Supplemental Data

Cell, Volume *133*

A Quantitative Spatiotemporal Atlas of Gene Expression in the *Drosophila* Blastoderm

Charless C. Fowlkes, Cris L. Luengo Hendriks, Soile V. E. Keränen, Gunther H. Weber, Oliver Rübel, Min-Yu Huang, Sohail Chatoor, Angela H. DePace, Lisa Simirenko, Clara Henriquez, Amy Beaton, Richard Weiszmann, Susan Celniker, Bernd Hamann, David W. Knowles, Mark D. Biggin, Michael B. Eisen, and Jitendra Malik

Measuring modes of geometric variation.
Supplementary Figure 1 shows example deformations required to bring individual PointClouds into alignment with the template during fine registration. Supplementary Figure 2 shows a principal components analysis (Mardia et al., 1979) of these spatial deformations. The figure indicates the directions along which nuclei were displaced relative to the average shape but the magnitude of the displacement has been greatly magnified for the purposes of visualization. The dominant mode of this geometric variation, which explained more than 45% of the variance in displacement fields for all cohorts, was an anisotropic "bulging" about the blastoderm center. We also measured overall shifts along the A-P axis in the location of the trunk domain specified by the marker genes (8% of variance), curving of the embryo axis (6% of variance), and more subtle variations in the asymmetry between anterior and posterior ends and dorsal and ventral sides of the PointClouds (Supplementary Figure 2, rows 2, 3, 4).

Bounding geometric variability remaining after registration.
In principle, a perfect registration method would remove all of the differences in expression between individuals that are attributable to geometric variation. Any remaining variation would then constitute measured expression differences between otherwise equivalent cells. Such non-geometric variability could arise from measurement error or from genuine regulatory variability, such as stochasticity in the response of targets to their regulators (Gregor et al., 2007). Quantifying non-geometric variability is important as it sets the maximum degree to which any registration method can possibly reduce the apparent expression variation among embryos. Although the true amount of regulatory variability between embryos is uncertain, our results describe an experimentally derived upper-bound and measure how close our registration results are to achieving this limit.

Figure 5E shows that the variability in the expression of *eve* as a function of *gt* inferred from the atlas data using *eve-ftz* and *ftz-gt* stains is nearly as small as the variability measured directly from *eve-gt* costained embryos. One possible concern with this three-way analysis is that the measurement error could be correlated with batches of embryos used in the experiment. For example, if the co-stained batch happened to have higher variability due to measurement error than the registered batch, the level of variability measured in co-stained embryos would no longer be a lower bound on level of apparent variability in registered embryos. We measured the average standard deviation after

registration for each of the three batches and found values of 0.116/0.0806 for the *eve/gt* co-stain, 0.177/0.137 for *eve/ftz*, and 0.129/0.142 for *ftz/gt*. Since the single gene variability within the co-stained batch was significantly lower for both *eve* (0.116 < 0.177) and *gt* (0.0806< 0.142) we conclude that our estimates of the geometric variability remaining after registration are potentially quite conservative.

Detecting bad matches with a quadratic penalty term

Let m be a matching that specifies for the ith point, located at t^i in our template, the corresponding boundary point, $m(i)$ detected at location $s^{m(i)}$ in an individual embryo. We evaluate the cost:

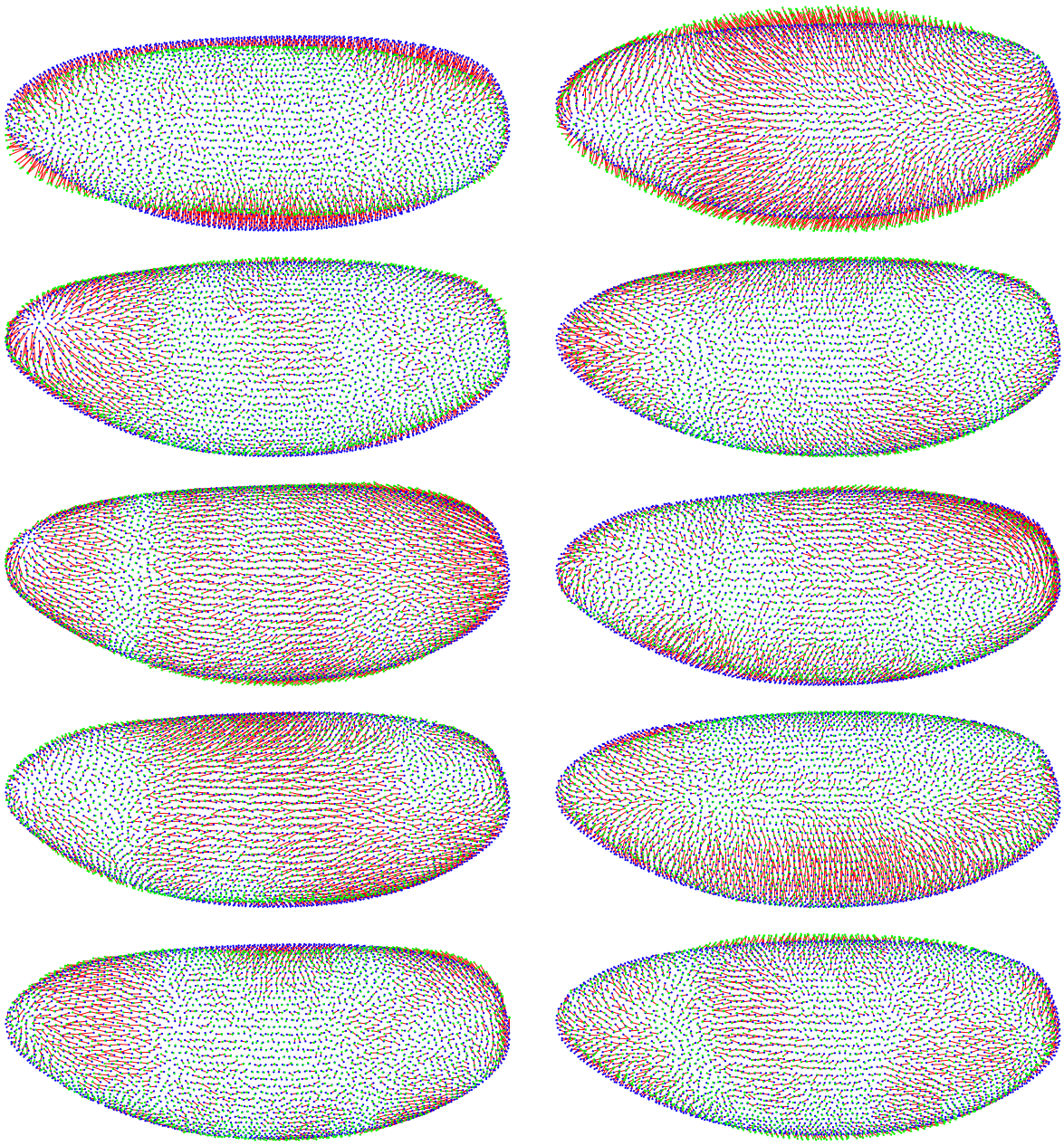
$$Q_i(m) = \sum_j \frac{\left\| (s^{m(i)} - t^i) - (s^{m(j)} - t^j) \right\|^2}{\left\| t^i - t^j \right\|^2}$$

and drop those points ***i*** for which Q_i exceeds a fixed threshold. We also use the total quadratic cost, $\sum_i Q_i$, to evaluate whether a reasonable match was found. A large total cost is indicative of an embryo with a poorly defined marker gene expression pattern. In such cases, the algorithm automatically defaults to using the coarse alignment.

