

# Stability-driven nonnegative matrix factorization to interpret spatial gene expression and build local gene networks

Siqi Wu<sup>a,b</sup>, Antony Joseph<sup>a,b,c</sup>, Ann S. Hammonds<sup>b</sup>, Susan E. Celniker<sup>b</sup>, Bin Yu<sup>a,d,1</sup>, and Erwin Frise<sup>b,1</sup>

<sup>a</sup>Department of Statistics, University of California, Berkeley, CA 94720; <sup>b</sup>Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; <sup>c</sup>Walmart Labs, San Bruno, CA 94066; and <sup>d</sup>Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

Contributed by Bin Yu, March 6, 2016 (sent for review October 26, 2015; reviewed by Richard Bonneau and Michael S. Waterman)

Spatial gene expression patterns enable the detection of local covariability and are extremely useful for identifying local gene interactions during normal development. The abundance of spatial expression data in recent years has led to the modeling and analysis of regulatory networks. The inherent complexity of such data makes it a challenge to extract biological information. We developed *stanMF*, a method that combines a scalable implementation of nonnegative matrix factorization (NMF) with a new stability-driven model selection criterion. When applied to a set of *Drosophila* early embryonic spatial gene expression images, one of the largest datasets of its kind, *stanMF* identified 21 principal patterns (PP). Providing a compact yet biologically interpretable representation of *Drosophila* expression patterns, PP are comparable to a fate map generated experimentally by laser ablation and show exceptional promise as a data-driven alternative to manual annotations. Our analysis mapped genes to cell-fate programs and assigned putative biological roles to uncharacterized genes. Finally, we used the PP to generate local transcription factor regulatory networks. Spatially local correlation networks were constructed for six PP that span along the embryonic anterior–posterior axis. Using a two-tail 5% cutoff on correlation, we reproduced 10 of the 11 links in the well-studied gap gene network. The performance of PP with the *Drosophila* data suggests that *stanMF* provides informative decompositions and constitutes a useful computational lens through which to extract biological insight from complex and often noisy gene expression data.

principal patterns | stability selection | sparse decomposition |  
spatial gene expression | spatially local networks

Biological processes in multicellular organisms depend on spatial and temporal control of gene expression. Gene products function in the context of other spatially localized gene products and these interactions have been well characterized for development and tissue differentiation. Recent studies of prenatal (1) and adult human brain (2) revealed widespread anatomical variability in gene networks, which is reflective of developmental processes and of the distribution of major cell types. Spatially resolved studies of tumors uncovered widespread intratumor heterogeneity (3–8). Given the importance of spatiotemporal gene expression, many efforts are underway to characterize it genome-wide. Systematic datasets include *Drosophila* gene expression during embryogenesis [Berkeley *Drosophila* Genome Project (BDGP) (9)], and oogenesis (10), subcellular mRNA localization (11), and in brain (12), imaginal discs (13), central nervous system (14), and other developmental model systems [e.g., *Xenopus* (15), *Ciona* (16), and mouse (17–19)].

Spatial datasets are complex and quickly surpass the human ability to interpret them. To represent, search, and analyze such large spatial expression datasets, they are commonly curated with defined controlled vocabulary (9, 17–21). Curation using ontologies is time-consuming and requires expert knowledge. Despite significant progress toward automatic computer annotation through supervised learning based on human labels (22–26), the subtleties

inherent in spatial expression patterns are difficult to capture and finding related patterns is challenging. An alternative, complementary to ontologies, is the spatial expression information extracted directly from images (12, 17–19, 22, 27–30). We discovered putative gene interactions by correlating gene expression and performing cluster analysis (27), and others have used sparse Gaussian graphical models (30) to do the same. Due to data complexity and the large size of image collections, image-based approaches are not routinely used for modeling.

Organ systems develop through the combinatorial action of gene regulatory networks (21, 31), and gene function and regulatory interactions can markedly differ depending on the spatial location (32). Studies of genomic enhancer elements have shown that wild-type spatial expression patterns are actually the product of multiple genomic elements. These previous studies dissected biological enhancers and discovered that complex expression patterns could be subdivided into smaller regions (33, 34). In *Drosophila*, clustering early embryonic gene expression patterns recovered groups of cells that likely interact with one another, contributing to the formation of organs and tissues (27, 33). These regions are similar to those identified in studies using laser ablation to determine cell lineage and function (35, 36). Yakoby et al. proposed an innovative method to model spatial gene expression in *Drosophila* follicle cells as a Boolean combination of smaller building blocks (10). Due to the small number of gene expression patterns in their work, they were able to produce

## Significance

Despite the abundance of spatial gene expression data, extracting meaningful information to reveal how genes interact remains a challenge. We developed *stanMF*, a method that combines a powerful unsupervised learning algorithm, nonnegative matrix factorization (NMF), with a new stability criterion that selects the size of the dictionary or the set of principal patterns (PP). We demonstrate that PP give rise to a novel and concise representation of the *Drosophila* embryonic spatial expression patterns and they correspond to biologically meaningful regions of the *Drosophila* embryo. Furthermore, this new representation was used to automatically predict manual annotations, categorize gene expression patterns, and reconstruct the local gap gene network with high accuracy.

Author contributions: S.W., B.Y., and E.F. designed research; S.W., A.J., and B.Y. performed research; S.W., S.E.C., and B.Y. contributed new reagents/analytic tools; S.W., A.J., A.S.H., B.Y., and E.F. analyzed data; and S.W., A.S.H., S.E.C., B.Y., and E.F. wrote the paper.

Reviewers: R.B., New York University; and M.S.W., University of Southern California.

The authors declare no conflict of interest.

<sup>1</sup>To whom correspondence may be addressed. Email: binyu@stat.berkeley.edu or erwin@fruitfly.org.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1521171113/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1521171113/-DCSupplemental).

building blocks manually. Such an approach is intuitive and conceptually supported by the aforementioned works on genomic enhancers.

In this paper, we describe stability-driven nonnegative matrix factorization (staNMF), a method that interprets, represents, and analyzes comprehensive spatial gene expression datasets. staNMF partitions biological spatial data or images into spatial building blocks, called principal patterns (PP). Specifically, we adapted a powerful unsupervised learning algorithm, NMF (37), to learn data-driven representations from large and complex datasets. We invented a new stability criterion to address the challenge of NMF model selection to arrive at staNMF, which is scalable, tolerates experimental noise, and moderates image registration variance. We applied staNMF to a dataset of spatial gene expression images during early *Drosophila* embryogenesis. The output of staNMF, i.e., PP, is humanly interpretable and biologically meaningful. Using the PP, we grouped genes into overlapping categories corresponding to regions involved in coherent developmental programs. Finally, we built spatially local networks based on the learned PP, correctly reproducing 10 of 11 links in the well-studied gap gene network. This PP-based approach can be applied to extract biological insight from other complex and noisy spatial gene expression datasets, for example, from the extant zebrafish and mouse brain studies.

## Results

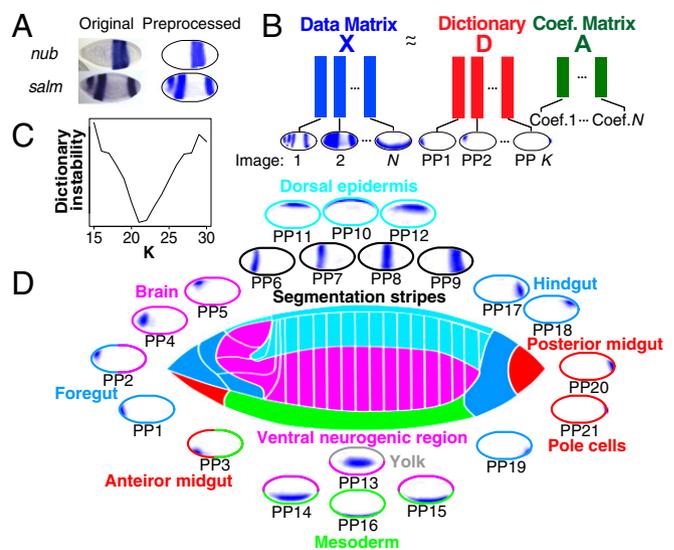
### Learned staNMF PP Correspond to Biologically Important Regions.

