Manu*, et al.* Canalization of gene expression and domain shifts in the Drosophila
965 blastoderm by dynamical attractors. *PLoS Comput Biol* **5**, e1000303 (2009).
966 29. Pignoni, F., Steingrimsson, E. & Lengyel, J. A. bicoid and the terminal system activate
967 tailless expression in the early Drosophila embryo. *Development* **115**, 239-251 (1992).
968 30. Dubuis, J. O., Samanta, R. & Gregor, T. Accurate measurements of dynamics and
969 reproducibility in small genetic networks. *Mol Syst Biol* **9**, 639 (2013).
970 31. Mjolsness, E., Sharp, D. H. & Reinitz, J. A connectionist model of development. *J Theor*
971 *Biol* **152**, 429-453 (1991).
972 32. Jaeger, J.*, et al.* Dynamic control of positional information in the early Drosophila embryo.
973 *Nature* **430**, 368-371 (2004).
974 33. Manu*, et al.* Canalization of gene expression in the Drosophila blastoderm by gap gene
975 cross regulation. *PLoS Biol* **7**, e1000049 (2009).
976 34. McHale, P., Rappel, W. J. & Levine, H. Embryonic pattern scaling achieved by oppositely
977 directed morphogen gradients. *Phys Biol* **3**, 107-120 (2006).
978 35. Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W. & Tank, D. W. Stability and
979 nuclear dynamics of the bicoid morphogen gradient. *Cell* **130**, 141-152 (2007).
980 36. Stauber, M., Taubert, H. & Schmidt-Ott, U. Function of bicoid and hunchback homologs
981 in the basal cyclorrhaphan fly Megaselia (Phoridae). *Proc Natl Acad Sci U S A* **97**,
982 10844-10849 (2000).
983 37. Wotton, K. R.*, et al.* Quantitative system drift compensates for altered maternal inputs to
984 the gap gene network of the scuttle fly *Megaselia abdita*. *Elife* **4**, (2015).
985 38. Wotton, K. R., Jimenez-Guri, E. & Jaeger, J. Maternal co-ordinate gene regulation and
986 axis polarity in the scuttle fly Megaselia abdita. *PLoS Genet* **11**, e1005042 (2015).

987 39. Pankratz, M. J., Hoch, M., Seifert, E. & Jackle, H. Kruppel requirement for knirps
988 enhancement reflects overlapping gap gene activities in the Drosophila embryo. *Nature*
989 **341**, 337-340 (1989).

990