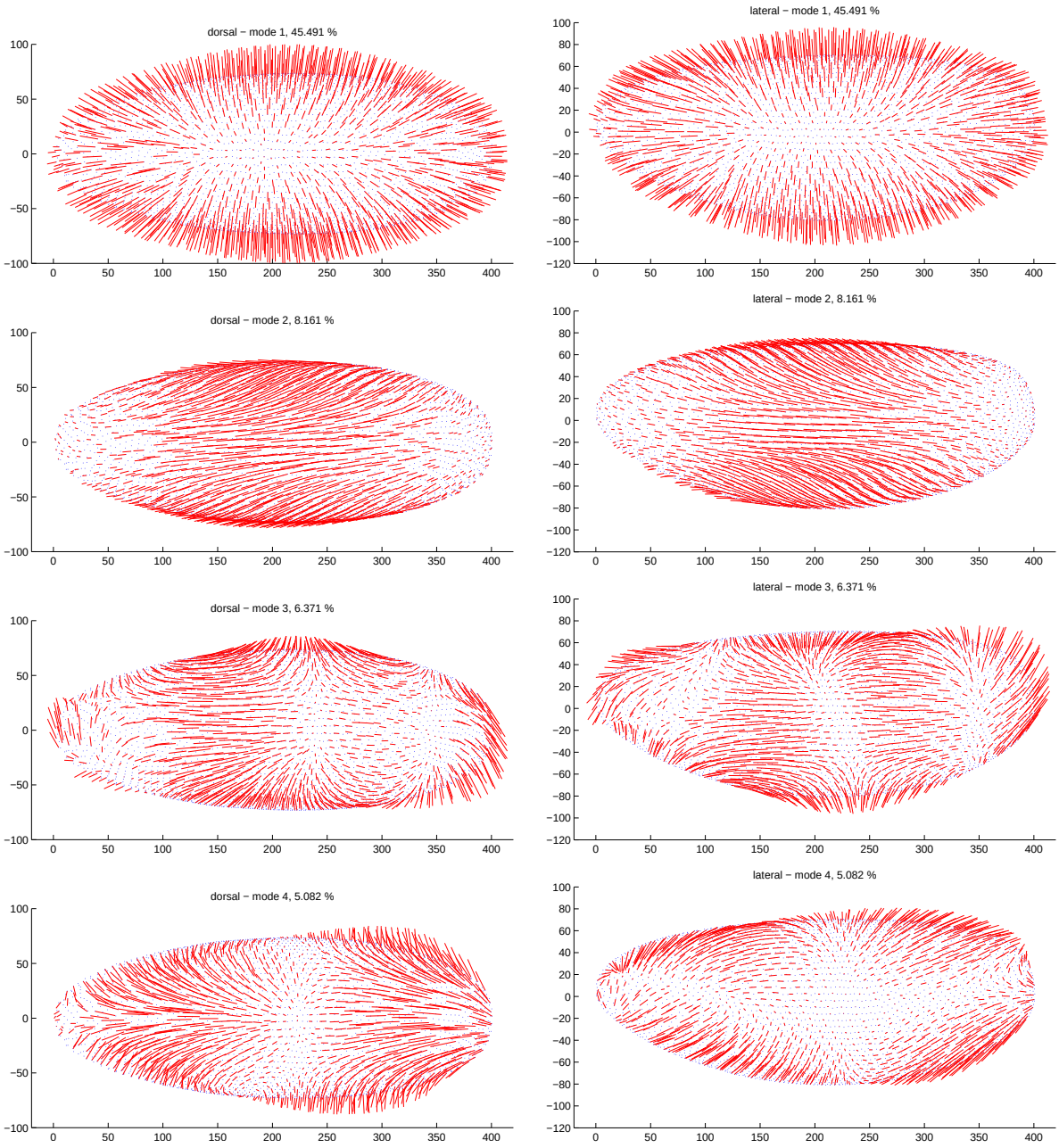
Supplemental References

Gregor, T., Tank, D.W., Wieschaus, E.F., and Bialek, W. (2007). Probing the limits to positional information. Cell 130(1), 153-164.

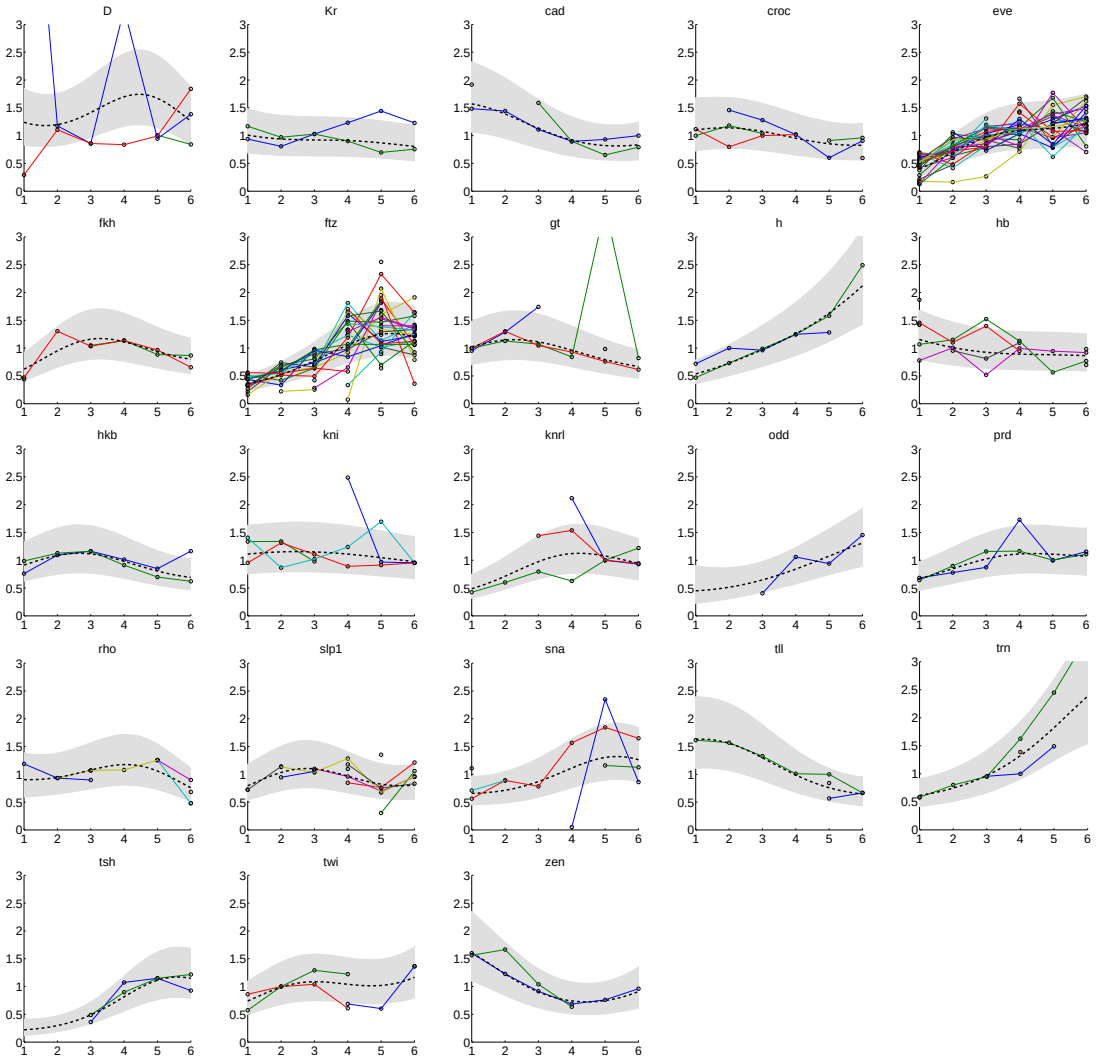
Mardia, K., Kent, J., Bibby, J. (1979) Multivariate Analysis (London: Academic Press)