We used a set of BDGP *Drosophila* lateral embryonic gene expression images at developmental stages 4–6 (1 h 20 min–3 h after egg laying at 25°C). The 1,640 images were derived from 701 genes with spatially restricted expression patterns. We extracted embryos using segmentation of images taken with differential interference contrast microscopy. The objects containing the embryos were resized in both axes, and registered to a predefined elliptical template. The intensities of the gene expression patterns were determined with a least-squares approach to distinguish the dye intensity from the background introduced by the imaging modality (SI Appendix). Each resulting image is an ellipse of 405 pixels with gene expression values between 0 and 1 (Fig. 1A).

In *Drosophila* development, cell fates are determined before any visible morphological features become apparent (35, 36) and are preceded by the coordinated coexpression of cohorts of genes in defined spatial regimes that divide the embryo into areas with unique regulatory profiles (27, 33). We model the embryo as a topological map where genes are either absent or expressed at a positive value. Thus, we think of each spatial gene expression as an additive and nonnegative linear combination of a set of regions of the embryo. To identify these additive and positively valued regions, we used NMF (37). For a given positive integer  $K$ , NMF finds a data-driven dictionary such that each expression image can be represented by a nonnegative linear combination of the  $K$  dictionary columns (Fig. 1B). We converted the pixel intensities of the preprocessed expression pattern into a linear vector and decomposed the vector with NMF, aiming to solve the following nonconvex optimization problem:

$$\min_{\mathbf{D} \geq 0, \mathbf{A} \geq 0} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2,$$

where  $\mathbf{D}$  is the sought-after dictionary,  $\mathbf{X}$  the linearized image data,  $\mathbf{A}$  a linear nonnegative coefficient matrix, and  $\|\cdot\|_F$  the matrix Frobenius norm. To account for replicate images of the same gene, we introduced a weight factor (SI Appendix). The columns of the dictionary are the PP. The nonnegativity constraints on both the dictionary and coefficients enforce the PP to have nonnegative contributions to the gene image, resulting in a “parts-based” representation. The constraints also implicitly impose sparsity on both the PP and the linear coefficients.



**Fig. 1.** Learning PP by staNMF from spatial gene expression patterns. (A) Expression patterns of two genes, *nub* and *sal*m, in *Drosophila* embryos. (B) For a given number  $K$ , NMF factorizes the nonnegative data matrix  $\mathbf{X}$ , the columns of which are gene expression images, into the product of two nonnegative matrices: dictionary  $\mathbf{D}$ , which contains the  $K$  PP, and coefficient matrix  $\mathbf{A}$ , which contains the nonnegative coefficients of the images. (C) staNMF identified  $K=21$  to be the optimal number of PP for  $15 \leq K \leq 30$ . (D) The *Drosophila* fate map (center) (35, 36), surrounded by the 21 PP learned by staNMF. The PP are arrayed according to the corresponding regions of the fate map. See SI Appendix for how we mapped the PP to the fate map.

Our aim was a generalizable method contingent only on data and with little need of prior knowledge. NMF depends on a single parameter, the number  $K$  of PP. We reasoned that a useful definition of an optimal NMF-generated dictionary would be reproducibly independent of the initialization values. To identify the number of PP, we optimized on a metric that measures the instability of the learned PP relative to the initial starting points. In detail, for each  $K$ , the NMF algorithm was repeated multiple times with an initial PP dictionary randomly sampled from the columns of  $\mathbf{X}$ . We measured the instability of the PP by computing the average dissimilarity of all learned dictionary pairs ( $\mathbf{D}$  and  $\mathbf{D}'$ ) using their cross-correlation matrix (C) and a new Amari-type quantity (38):

$$\text{diss}(\mathbf{D}, \mathbf{D}') = \frac{1}{2K} \left( 2K - \sum_{j=1}^K \max_{1 \leq k \leq K} C_{kj} - \sum_{k=1}^K \max_{1 \leq j \leq K} C_{kj} \right).$$

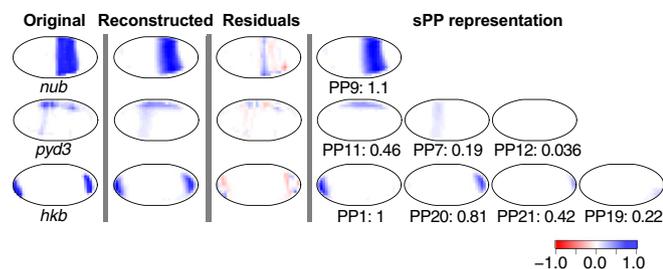
We repeated the process for each  $K$  and selected the  $K$  where the learned dictionaries achieve the lowest instability (SI Appendix). Among all dictionaries with the optimal size  $K$ , the dictionary with the minimum NMF objective function value was chosen. We called this stability-based NMF model selection method “staNMF” and validated our method with multiple synthetic datasets (SI Appendix, Figs. S4, S6, and S7). When applied to our spatial gene expression pattern dataset with range  $15 \leq K \leq 30$ , staNMF identified the number of PP with the lowest instability at  $K=21$  (Fig. 1C; see also SI Appendix).

The 21 learned PP divided the *Drosophila* embryo into contiguous pretissue and organ regions (Fig. 1D). Compared with principal component analysis, independent component analysis (39), and a recently proposed sparse Bayesian factor model (24), only PP recapitulate the underlying biology of cell and tissue fate maps (SI Appendix, Figs. S8 and S9). Each PP is spatially coherent: The intensity is locally continuous and the regions defined by the PP are interconnected. We grouped the 21 PP into four

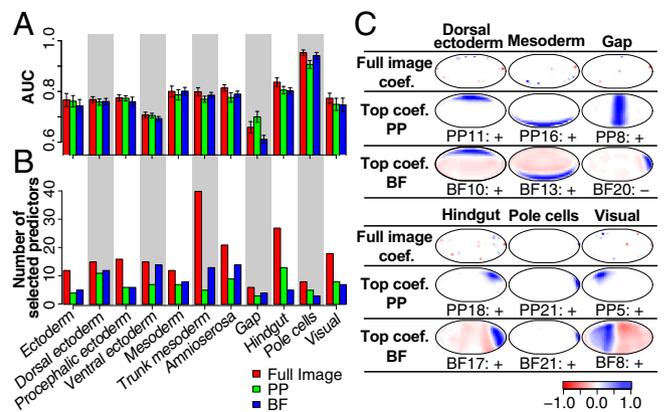
categories: PP1–5: anterior patterns; PP6–9: vertical (gap) segmentation stripes; PP10–16: horizontal ventral–dorsal patterns; and PP17–21: posterior patterns. We compared the PP and the categories to the *Drosophila* fate map (35, 36), an experimentally determined functional mapping of spatial regions before availability of gene expression data. We associated each PP or a group of PP with a region in the fate map of similar size and shape (Fig. 1D; see also *SI Appendix*). We found that the PP refined the fate map in the dorsal epidermal region, the ventral neurogenic region, the mesoderm, and the hindgut. Some of the refinements are already biologically supported. For example, the vertical stripes are known to be the result of gap, pair-rule, polarity, and segmentation genes that eventually establish 14 refined stripes that become morphologically distinguishable in a later-stage embryo (40).

**$l_1$  Regularization to Provide Sparse PP Data Representations.** We evaluated the ability of PP to provide a compact representation for spatial gene expression patterns. A sparse decomposition of complex expression patterns into additive smaller components offers a simple and intuitive computational representation of spatial gene expression (Fig. 2). Whereas the nonzero coefficients of the matrix **A** (Fig. 1B) provided such a decomposition, they tended to select more PP than necessary in our simulations (*SI Appendix*, Fig. S11). Instead, we used the least absolute shrinkage and selection operator (LASSO) (41) for PP selection. Nonnegative constraints were put on the linear coefficients. For each expression pattern, we chose the LASSO regularization parameter using a 10-fold cross-validation and refitted the coefficients with nonnegative least squares on the selected PP (*SI Appendix*). We call these coefficients “sparse PP” (sPP) coefficients or representation. The average number of PP chosen by this procedure is 10.4, and the average correlation between the original expression pattern and the reconstructed pattern is 0.854 (*SI Appendix*, Fig. S12 A and B). Considering the small number of the selected PP, the correlation measure indicates that our model selection and fitting procedure achieved a reasonably good reconstruction quality. As expected, the correlation increases as the number of PP increases (*SI Appendix*, Fig. S12C). We investigated cases with poor performance and found such gene expression patterns are either faint or have poorly defined boundaries. In addition, nonsparse representations almost always correspond to ubiquitously expressed genes (*SI Appendix*, Fig. S12D). As illustrated by the residual images, errors are most likely to occur at expression pattern boundaries (Fig. 2).

**PP Provide a Data-Driven Alternative to Human Expert Annotations.** Expert curators annotated BDGP spatial gene expression patterns with a controlled vocabulary whose terms represent anatomical regions of the developing embryo, similar to the fate map



**Fig. 2.** Sparse decomposition of spatial gene expression patterns using the 21 PP. Shown are a sample of three gene expression patterns (original), their reconstructed patterns using the sPP representation (reconstructed), the difference between the original and the reconstructed patterns (residuals) and the contributions from the 21 PP (sPP representation).



**Fig. 3.** Predicting annotation terms based on 405 image pixels, the sPP, and the BF (sparse BF) representations. (A) Prediction accuracy as evaluated by the AUC value. Data are expressed as mean  $\pm$  SEM. (B) Number of selected predictors in the optimal model. (C) Interpretability of the L1LR under all three representations. The pixel-based full image representation: all coefficients are shown as pixel values within the embryo; the sPP-based and the BF-based representations: only the PP or BF that corresponds to the largest L1LR coefficient is shown. “+”: the largest L1LR coefficient is positive; “-”: the largest L1LR coefficient is negative.

discussed above. To compare the 21 learned PP with the anatomical vocabulary, we used the sPP coefficients as predictors in a supervised learning approach and labeled each image with annotation terms. We selected 11 stage 4–6 annotation terms with more than 100 images: ectoderm anlage in statu nascendi (AISN), dorsal ectoderm AISN, procephalic ectoderm AISN, ventral ectoderm AISN, mesoderm AISN, trunk mesoderm AISN, amnioserosa AISN, gap, hindgut AISN, pole cells, and visual AISN. For each of the 11 terms, we labeled images annotated this term as “1,” the rest as “0,” and fitted an  $l_1$ -penalized logistic regression (L1LR) with the sPP coefficients as predictors. A 10-fold cross-validation was performed for each term to choose the regularization parameter in the L1LR. To compare with sPP, we also trained L1LR using the full expression pattern with 405 pixels, and the sparse Bayesian factors (BF) of ref. 24 (*SI Appendix*). We generated 21 BF to compare directly to the 21 PP.

The prediction performance of the three methods is very similar, as measured by the cross-validation AUC (area under the receiver operating characteristic curve, valued between 0 and 1) (Fig. 3A). On average, the AUC value for the sPP representation is 0.772, compared with 0.787 for the pixel-based representation and 0.767 for the BF representation. Taking into account the SE of the AUC for each annotation term, none of the three methods significantly outperforms the others. In terms of model complexity, on average 17 predictors are selected for the pixel-based L1LR, 7 for our sPP-based approach, and 8 for the BF-based model (Fig. 3B).

Among the three L1LR models, the sPP-based model is the most interpretable and most biologically meaningful (Fig. 3C). For the pixel-based model, we created a visualization of the 405 predictors for each annotation term by plotting the L1LR coefficient values as pixels in our elliptic embryo shape. To compare with this visualization, we selected the (top L1LR) PP and BF corresponding to the largest L1LR coefficients for their respective L1LR models. The pixel-based predictors consist of scattered points and the top L1LR BF contains negative values, both of which are difficult to interpret. In contrast, the top L1LR PP consistently showed the annotation term exactly as a curator would annotate the gene expression (Fig. 3C; see *SI Appendix*, Figs. S13–S17 for all L1LR results as evidence for the benefits of

the sPP models). In addition, the pixel-based and the BF-based representations resulted in unstable predictor sets between cross-validation runs (*SI Appendix, Fig. S18*). This instability further reduces the interpretability of the two models. In the automatic label prediction task, the better interpretability of the top L1LR PP and the comparable prediction accuracy of sPP-based models provided additional support for the PP as a biologically attractive decomposition.

**PP Associated Gene Functions and Relationships Between Spatial Regions.** We use the term “function” as defined by the experimentally generated fate map that describes the locations of larval/adult progenitor cells in the blastoderm. These cells give rise to particular tissues and organs during development. By systematically associating genes to PP, we can group known and uncharacterized genes and assign putative biological roles. We assigned a gene to PP category  $k$  if the  $k$ th sPP coefficients of the gene exceeded 0.1. The number of genes in each of the 21 PP categories is, on average, 300 genes ranging from 184 to 395. PP categories 6–9 contain fewer, on average, 223 genes (Fig. 4A, *Right* and *SI Appendix, Fig. S19* and *Tables S2* and *S3*). In addition, we also found a significant presence of previously uncharacterized computed genes (CG) in all PP categories: The average percentage of CG per PP category is 23.4%.

To directly relate genes to each other, we created a heatmap visualization of the sPP coefficients for 667 genes that belong to at least one PP category. We ordered the genes by first associating each of them to the PP with the maximum sPP coefficient, and then performing a hierarchical clustering of the genes assigned to the same PP (Fig. 4A, *Left*; see also *SI Appendix*). A surprisingly large fraction of genes (17.8%) exhibit their strongest expression in PP21 (pole cells) and have limited expression in other PP. We found that only 5.8% of the 156 transcription factors are among these PP21 specific genes, confirming previous results (21). Of the 667 genes, 4.5% have their strongest expression in segmentation patterns PP6–9, suggesting that only a small number of genes are dedicated to segmentation. Furthermore, 93.3% of these genes have been characterized, implying that we know most segmentation genes. We found genes with known roles in foregut development (*croc*, *hkb*, and *kni*) associated

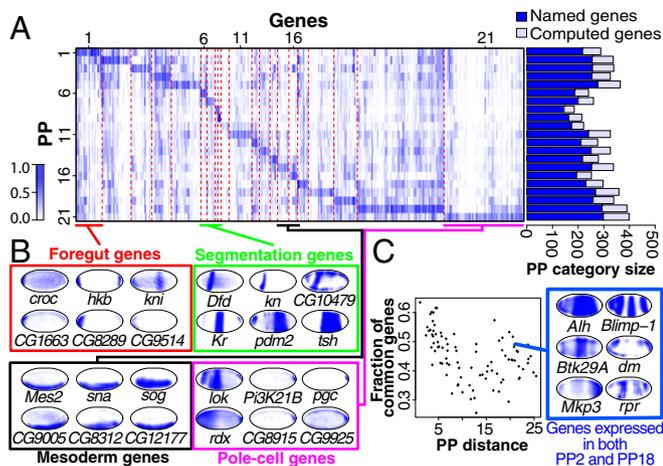
with PP1, segmentation specific genes (*Dfd*, *kn*, *Kr*, and *tsh*) associated with PP6–9, genes essential for mesoderm/ectoderm development (*Mes2*, *sna*, and *sog*) associated with PP15, genes essential for pole-cell formation associated with posterior PP21 (*lok*, *pgc*, and *rdx*) as well as previously uncharacterized genes such as *CG1663*, *CG8289*, *CG9514*, and *CG10479* in these PP categories (Fig. 4B). With additional later-stage organ system annotation data (21), we found genes expressed in PP16 (mesoderm) in stages 4–6 become expressed in the central nervous system (CNS) starting at stage 9 (*trx*, *sna*, *Traf4*, and *Caf1*). Early mesoderm genes with function during CNS development have been shown before (42), but here we demonstrate a systematic secondary function of mesoderm specific genes, including previously uncharacterized genes (e.g., *CG11247*).