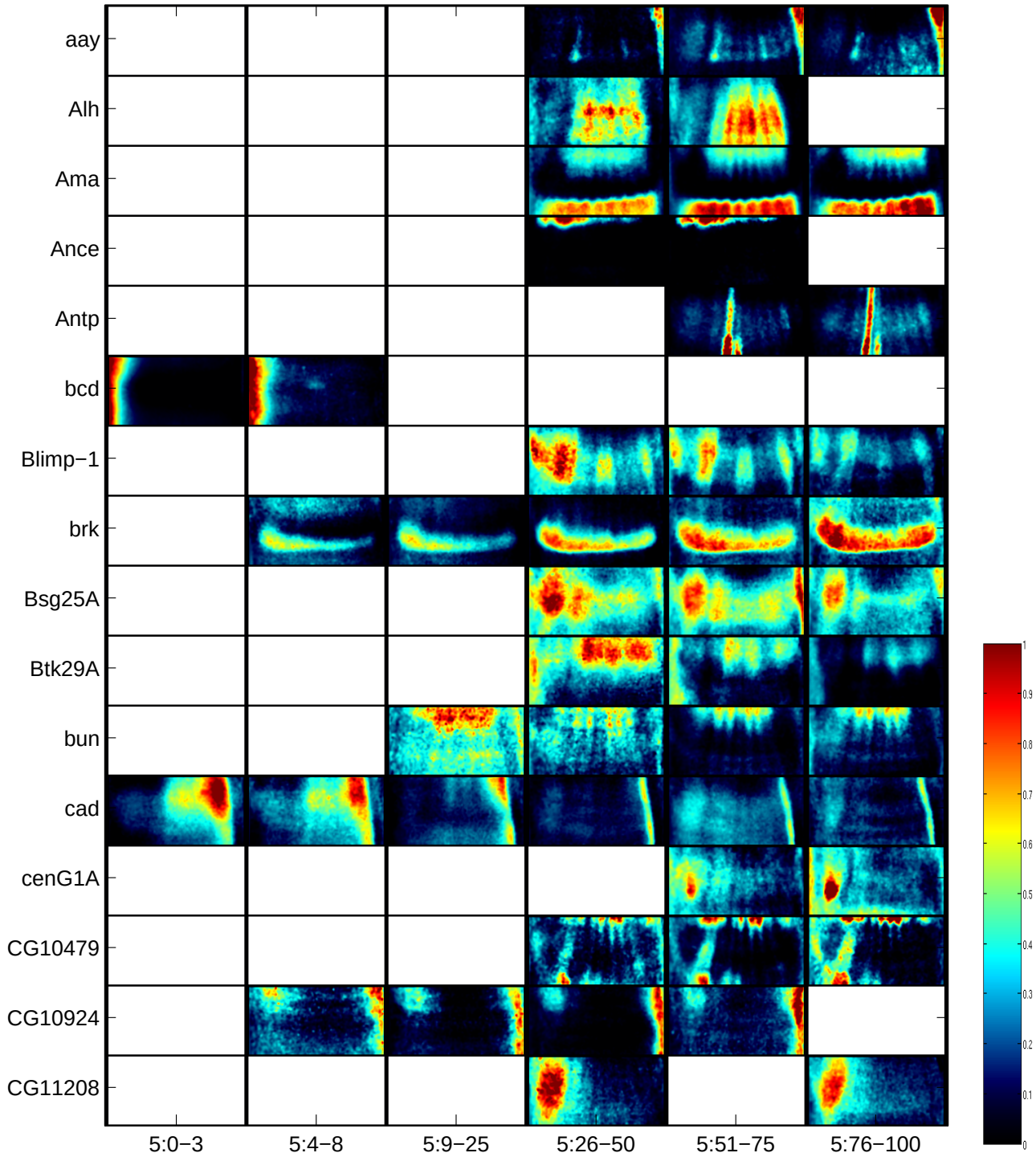
Supplementary Figure 1. Deformations required to map individual PointClouds onto the standard morphological template after coarse alignment. Detected nuclei locations in individual PointClouds (green) and average locations in the morphological template (blue) are shown in lateral orthographic projections with anterior to the left and dorsal up. Each red line indicates the correspondence between a nucleus in the individual PointCloud and one or more nuclei in the template. There is a range of variability in the shapes of individual blastoderm embryos at the same developmental stage (here stage cohort 5:0-3%). These include variation in overall roundness (1st row) shifts of nuclei towards the poles (2nd and 3rd row) a-p and d-v shifts along the trunk (4rd row) and combinations of these local motions (5th row)



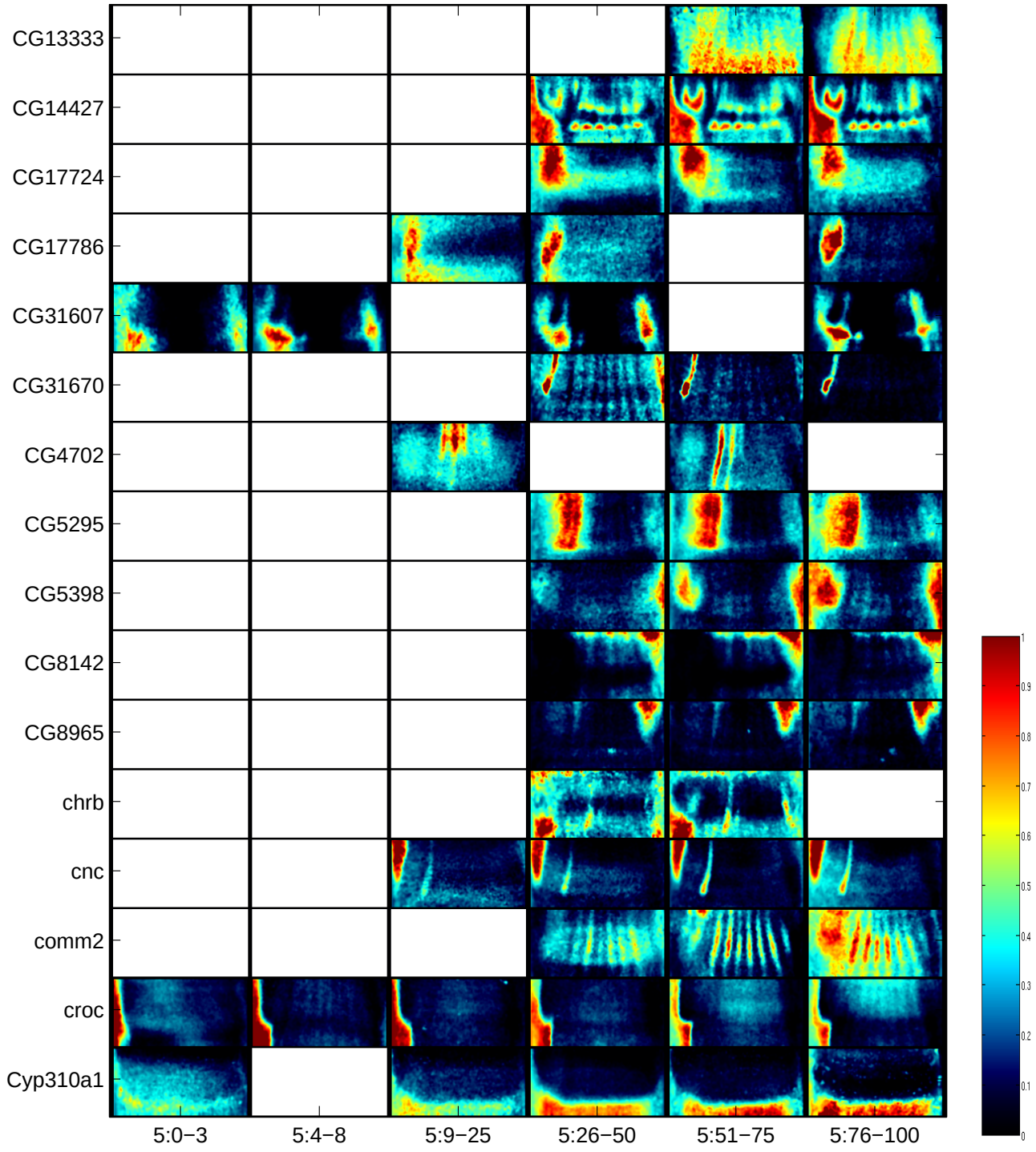
Supplementary Figure 2. A linear analysis of the deformations required to bring embryos into alignment for all embryos. Principal components analysis (Mardia et al., 1979) was performed on the set of deformation fields required to bring each embryo in to correspondence with the template. Each row shows lateral (right panels) and dorsal (left panels) views of a different mode of variation. The displacement fields show the direction of motion but the magnitude has been exaggerated by scaling each mode to have the same norm. The principal mode of variation (capturing 45% of the variance) is change in the overall roundness of the embryo. Successive modes capture shifts of the trunk patterning system relative to the egg endpoints (row 2), curving of the central axis (row 3) and differential bulging at the poles (row 4).