Next, we investigated the relationship between the PP that span the anterior–posterior axis, i.e., PP1–9, PP17–21. We plotted the fraction of common genes in a pair of PP categories, defined as the Jaccard distance between the two categories, in relation to the pairwise PP centroid distance (Fig. 4C and *SI Appendix, Fig. S20*). Our results show that when the PP distance is small, the fraction of common genes is high. However, after the initial decrease, the fraction of common genes increases as the PP distance increases. An example is the set of genes (49% or 227) shared between the distant PP2 and PP18 that map to anterior foregut/brain and posterior hindgut (Fig. 4C). These genes include known foregut and hindgut development genes such as *Alh*, *Blimp-1*, *Btk29A*, *dm*, *Mkp3*, and *rpr*. This finding substantiates the previously identified common origins and gene expression signatures of foregut and hindgut that were based on manual annotations (21, 36). Similarly, 229 genes (52%) are shared between PP3 (anterior midgut/mesoderm) and PP19 (hindgut), including known midgut and hindgut genes, *ry*, *Ect4*, *Sdc*, *Pcl*, *larp*, and *emc*, suggesting a more general link between the anterior and posterior patterns.

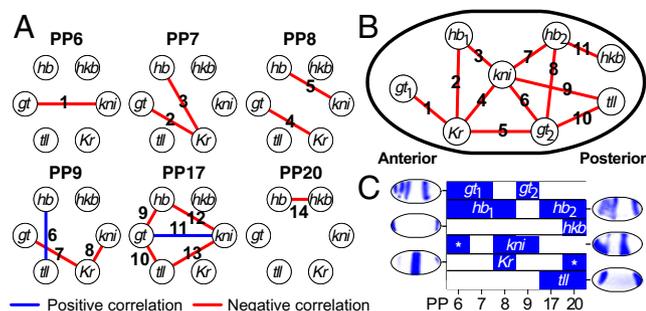
**PP-Based Correlation Network Inference Leads to Accurate de Novo Reconstruction of the *Drosophila* Gap Gene Network.** Associations between two genes are routinely described by their correlation to each other (43). In terms of spatial relationships, positive gene interactions exhibit spatial overlap whereas repressive gene interactions exhibit spatial exclusivity. Below, we used the learned PP to construct spatially local correlation networks (SLCN) for 156 transcription factors (TF) from our expression data.

The *Drosophila* gap gene network has been studied for decades (40, 44, 45). It controls embryonic patterning by regulating the genes required to establish the anterior–posterior segmentation stripes and is primarily driven by well-studied activating and repressive interactions between eight TF. To reconstruct this network solely from our expression data of 156 spatially restricted TF, we selected six PP (PP6–9, PP17, and PP20) corresponding to the domains of the gap gene network. We called these six PP “gap-PP.” For each gap-PP, we computationally constructed an SLCN. First, we identified all TF with sPP coefficient greater than 0.1 in the gap-PP, or its directly adjacent PP. Next we computed the weighted correlation between all pairs of selected TF using the pixel intensities in the gap-PP as weights. Finally, links of each resulting correlation network were filtered by thresholding the weighted correlations at a two-tail 5% cutoff, producing six SLCN with interactions among known and previously uncharacterized genes (*SI Appendix, Figs. S21–S24*). See *SI Appendix* for details of the SLCN construction.

We evaluated our SLCN construction by comparing interactions found in the six SLCN to known regulatory interactions of selected trunk and terminal gap genes, *giant(gt)*, *hunchback(hb)*, *knirps(kni)*, *Krüppel(Kr)*, *huckebein(hkb)*, and *tailless(tll)*. We compared the subnetworks of the SLCN containing only the six genes (Fig. 5A) to a schematic network diagram (Fig. 5B), as originally depicted in ref. 44. Although the diagram indicates that some gene interactions are contingent on spatial position, it



**Fig. 4.** PP-based gene categorization. (A) Left heatmap: PP expression profile of genes. Each column corresponds to the sPP coefficients of one gene. Between the red dashed lines are the genes with the strongest expression in the same PP (indicated by numbers on horizontal axis). Right barplot: numbers of named and computed genes (CG) in each PP category. (B) Genes with known functions and CG were found in the associated PP categories. (C) The relationship between the fraction of common genes in a pair of PP categories and the centroid distance of the two PP, for PP1–9, PP17–21. Each dot in the plot corresponds to a PP pair. Shown also are six genes expressed in both PP2 (brain/foregut) and PP18 (hindgut), a pair of distant PP.



**Fig. 5.** Modeling and validation of the *Drosophila* gap gene network with SLCN. (A) The SLCN for six gap genes. For each of the six gap-PP, the subnetwork of the SLCN that contains the six gap genes is shown. Links are numbered from 1 to 14. (B) The gap gene network diagram depicting repressive interactions of six genes (44). Links are numbered from 1 to 11 and multiple occurrences of the same gene are subscripted by numbers (e.g.,  $hb_1$  and  $hb_2$ ). The directions of the interactions are not indicated. (C) Expression patterns of the six gap genes and their linearly ordered PP representation. For each gene, the regions depicted in blue are the gap-PP with sPP coefficient greater than or equal to 0.1. The "\*" symbol indicates a region of gene expression with no match in B.

does not provide precise locations of the interactions. To compare with our networks, we devised a method to match the links in the diagram to our SLCN. For each gap gene, we first created a linearly ordered PP representation by placing the six gap-PP anterior to posterior and associating a gap-PP to the gene if the sPP coefficient for the gap-PP exceeded a threshold of 0.1 (Fig. 5C). The gap-PP associated with each gap gene were then merged into one or more connected PP groups. Based on its relative location in the diagram, we then matched each gene node in the schematic diagram to a connected PP group for the same gene. We considered an interaction between two gene nodes in the schematic network diagram as successfully identified by our method if the same interaction exists in the any SLCN associated with the overlapping PP in the connected PP groups of the two gene nodes (SI Appendix, Fig. S25 and Table S4).

For example, the diagram depicts a repressive link between the anterior component of *gt* (i.e.,  $gt_1$ ) and *Kr*. Using our linearly ordered PP representation, we found the connected PP groups for  $gt_1$  and *Kr* are PP6/7 and PP8, respectively (Fig. 5C). We searched for the *gt*-*Kr* interaction only in the SLCN of PP7 and PP8, because PP6 and PP8 do not overlap. In both networks, we found a repressive interaction (or negative correlation). Hence we considered the anterior *gt*-*Kr* link of the schematic gap gene network diagram as being identified with our model. See SI Appendix, Fig. S25 for the validation of the remaining links.

For the six gap genes, our SLCN reconstruction identified 14 interactions (Fig. 5A). Eight out of 11 links in the gap gene network diagram have a one-to-one mapping with 8 of the 14 SLCN interactions. In addition, the two *gt*-*Kr* links in the gap gene network (link 1 and 5 in Fig. 5B) are found in the SLCN of PP7–9 (links 2, 4, and 7 in Fig. 5A). The remaining *kni*- $gt_2$  link (link 6 in Fig. 5B) has no corresponding link in the SLCN. Three of the 14 SLCN links do not correspond to any interactions in the network diagram: the two *gt*-*kni* links in PP6 and PP17 (link 1 and link 11 in Fig. 5A), and the *hb*-*tll* link in PP9 (link 6 in Fig. 5A). Therefore, our SLCN recovered 10 out of 11 interactions in the gap gene network and discovered 3 interactions not described by the diagram. In contrast, when using correlation over the whole embryo to construct the TF network, we recovered only three out of nine unique links of the gap gene network (SI Appendix, Fig. S27).

## Discussion

We proposed staNMF that combines NMF with a new stability-based model selection criterion to decompose spatial gene

expression patterns into local PP. When we applied staNMF to *Drosophila* embryonic expression data at early stages 4–6, the learned PP correspond to preorgan regions, and thus provide an informative representation of spatial gene expression data. We demonstrated that PP are a data-driven alternative to manual curation and facilitate the categorization of gene expression patterns. Our PP-based sparse representations (sPP) reduce large datasets to manageable scales. They allow suitable human interrogation and downstream computation on desktop computers while preserving quantitative relationships of full datasets. In addition, staNMF's utility was further substantiated by the agreement between our PP-based spatially local networks and the well-studied gap gene network.