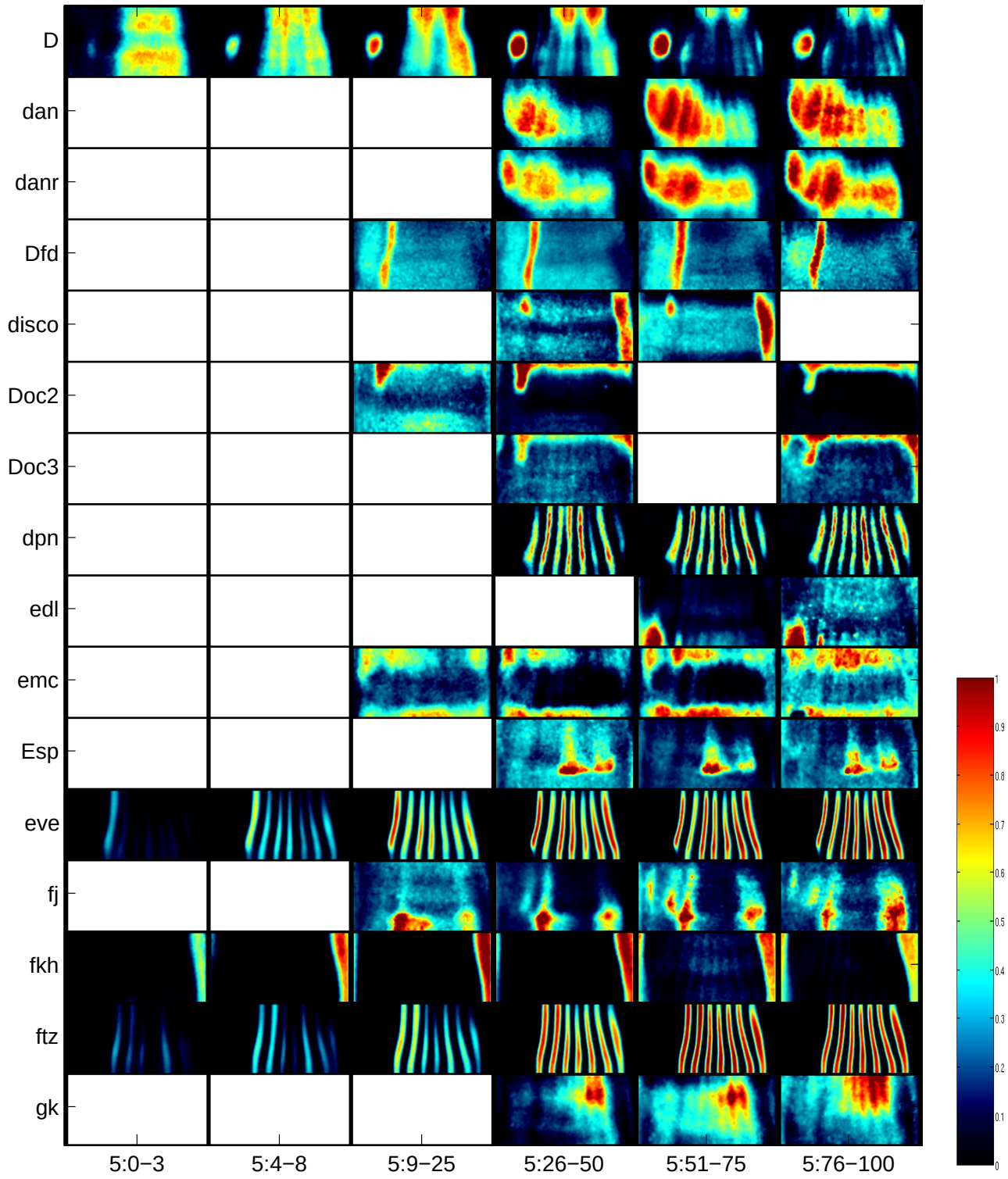
Supplementary Figure 3. Calibrating expression measurements across time. Each plot shows the estimated maximum (99th percentile) expression level for each of 23 genes over six temporal cohorts during the 50 minutes prior to gastrulation. Individual colored curves show estimates using embryos from a given hybridization batch. A separate gain parameter is estimated in order to align batches and remove the effects of variable reaction efficiency or choice of fluorophore. The black curve shows a fit to log expression levels given by Gaussian process regression using a squared exponential covariance function with characteristic length of 3 (in the graph units, roughly 30 minutes) and independent noise model with standard deviation of 0.3



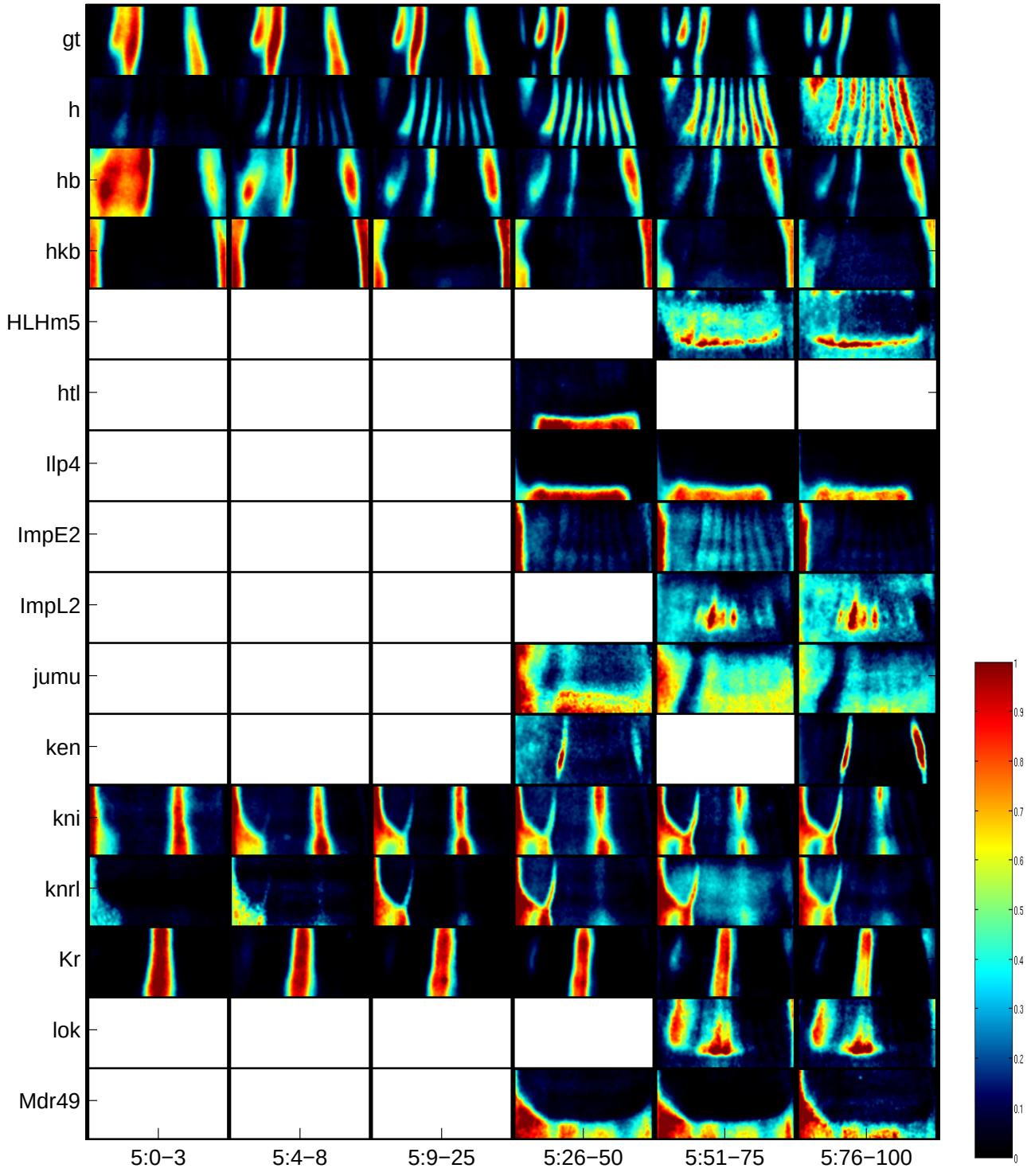
Supplementary Figure 4. A summary of the average mRNA expression patterns for all genes in the initial atlas release. Temporal cohorts, staged by percent membrane invagination, are arranged from left to right with each row corresponding to a different gene. Each rectangle shows a lateral view of the blastoderm in cylindrical projection with dorsal at top, anterior to the left.