Model selection or identification of a well-reasoned number of components for unsupervised learning has been a challenging problem. staNMF's underlying idea of stability was previously used to identify the number of clusters in cluster analysis using NMF (46). Our contribution is to use an Amari-type measure to evaluate dictionary stability, rather than clustering stability as in the previous work. We experimented with the method of ref. 46, and found that it failed to identify the correct number of PP in a number of synthetic datasets (SI Appendix). Recent Bayesian model selection approaches (47, 48) introduced additional hyperparameters, which in practice are generally not known in advance. We believe staNMF is an important advance because it does not depend on tuning parameters, and has been demonstrated to work well in both simulations and our *Drosophila* spatial gene expression data.

Our SLCN identified three network links previously not described in ref. 44. In PP6, we found a repression link between *gt* and *kni*. Gene expression images of *gt* and *kni* revealed a clear complementary pattern toward the anterior end with a negative local correlation of  $-0.720$  in PP6 (SI Appendix, Fig. S25D). In the PP17 SLCN, an activation link between *kni* and *gt* was identified. Because our images covered an interval of around 1.5 h, the posterior part of *kni* expression pattern at the early developmental stages 4–6 might have been aligned to the *gt* gene posterior end at a later time point (SI Appendix, Fig. S25D). Experiments will be needed to confirm or refute these predicted links. Finally, although not described in ref. 44, the predicted *hb*-*tll* activation link in PP9 is supported by ref. 49.

Given the successes of our PP-based approach in the well-characterized early *Drosophila* embryo, we expect staNMF to be broadly applicable to derive meaningful data-driven representations of spatial gene expression for other systems such as zebrafish, *Caenorhabditis elegans*, and human histological samples. In conclusion, we have demonstrated with ample evidence the utility of sPP as an effective computational lens to reveal hidden structures in complex gene expression data.

## Datasets and Software

Data are available as Datasets S1–S6, and code and datasets are provided under "Principal Patterns" on our website: [insitu.fruitfly.org/downloads](http://insitu.fruitfly.org/downloads).

**ACKNOWLEDGMENTS.** We are indebted to BDGP members for their advice. We thank Ben Brown, Julien Mairal, and Sivaraman Balakrishnan for helpful discussions and comments. B.Y. acknowledges partial research support from National Science Foundation (NSF) Grants DMS-1107000, CDS&E-M55 1228246, and DMS-1160319 (Focused Research Groups in the Mathematical Sciences), Army Research Office Grant W911NF-11-1-0114, Air Force Office of Scientific Research Grant FA9550-14-1-0016, and the Center for Science of Information, an NSF Science and Technology Center, under Grant Agreement CCF-0939370. National Institutes of Health Grants R01 GM076655 (to S.E.C. and A.S.H.) and R01 GM097231 (to E.F., S.E.C., S.W., and A.J.) supported this work. Work at Lawrence Berkeley National Laboratory was conducted under Department of Energy Contract DEAC02-05CH11231. S.W. was partially supported by National Human Genome Research Institute Grant 1U01HG007031-01, and Citadel Fellowship in the Department of Statistics at University of California, Berkeley.

1. Miller JA, et al. (2014) Transcriptional landscape of the prenatal human brain. *Nature* 508(7495):199–206.
2. Hawrylycz MJ, et al. (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489(7416):391–399.
3. Almendro V, et al. (2014) Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Reports* 6(3):514–527.
4. Almendro V, et al. (2014) Genetic and phenotypic diversity in breast tumor metastases. *Cancer Res* 74(5):1338–1348.
5. Bedard PL, Hansen AR, Ratain MJ, Siu LL (2013) Tumour heterogeneity in the clinic. *Nature* 501(7467):355–364.
6. Gerlinger M, et al. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 366(10):883–892.
7. de Bruin EC, et al. (2014) Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 346(6206):251–256.
8. Zhang J, et al. (2014) Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 346(6206):256–259.
9. Tomancak P, et al. (2002) Systematic determination of patterns of gene expression during drosophila embryogenesis. *Genome Biol* 3(12):research0088.1-88.14.
10. Yakoby N, et al. (2008) A combinatorial code for pattern formation in Drosophila oogenesis. *Dev Cell* 15(5):725–737.
11. Lécuyer E, et al. (2007) Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* 131(1):174–187.
12. Jenett A, et al. (2012) A GAL4-driver line resource for Drosophila neurobiology. *Cell Reports* 2(4):991–1001.
13. Jory A, et al. (2012) A survey of 6,300 genomic fragments for cis-regulatory activity in the imaginal discs of Drosophila melanogaster. *Cell Reports* 2(4):1014–1024.
14. Manning L, et al. (2012) A resource for manipulating gene expression and analyzing cis-regulatory modules in the Drosophila CNS. *Cell Reports* 2(4):1002–1013.
15. Pollet N, et al. (2005) An atlas of differential gene expression during early Xenopus embryogenesis. *Mech Dev* 122(3):365–439.
16. Imai KS, Hino K, Yagi K, Satoh N, Satou Y (2004) Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: Towards a comprehensive understanding of gene networks. *Development* 131(16):4047–4058.
17. Smith CM, et al. (2007) The mouse gene expression database (gxd): 2007 update. *Nucleic Acids Res* 35(Database issue):D618–D623.
18. Richardson L, et al. (2009) Emage mouse embryo spatial gene expression database: 2010 update. *Nucleic Acids Res* 38(Database issue):D703–D709.
19. Lein ES, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445(7124):168–176.
20. Tomancak P, et al. (2007) Global analysis of patterns of gene expression during Drosophila embryogenesis. *Genome Biol* 8(7):R145.
21. Hammonds AS, et al. (2013) Spatial expression of transcription factors in Drosophila embryonic organ development. *Genome Biol* 14(12):R140.
22. Peng H, et al. (2007) Automatic image analysis for gene expression patterns of fly embryos. *BMC Cell Biol* 8(Suppl 1):S7.
23. Zhou J, Peng H (2007) Automatic recognition and annotation of gene expression patterns of fly embryos. *Bioinformatics* 23(5):589–596.
24. Pruteanu-Malinici I, Mace DL, Ohler U (2011) Automatic annotation of spatial expression patterns via sparse Bayesian factor models. *PLOS Comput Biol* 7(7):e1002098.
25. Pruteanu-Malinici I, Majoros WH, Ohler U (2013) Automated annotation of gene expression image sequences via non-parametric factor analysis and conditional random fields. *Bioinformatics* 29(13):i27–i35.
26. Yuan L, et al. (2014) Automated annotation of developmental stages of Drosophila embryos in images containing spatial patterns of expression. *Bioinformatics* 30(2):266–273.
27. Frise E, Hammonds AS, Celniker SE (2010) Systematic image-driven analysis of the spatial Drosophila embryonic expression landscape. *Mol Syst Biol* 6:345.
28. Puniyani K, Faloutsos C, Xing EP (2010) SPEX2: Automated concise extraction of spatial gene expression patterns from Fly embryo ISH images. *Bioinformatics* 26(12):i47–i56.
29. Mace DL, Varnado N, Zhang W, Frise E, Ohler U (2010) Extraction and comparison of gene expression patterns from 2D RNA in situ hybridization images. *Bioinformatics* 26(6):761–769.
30. Puniyani K, Xing EP (2013) GINI: From ISH images to gene interaction networks. *PLOS Comput Biol* 9(10):e1003227.
31. Adryan B, Teichmann SA (2010) The developmental expression dynamics of Drosophila melanogaster transcription factors. *Genome Biol* 11(4):R40.
32. Schulz C, Tautz D (1994) Autonomous concentration-dependent activation and repression of Krüppel by hunchback in the Drosophila embryo. *Development* 120(10):3043–3049.
33. Kvon EZ, et al. (2014) Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* 512(7512):91–95.
34. Stanojevic D, Small S, Levine M (1991) Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science* 254(5036):1385–1387.
35. Lohs-Schardin M, Cremer C, Nüsslein-Volhard C (1979) A fate map for the larval epidermis of Drosophila melanogaster: Localized cuticle defects following irradiation of the blastoderm with an ultraviolet laser microbeam. *Dev Biol* 73(2):239–255.
36. Hartenstein V (1993) *Atlas of Drosophila Development* (Cold Spring Harbor Laboratory Press, Plainview, NY).
37. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755):788–791.
38. Amari S, Cichocki A, Yang H (1996) A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA), pp 757–763.
39. Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Networks* 10(3):626–634.
40. Nüsslein-Volhard C, Wieschaus E (1980) Mutations affecting segment number and polarity in Drosophila. *Nature* 287(5785):795–801.
41. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Royal Statist Soc B* 58(1):267–288.
42. Ashraf SI, Hu X, Roote J, Ip YT (1999) The mesoderm determinant snail collaborates with related zinc-finger proteins to control Drosophila neurogenesis. *EMBO J* 18(22):6426–6438.
43. Wang YX, Huang H (2014) Review on statistical methods for gene network reconstruction using expression data. *J Theor Biol* 362:53–61.
44. Jaeger J (2011) The gap gene network. *Cell Mol Life Sci* 68(2):243–274.
45. Nüsslein-Volhard C, Kluding H, Jürgens G (1985) Genes affecting the segmental subdivision of the Drosophila embryo. *Cold Spring Harb Symp Quant Biol* 50:145–154.
46. Brunet JP, Tamayo P, Golub TR, Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci USA* 101(12):4164–4169.
47. Tan VY, Févotte C (2013) Automatic relevance determination in nonnegative matrix factorization with the  $\beta$ -divergence. *IEEE Trans Pattern Anal Mach Intell* 35(7):1592–1605.
48. Sun M, Zhang X (2015) A stable approach for model order selection in nonnegative matrix factorization. *Pattern Recognit Lett* 54:97–102.
49. Margolis JS, et al. (1995) Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element. *Development* 121(9):3067–3077.