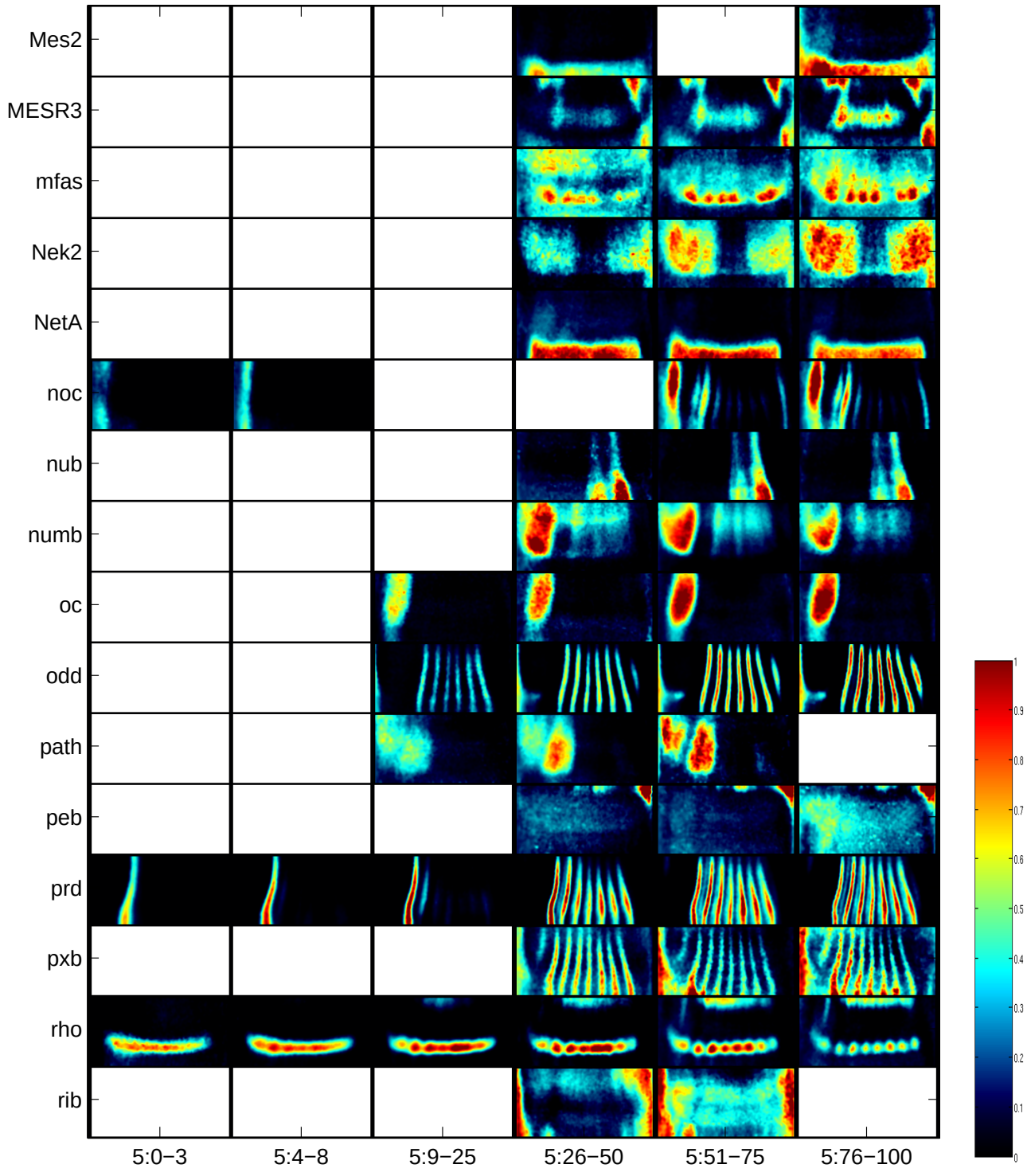
Supplementary Figure 4. A summary of the average mRNA expression patterns for all genes in the initial atlas release. (continued)



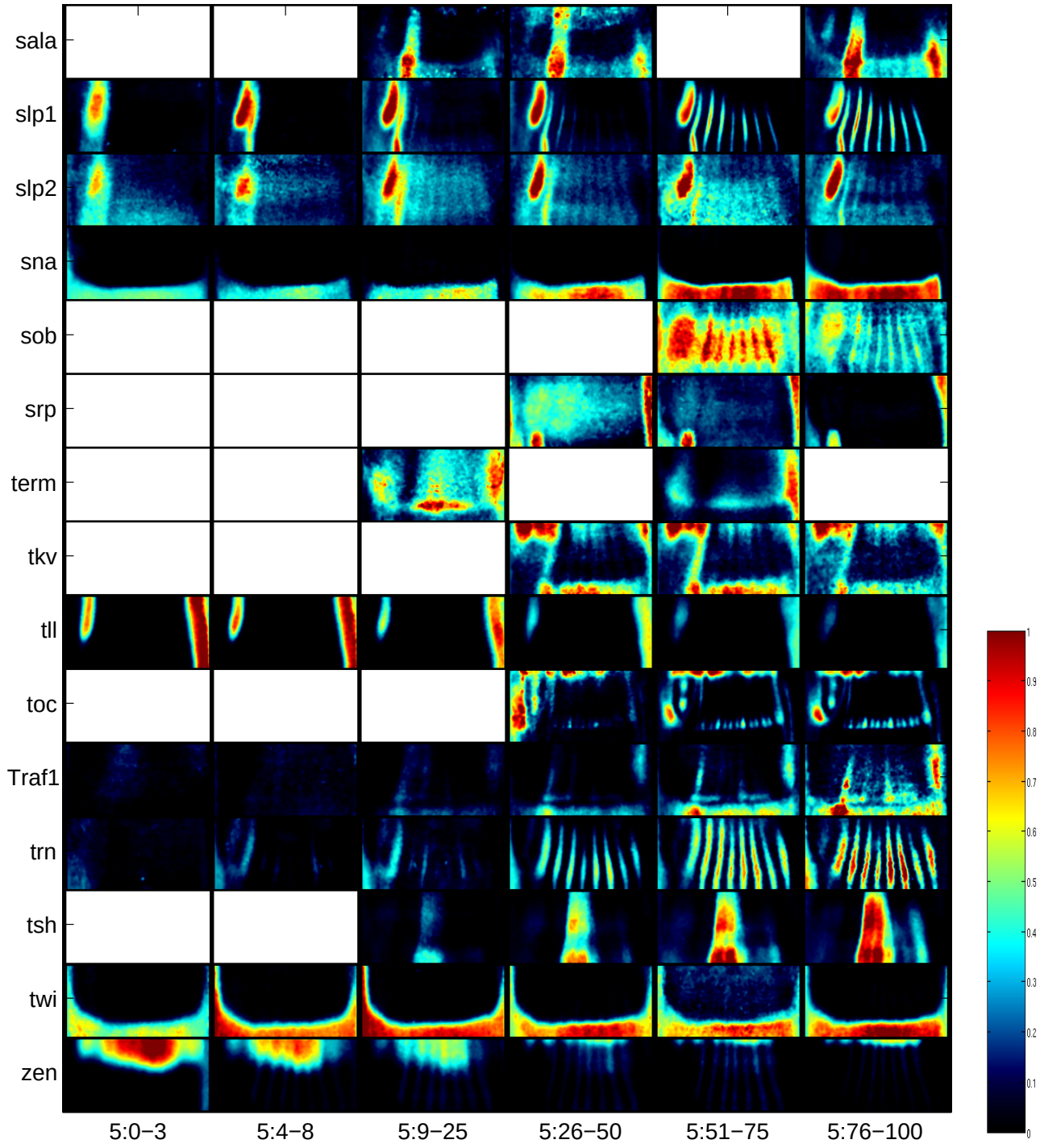
Supplementary Figure 4. A summary of the average mRNA expression patterns for all genes in the initial atlas release. (continued)