Stability driven nonnegative matrix factorization to interpret  
spatial gene expression and build local gene networks

## **Supporting Information Appendix**

Siqi Wu, Antony Joseph, Ann S. Hammonds, Susan E. Celniker, Bin Yu and Erwin Frise

# Table of Contents

<b>1</b>	<b>Methods</b>	<b>4</b>
1.1	Data and preprocessing . . . . .	4
1.2	Expression pattern extraction . . . . .	4
1.3	Nonnegative matrix factorization (NMF) . . . . .	5
1.4	StaNMF: stability based NMF model selection . . . . .	6
1.5	Aligning PP on the <i>Drosophila</i> fate map . . . . .	9
1.6	Representing spatial expression patterns by the learned PP . . . . .	10
1.7	Predicting annotation terms . . . . .	11
1.8	PP associated gene functions: heatmap visualization of sPP coefficients . . . . .	11
1.9	Constructing spatially local correlation networks (SLCN) . . . . .	11
1.10	Correlating genes on the whole embryo . . . . .	13
<b>2</b>	<b>List of Datasets</b>	<b>14</b>
<b>3</b>	<b>Supporting Figures and Tables</b>	<b>15</b>

## List of Supporting Figures

S1	Visual evaluation of embryo registration. . . . .	15
S2	Extracting gene expression patterns from images obtained through differential interference contrast (DIC) microscopy. . . . .	16
S3	Construction of the dictionary in Simulation Experiment 1. . . . .	17
S4	NMF model selection using staNMF: Simulation Experiment 1. . . . .	18
S5	NMF model selection using Brunet et al.’s clustering instability criterion [1]: Simulation Experiment 1. . . . .	19
S6	NMF model selection: Simulation Experiment 2 – the Swimmer dataset. . . . .	20
S7	Using StaNMF and Brunet et al.’s stability criterion on the <i>Drosophila</i> spatial gene expression data and the corresponding denoised data. . . . .	21
S8	A comparison between the 21 principal patterns (PP) and the 21 sparse Bayesian Factors (BF) . . . . .	22
S9	Principal component analysis (PCA) and independent component analysis (ICA) for the <i>Drosophila</i> gene expression data. . . . .	23
S10	The PP learned with parameter $K = 22$ instead of the optimal $K = 21$ . . . . .	24
S11	Effectiveness of the LASSO+NLS model selection and fitting procedure. . . . .	25
S12	Spatial gene expression reconstruction quality of the sparse PP representation . . . .	26
S13	Interpretability of the L1LR model under the pixel-based, the sPP and the BF representations. . . . .	27
S14	Top L1LR PP for all 11 annotation terms. . . . .	28
S15	The full L1LR coefficients for annotation prediction using the sPP representation. . .	29
S16	Top L1LR BF for all 11 annotation terms. . . . .	30
S17	The full L1LR coefficients for annotation prediction using the BF representation. . .	31
S18	Stability analysis of the set of selected L1LR predictors for three representations . .	32
S19	Number of genes expressed in each of the 21 PP. . . . .	35
S20	The relationship between the fraction of common genes in a pair of PP categories and the centroid distance of the two PP . . . . .	36
S21	Histograms of local correlations for the six gap-PP. . . . .	37
S22	Spatially local correlation networks (SLCN) for transcription factors (TF): PP6 and PP7. . . . .	38
S23	Spatially local correlation networks (SLCN) for transcription factors (TF): PP8 and PP9. . . . .	39
S24	Spatially local correlation networks (SLCN) for transcription factors (TF): PP17 and PP20. . . . .	39
S25	Validating the SLCN with the gap gene network. . . . .	40
S26	Sub-networks of the six SLCN that contain the six gap genes, for sPP coefficient threshold in $\{0.05, 0.1, 0.2, 0.3\}$ . . . . .	42
S27	Correlating transcription factors (TF) over the whole embryo . . . . .	43

## List of Supporting Tables

S1	The adjacent PP for the six gap-PP. . . . .	12
S2	Genes in PP categories 1-10. . . . .	33

S3	Genes in PP categories 11-21. . . . .	34
S4	Validating the SLCN with the gap gene network. . . . .	41

# 1 Methods

## 1.1 Data and preprocessing

We collected images from the Berkeley *Drosophila* Genome Project (BDGP) database (March 2013) [2] labeled as developmental stages 4 – 6 with lateral view embryos. We removed images corresponding to genes annotated solely with the terms “ubiquitous”, “maternal” or “no staining”. Compared to restricted zygotic expression patterns, these images had almost uniform expression intensities throughout the embryos and were therefore less meaningful for our analysis. The resulting dataset contained 1640 images.

We used our previous segmentation algorithm [3] to detect the outline of the embryo in each image. Based on the detected outline, we employed SPEX2 [4] to register the embryo onto an ellipse template with a long axis of 64 pixels and a short axis of 32 pixels, contained in a rectangle image of  $64 \times 128$  pixels. We then extracted the gene expression pattern of the embryo using a least squares (LS) approach detailed in the next section. By taking the average gene expression intensity of every four pixels that form a square, the ellipse-registered embryo was down-sampled by a factor of four to fit in an ellipse template with a long axis of 16 pixels and a short axis of 8 pixels. Such an ellipse template can be embedded in a rectangle image of  $16 \times 32$  pixels for visualization. Inside the rectangle image, there are 405 pixels within the ellipse and  $16 \times 32 - 405 = 107$  pixels outside. To validate our registration pipeline, we selected replicates of the same gene and genes with known adjacent expression patterns, superimposed them, and visually evaluated the matches to deem them satisfactory (Fig. S1). The dataset that contains the preprocessed expression patterns for 1640 images can be found in *Dataset S1*.

## 1.2 Expression pattern extraction

The *Drosophila* embryonic gene expression images were captured using differential interference contrast (DIC) microscopy. However, the shadows induced by DIC are frequently indistinguishable from expression patterns [3]. We developed a least squares (LS) based method using the color channels to differentiate spatial gene expression from background. Using Adobe<sup>®</sup> Photoshop<sup>®</sup>, we created a training set of 32 images by manually selecting the regions of the embryos with gene expression, as detected by the blue dye (Nitro blue tetrazolium and 5,5'-dibromo-4,4'-dichloro-indigo). For each pixel inside the ellipse template, we averaged the three RGB values as proxy for gene expressing intensity and set the intensity outside the segmented region to zero. We then standardized the expression intensity  $g$  of each pixel using the formula  $(255 - g)/255$ . Here, 255 is the number of possible gray-scale values. Under this standardization scheme, the maximum gene expression intensity is one and the minimum is zero.