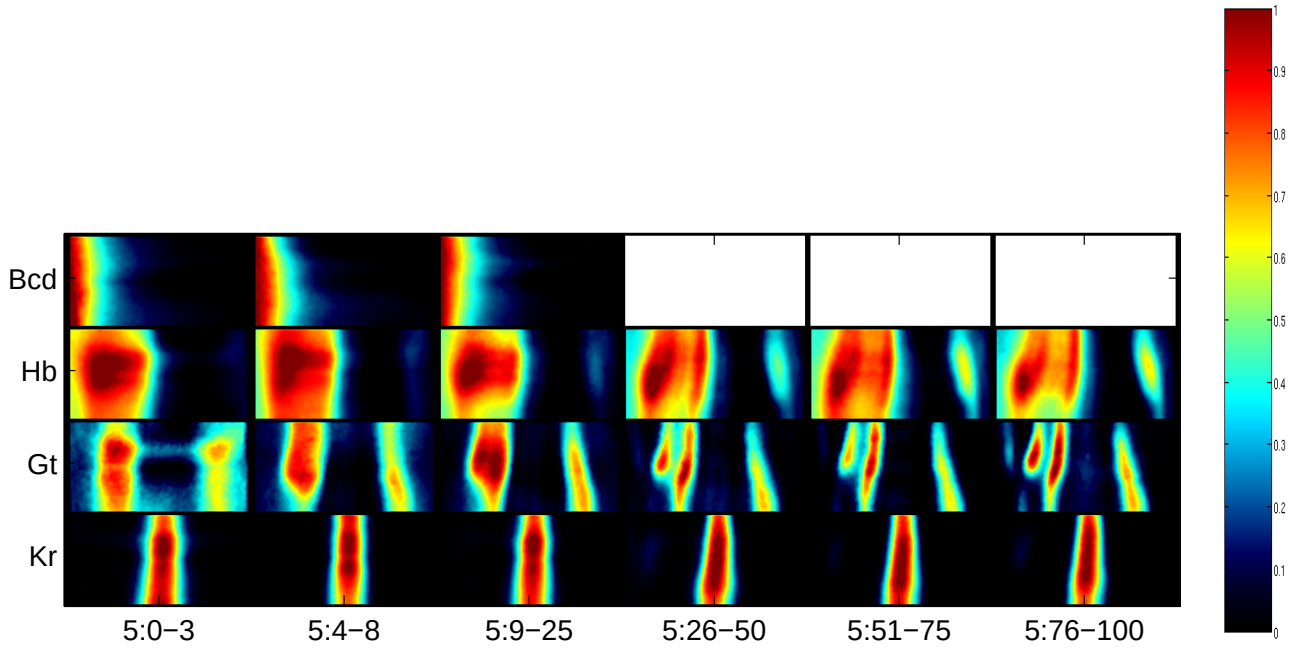
Supplementary Figure 4. A summary of the average mRNA expression patterns for all genes in the initial atlas release. (continued)



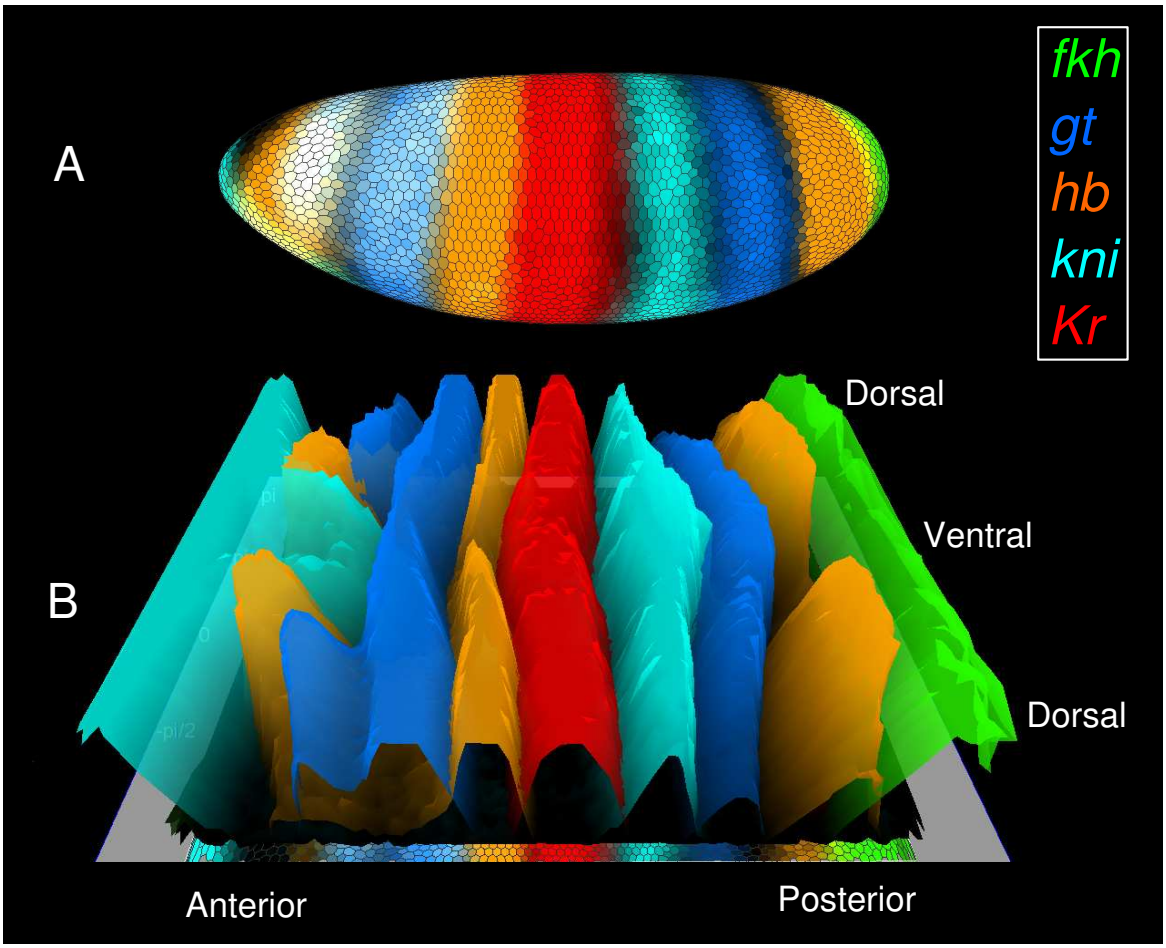
Supplementary Figure 4. A summary of the average mRNA expression patterns for all genes in the initial atlas release. (continued)



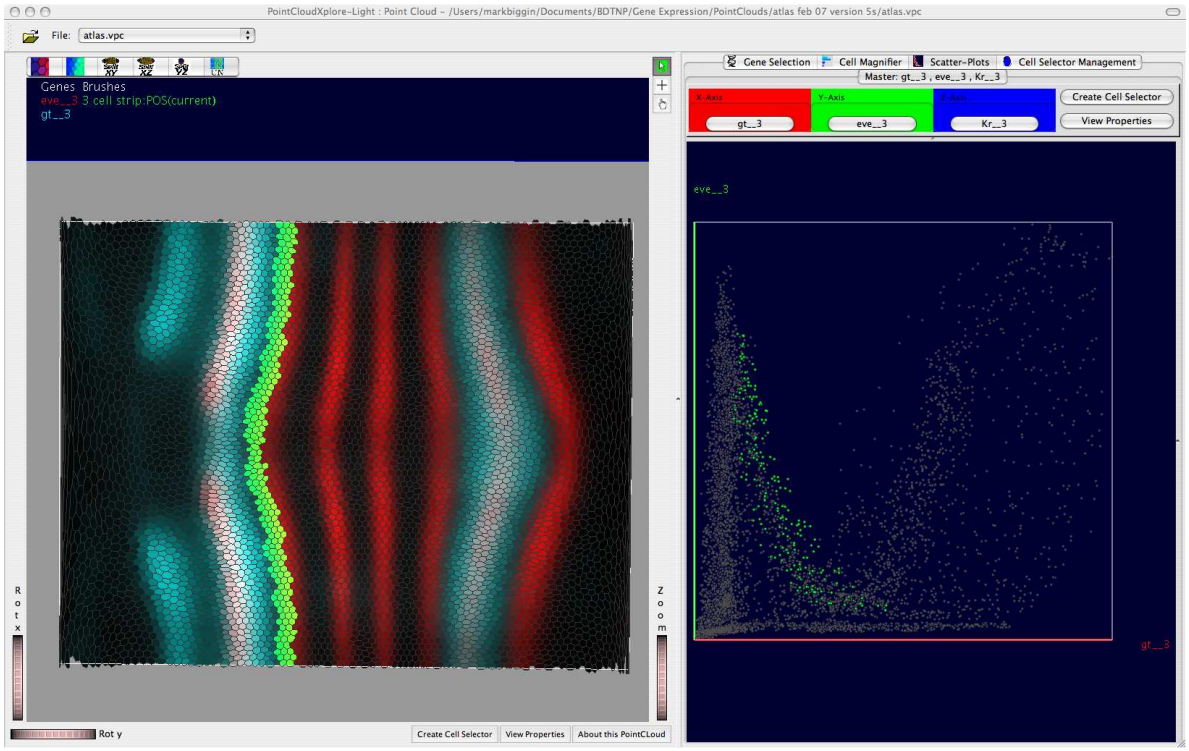
Supplementary Figure 4. A summary of the average mRNA expression patterns for all genes in the initial atlas release. (continued)



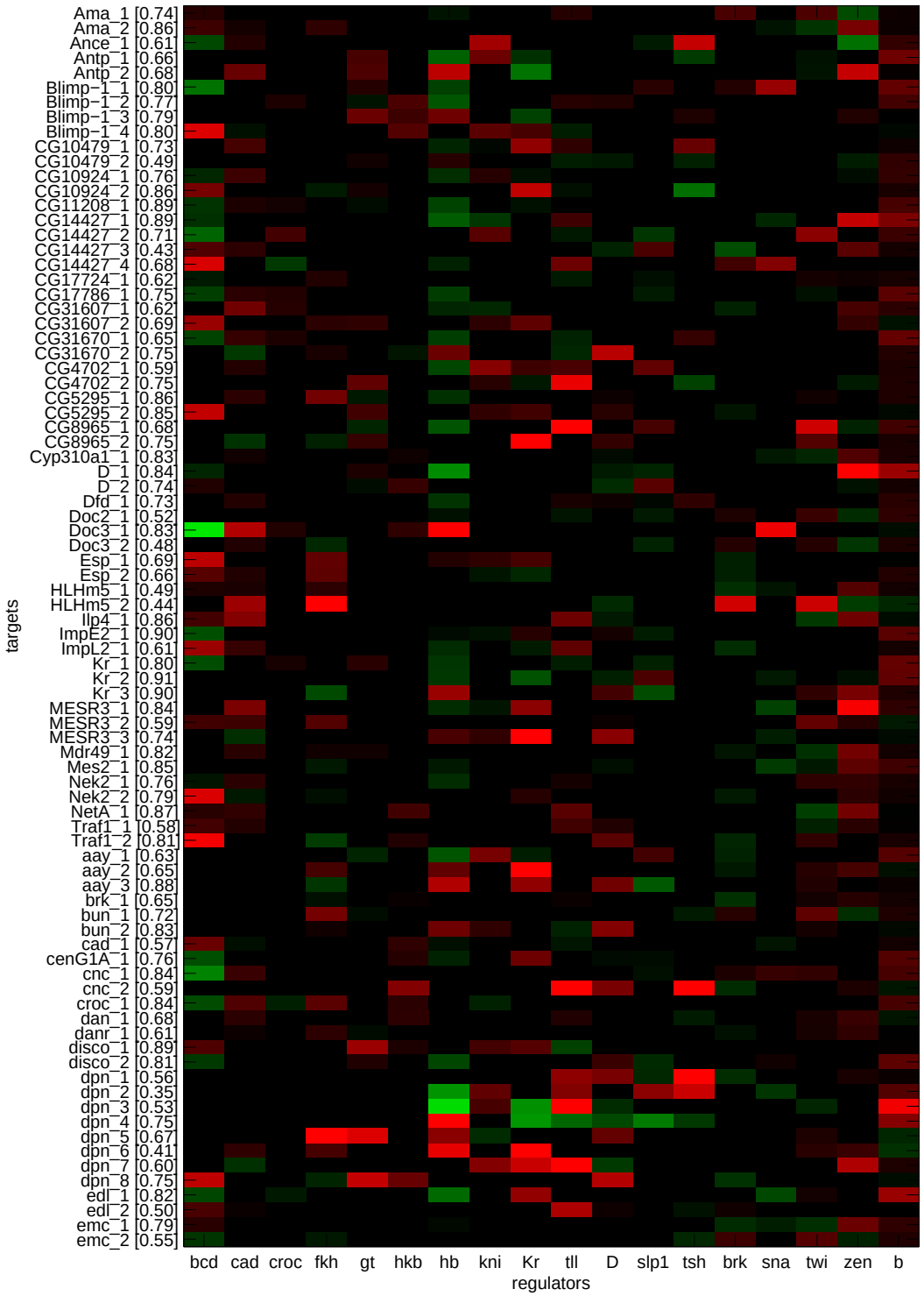
Supplementary Figure 5. Average protein patterns for BCD,HB, GT and KR used in modeling gene regulation. Temporal cohorts, staged by percent membrane invagination, are arranged from left to right with each row corresponding to a different gene. Each rectangle shows a lateral view of the blastoderm in cylindrical projection with dorsal at top, anterior to the left.



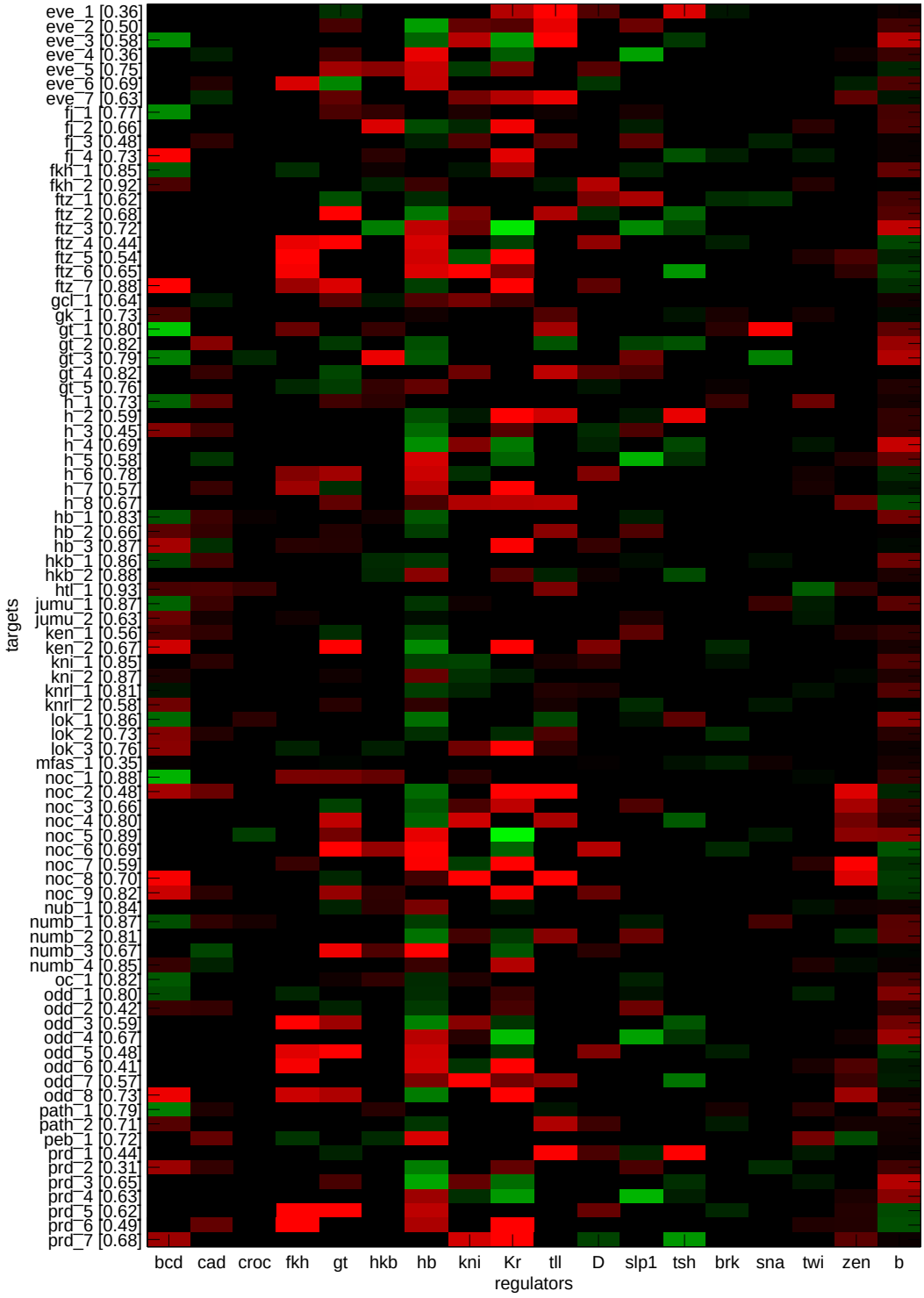
Supplementary Figure 6. PointCloudXplore allows visualization of quantitative 3D expression data. Expression of the gap genes *fkh, gt, hb, kni,* and *Kr* is shown for the stage 5: 4-8% cohort. Panel A shows a 3D model of the blastoderm surface (anterior left, dorsal down) with each nucleus colored according to the expression level of the 5 genes. Panel B shows a cylindrical projection of the entire blastoderm (anterior left, ventral center, dorsal upper and lower edges). The heights of each surface plot indicate the average expression level of the gene recorded at that point on the VirtualEmbryo, making readily visible the quantitative changes in expression of these gap genes along both the A-P and D-V axes.