For  $i = 1, \dots, 32$ , we represented the  $i$ -th manually processed image as a vector  $\mathbf{s}_i \in [0, 1]^{8192}$ . The vector length 8192 was derived from  $128 \times 64$  – the size of the rectangle image that contained the ellipse template. We modeled the gene expression intensity of each pixel as a linear combination of the color information of its neighbor pixel at different scales. More specifically, for each pixel and each color channel, we generated up to the fourth moment of the intensity at the pixel and within a disk centered at the pixel with radius 2, 4, and 8. Denote the feature matrix by  $\mathbf{Z}_i \in \mathbb{R}^{8192 \times 49}$ . Each row of  $\mathbf{Z}_i$  corresponds to a pixel and each column a feature (1 column for the intercept, 4 moments  $\times$  4 radii  $\times$  3 channels = 48 features). To obtain the linear coefficients  $\mathbf{b} \in \mathbb{R}^{49}$ , an LS

was used based on the 32 images:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b} \in \mathbb{R}^{49}} \sum_{i=1}^{32} \|\mathbf{s}_i - \mathbf{Z}_i \mathbf{b}\|_2^2.$$

The correlation between the predicted gene expression and the manually extracted gene expression was 0.9832. This number was quite high considering the fact that we had in our training data  $8192 \times 32 = 262144$  pixels in total but only 48 features.

To extract the gene expression pattern for a new image, we first computed the feature vector  $\mathbf{Z}_{new}$  and set  $\mathbf{s}_{new} = \mathbf{Z}_{new} \hat{\mathbf{b}}$ . Since the gene expression intensity should value between zero and one, we further truncated each entry of  $\mathbf{s}_{new}$  to be in  $[0, 1]$ . We then used the resulting vector as the extracted gene expression pattern for further analysis. We tested our gene expression extraction procedure on a number of images not in the training set. The results indicated that our method performed very well in terms of the correlation between the gene expression pattern extracted by the curator and the one predicted by the LS method (Fig. S2).

### 1.3 Nonnegative matrix factorization (NMF)

Nonnegative Matrix Factorization, or NMF, is a popular unsupervised learning algorithm that can learn parts-based representation from the input data [5]. NMF has been applied to many fields such as image processing and computer vision [6], text mining [7], audio signal separation [8] and bioinformatics [1], see [9] for a recent review of applications of NMF. Recently, a new nonnegative tensor factorization algorithm similar to NMF was developed and applied to parts of the BDGP gene expression data [10]. However, they did not address the issue of model selection and their derived dictionary lacked biological interpretability. Here, we designed an NMF model selection method and our dictionary (PP) can be validated using the *Drosophila* fate map and human annotations.

Denote by  $\mathbb{R}_+$  the nonnegative real line. Let  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$  be the data matrix where each column represents a data vector. For a given positive integer  $K$ , NMF finds a dictionary  $\mathbf{D} \in \mathbb{R}_+^{d \times K}$  under which each vector  $\mathbf{x}_i$  has nonnegative representations: i.e.  $\mathbf{x}_i \approx \mathbf{D} \mathbf{a}_i$  for a nonnegative vector  $\mathbf{a}_i \in \mathbb{R}_+^K$ . More precisely, NMF aims at solving the following nonconvex optimization problem:

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{A}=[\mathbf{a}_1, \dots, \mathbf{a}_N]} \|\mathbf{X} - \mathbf{D} \mathbf{A}\|_F^2 &= \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{D} \mathbf{a}_i\|_2^2, \\ \text{subject to } \mathbf{D} &\geq 0, \|\mathbf{D}[k]\|_\infty \leq 1 \text{ for } k = 1, \dots, K, \\ &\text{and } \mathbf{a}_i \geq 0, \text{ for } i = 1, \dots, N. \end{aligned}$$

Here,  $\mathbf{D}[k]$  is the  $k$ -th column of the dictionary  $\mathbf{D}$ . Note that the above formulation of NMF does not require the data matrix  $\mathbf{X}$  to be nonnegative in every entry. For some numerical examples,  $\mathbf{X}$  is the product of two nonnegative matrices plus noise and hence can be negative in some entries. See, e.g. our Simulation Experiment 1 in the below section. In our *Drosophila* data application,  $\mathbf{x}_i$  is a vector of length 405 that corresponds to the  $i$ -th preprocessed spatial gene expression pattern. The columns of the dictionary  $\mathbf{D}$  are the Principal Patterns (PP). To account for the possibility of multiple replicates for one gene, we used a weighted version of NMF by changing the objective function to the following form:

$$\sum_{i=1}^N \mathbf{w}[i] \|\mathbf{x}_i - \mathbf{D} \mathbf{a}_i\|_2^2,$$

where the weight for the  $i$ -th image  $\mathbf{w}[i]$  is the reciprocal of the number of replicates of the gene that corresponds to the  $i$ -th image (from now on, we will denote the  $j$ -th entry of a vector  $\mathbf{v} \in \mathbb{R}^m$  as  $\mathbf{v}[j]$ ). Note that the above objective function can be rewritten as  $\mathbf{w}[i] \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2^2 = \|\sqrt{\mathbf{w}[i]}\mathbf{x}_i - \mathbf{D}(\sqrt{\mathbf{w}[i]}\mathbf{a}_i)\|_2^2$ . Therefore we can simply set  $\mathbf{x}'_i = \sqrt{\mathbf{w}[i]}\mathbf{x}_i$  and use any algorithm that solves original NMF formulation, with  $\mathbf{X}' = [\mathbf{x}'_1, \dots, \mathbf{x}'_N]$  as the new data matrix. Denote by  $(\hat{\mathbf{D}}, \hat{\mathbf{A}}')$  the output of the NMF algorithm. The nonnegative coefficient matrix  $\hat{\mathbf{A}}$  can be retrieved by the scaling the  $i$ -th column of the matrix  $\hat{\mathbf{A}}'$  by the factor  $\mathbf{w}[i]^{-1/2}$ .

To compute NMF, we used the SPAMS package with the MATLAB interface [11]. SPAMS implemented a number of online algorithms for dictionary learning and matrix factorization. The package is fast, and scales to large numbers of data points. The NMF algorithm in this package required the input of an initial guess for the dictionary, which we constructed by randomly sampling  $K$  columns from the data matrix  $\mathbf{X}$ . To compute the dictionary, SPAMS performed alternating minimization: given the current iteration of the dictionary  $\mathbf{D}$ , update the nonnegative coefficients  $\mathbf{a}_i$ 's using nonnegative least squares (NLS); and given the nonnegative coefficients, update the dictionary  $\mathbf{D}$  by solving another NLS. We ran the algorithm until convergence, which took about 200 to 300 iterations for the *Drosophila* spatial gene expression dataset. Since the optimization problem is nonconvex, the output dictionary  $\hat{\mathbf{D}}$  depended on the initial dictionary. In the next section, we will explain how we made use of this property to choose the number  $K$  of dictionary columns.

#### 1.4 StaNMF: stability based NMF model selection

In this section, we will address the issue of choosing the number  $K$  of dictionary columns in NMF. As discussed in the previous section, SPAMS solved NMF by an alternating minimization algorithm. As a result, the output dictionary depended on the initial value. We reasoned that the  $K$  should be chosen such that the output dictionary is most reproducible, or stable, for different initializations. We proposed staNMF, a model selection procedure that combined multiple runs of NMF with a new Amari-type criterion to measure the instability of the output dictionaries.