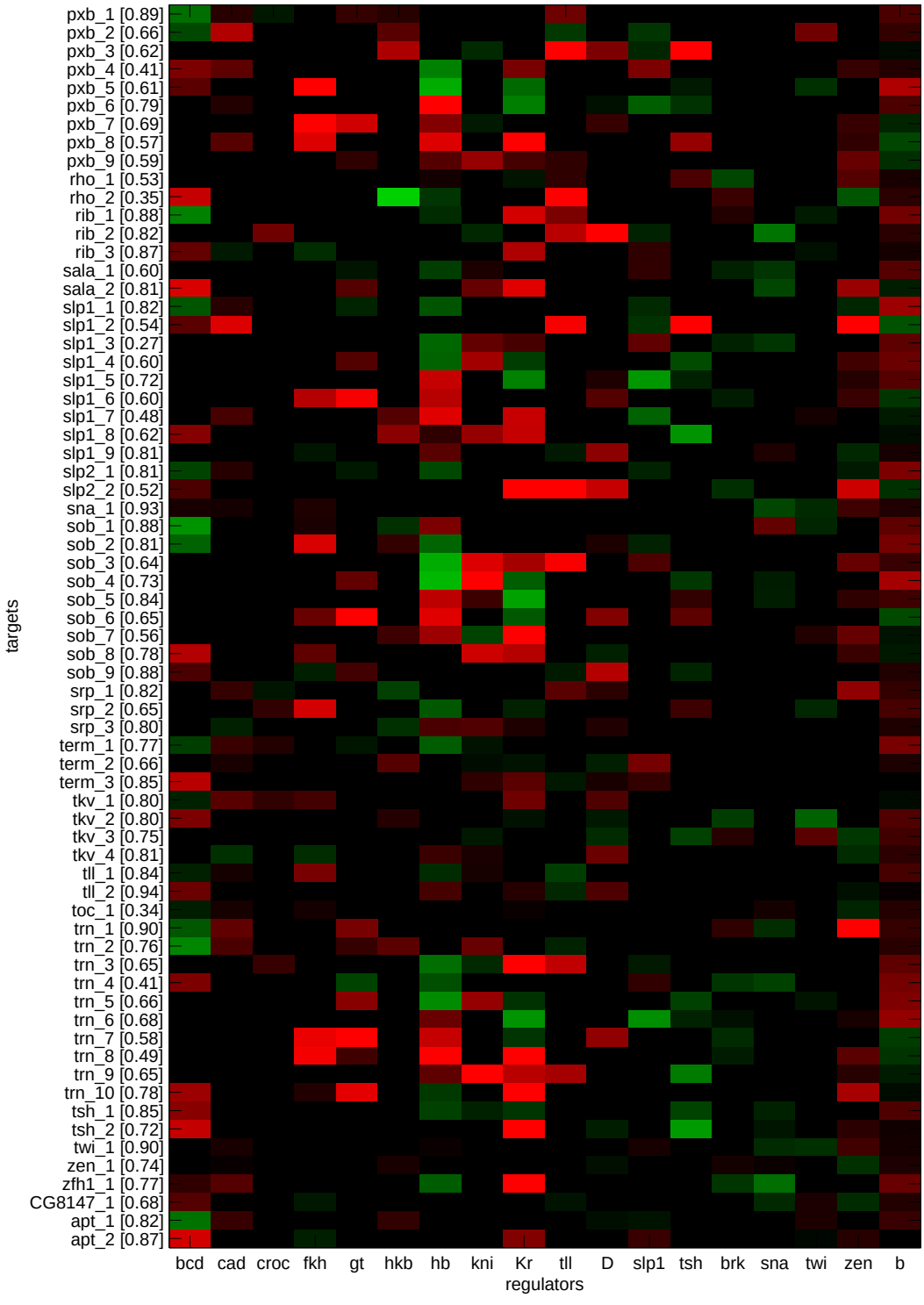
Supplementary Figure 7. Scatter plots in PointCloudXplore (PCX) allow the user to interactively explore average regulatory relationships between multiple genes. The left panel shows a cylindrical projection of the virtual embryo with levels of gt and eve expression displayed by the intensity of blue and red respectively. A set of cells along eve stripe 2 (green highlight) have been selected by the user by simply clicking on each cell to produce a so called cell selector. These cells simultaneously appear highlighted in green in the scatter plot (right panel) which shows the relative expression of eve as a function of gt. Data for all non-selected cells is plotted as grey points. Note that the result in the selected cells is similar to the plot in Figure 6 A, but in this case was generated via a simple graphical user interface and could just have easily been generated for any other subset of cells or combination of gene pairs. This, and other interactive tools, included in PCX for making sense of high-dimensional spatial expression data, is an integral part of our public data release.



Supplementary Figure 8. Predicted regulatory relations for all targets. Coefficients are shown for 17 regulators (columns) determined by fitting each target module (rows). Each row contains 6 non-zero entries corresponding to selected regulators for that target. Green indicates activation, red indicates repression, black indicates no interaction. The rightmost column indicates the constant offset b. Quality of fit (R^2) values are specified in brackets for each row.



Supplementary Figure 8. Predicted regulatory relations for all targets. (continued)



Supplementary Figure 8. Predicted regulatory relations for all targets. (continued)

gene	5:0-3	5:4-8	5:9-25	5:26-50	5:51-75	5:75-100
cad	-1.9	-5.7	-0.0	-7.7	-3.2	-5.3
gt	-7.0	-8.1	-12.5	-19.2	-7.4	-17.0
hb	-6.5	-6.8	-3.6	-7.7	-7.4	-4.9
kni	-3.2	-2.6	-10.3	-1.2	-1.2	-2.2
Kr	-4.7	-7.9	-6.4	-16.4	-13.5	-7.3
eve	-13.9	-36.6	-43.3	-48.1	-45.6	-47.9
ftz	-20.2	-38.1	-39.2	-46.0	-46.1	-47.4
h	-3.3	-18.6	-21.9	-17.6	-15.8	-7.9
prd	-10.5	-23.6	-22.8	-36.4	-23.1	-44.1
slp1	-1.3	-4.1	-4.8	-0.9	-17.1	-27.6
trn	-0.0	-7.2	-3.3	-24.6	-20.7	
odd				-37.5	-52.3	-46.7
croc		0.8	1.6	5.2	-2.2	-1.1
fkh	0.3	-4.0	-4.4	-7.1		-4.5
hkb	-2.2	-3.0	-6.6	-8.8	-4.8	-16.9
tll	-5.3	-3.5	-3.1	-11.6	-5.3	-1.6
rho		0.4	-3.5	-0.7	-1.0	-2.9
sna	-0.0	3.5		1.1	-3.3	-2.6
twi	-1.9	-0.9	-1.2	3.1	3.4	-1.0
zen	-0.9	-0.1	1.7	-6.6	-15.5	-13.8
D	-0.0	-3.2	-3.8	-8.3	-3.5	-0.7
knrl			-1.7	-7.8	1.3	-3.9
tsh			-0.8	-5.0	-13.3	-13.9
ave:	-3.60	-7.36	-8.25	-13.48	-12.94	-13.95

Supplementary Table 1. Percentage change in the average standard deviation of expression levels before and after fine registration. For each gene and cohort, the table entry indicates the change in the apparent average standard deviation of expression level across individual embryos in going from coarse alignment to fine registration. Values are given as a percentage of the measured variation for coarsely registered embryos. Registration removes significant variability for most genes, with the exception of d-v genes whose boundary locations are not well extrapolated from our marker genes. Interestingly, the importance of registration increases over time as pattern boundaries become more precisely localized.