For each  $K$ , we ran the NMF algorithm  $B$  times. Typically,  $B = 100$  for the *Drosophila* gene expression data and other simulated examples. For each NMF run, we chose an initial dictionary whose columns were randomly sampled from those of  $\mathbf{X}$ . The  $B$  runs of NMF generated output dictionaries  $\hat{\mathbf{D}}_b$  for  $b = 1, \dots, B$ . Now we introduce a measure that quantifies the stability of the  $B$  dictionaries. Let  $\mathbf{C} \in \mathbb{R}^{K \times K}$  be the cross correlation matrix between the columns of two dictionaries  $\mathbf{D}_1$  and  $\mathbf{D}_2$  having the same number of columns  $K$ . For a matrix  $\mathbf{H} \in \mathbb{R}^{m \times n}$ , denote by  $\mathbf{H}[j, k]$  its  $(j, k)$ -th entry. Since the columns of a dictionary are permutation invariant, to measure dissimilarity between  $\mathbf{D}_1$  and  $\mathbf{D}_2$ , we designed the following Amari-type quantity:

$$\begin{aligned} diss(\mathbf{D}_1, \mathbf{D}_2) &= \frac{1}{2} \left( \frac{1}{K} \sum_{j=1}^K \left( 1 - \max_{1 \leq k \leq K} \mathbf{C}[k, j] \right) + \frac{1}{K} \sum_{k=1}^K \left( 1 - \max_{1 \leq j \leq K} \mathbf{C}[k, j] \right) \right) \\ &= \frac{1}{2K} \left( 2K - \sum_{j=1}^K \max_{1 \leq k \leq K} \mathbf{C}[k, j] - \sum_{k=1}^K \max_{1 \leq j \leq K} \mathbf{C}[k, j] \right). \end{aligned}$$

Note that when  $\mathbf{D}_2$  can be transformed into  $\mathbf{D}_1$  by column permutation,  $diss(\mathbf{D}_1, \mathbf{D}_2) = 0$ . Such a definition was inspired by Amari et al. [12], who used a comparable quantity to measure the

performance of their blind signal separation algorithm. The discrepancy of all  $B$  dictionaries for  $K$  was measured by the average Amari-type error of all  $B(B - 1)/2$  pairs of dictionaries:

$$\Upsilon(K) = \frac{2}{B(B - 1)} \sum_{1 \leq b < b' \leq B} \text{diss}(\hat{\mathbf{D}}_b, \hat{\mathbf{D}}_{b'}).$$

We selected  $K$  that achieved a small  $\Upsilon(K)$ , i.e. a small discrepancy or instability. Once the parameter  $K$  was determined, we selected the learned dictionary with the minimum NMF square loss among all  $B$  dictionaries.

The idea of using stability for NMF model selection was first introduced by Brunet et al. [1]. However, their stability metric was substantially different from ours. In their paper, NMF was used for cluster analysis. They proposed to choose  $K$  such that their NMF cluster assignment is most stable. Given a dictionary  $\mathbf{D}$  with  $K$  columns, they assigned the data vector  $\mathbf{x}_i$  to the  $k$ -th cluster, if the nonnegative coefficient for the  $k$ -th dictionary column has the highest value among all  $K$  coefficients. If more than one dictionary column share the same coefficient value, the data point is assigned to any of the corresponding clusters with equal probability. For the clustering defined by NMF, they constructed the connectivity matrix  $S$ , whose  $(i, j)$ -th entry is set to one if the  $i$ -th and the  $j$ -th data points belong to the same cluster, and zero otherwise. Based on the  $B$  NMF runs, they computed the consensus matrix,  $\bar{S}$ , which was defined as the average of all connectivity matrices. They then used the cophenetic correlation coefficient based on  $\bar{S}$  to measure the clustering stability of NMF. In their paper, the cophenetic correlation coefficient was defined as the Pearson correlation coefficient of (1) the distance between the  $i$ -th and  $j$ -th data points as measured by  $1 - \bar{S}[i, j]$  and (2) the distance between the  $i$ -th and  $j$ -th data points induced by the average linkage hierarchical clustering using  $\bar{S}$  as the similarity matrix, for all  $1 \leq i < j \leq N$  (recall that  $N$  is the number of data points). The closer the cophenetic correlation coefficient to 1, the more stable the clustering assignment. To compare with our method, we used the equivalent one minus the cophenetic correlation coefficient, which is now a measure for *clustering instability*, and strived for a minimum value.

As defined earlier, our proposed criterion  $\Upsilon(K)$  measured the *instability of the output dictionaries* with respect to different initial inputs. We tested our staNMF as well as Brunet et al.’s method on a number of synthetic data with a known ground truth dictionary. While both methods identified the same  $K$  for some examples (Simulation Experiment 1 and 3), it is not surprising that Brunet et al.’s method failed on the others (e.g. Simulation Experiment 2), as their method was designed for the purpose of cluster analysis. Our staNMF performed consistently well, demonstrating the reliability of our method to identify the dictionary for applications with data similar to ours. When applied to the *Drosophila* spatial gene expression data, both methods agreed on  $K = 21$ , which, gave additional supporting evidence that the number of PP suggested by staNMF was optimal for our dataset. In what follows, we will detail our simulation experiments and the real data application.

### *Simulation Experiment 1*

In this experiment, we investigated how the above-mentioned stability based methods behave for dictionary with different coherence and linear coefficients with various sparsity. It has been shown in the dictionary learning literature that increased dictionary coherence, or collinearity between dictionary columns, might lead to ill-posedness of the dictionary learning formulation [13–16]. Empirically, we also found it hard to use NMF to recover the dictionary if the columns were highly

collinear. Therefore, we suspected that it is difficult for the stability-based methods to identify the correct  $K$  as the coherence of the dictionary increases.

We generated our data in the following way. Let  $\text{coh} \in \{0, 1, \dots, 10\}$ . We constructed the ground truth dictionary  $\mathbf{D}_0 \in \mathbb{R}^{200 \times 20}$  as:

$$\mathbf{D}_0[j, k] = \begin{cases} 1, & \text{for } 1 + (k - 1)(10 - \text{coh}) \leq j \leq 10 + (k - 1)(10 - \text{coh}), \\ 0, & \text{otherwise.} \end{cases}$$

See Fig. S3 for an illustration of the above dictionary construction. Under this construction, each dictionary column has exactly 10 entries equal to one and the remaining 190 entries equal to zero. Furthermore, two consecutive dictionary columns share  $\text{coh}$  entries that are equal to one in common. Thus, the parameter  $\text{coh}$  controls the coherence of the dictionary, which is defined as the maximum absolute inner product between dictionary columns. Next, we generated the entries of the coefficient matrix  $\mathbf{A}_0 \in \mathbb{R}^{20 \times 1000}$  as independent and identical Bernoulli random variable with success probability  $0 < \mathbf{p} \leq 1$ . Set the data matrix  $\mathbf{X} = \mathbf{D}_0 \mathbf{A}_0 + \mathbf{E}$ , where  $\mathbf{E} \in \mathbb{R}^{200 \times 1000}$  was a noise matrix with entries drawn independently and identically from a Gaussian distribution with mean zero and standard deviation 0.1.

For each combination of  $(\mathbf{p}, \text{coh})$ , we ran NMF  $B = 100$  times for  $10 \leq K \leq 30$  and then applied our stability criterion. The results shown in the Fig. S4 indicated that when the dictionary coherence was low, the measure for dictionary instability,  $\Upsilon(K)$ , had a clear minimum at  $K = 20$  which was the true number of dictionary columns. However, as the dictionary coherence increased, for example,  $\text{coh} = 6$ ,  $\Upsilon(K)$  as a function of  $K$  changed shape and multiple local minima emerged. This observation supported our previous conjecture that a higher dictionary coherence made it more difficult for the stability-based method to identify the correct number of  $K$ . It is unclear how the sparsity parameter  $\mathbf{p}$  affected our stability criterion.

Brunet et al.’s method behaved similar on the same data (Fig. S4). However, we found that the clustering instability measure was versatile across the range of  $K$  and had too many abrupt local minima. On the other hand, our measure of dictionary instability  $\Upsilon(K)$  was much more continuous and predictable. For example, for  $\mathbf{p} = 1$  and  $\text{coh} = 2$ , Brunet et al.’s stability curve had two almost identical local minima: one at  $K = 10$  and the other at  $K = 20$ . In this case, their method was not robust: slight contamination of the data might mislead their method to consider  $K = 10$  as the best number of dictionary columns. For the same example, staNMF gave a very clear minimum at  $K = 20$ .

*Simulation Experiment 2: the Swimmer data*

In this example, we evaluated staNMF with a dataset that has been widely applied in the NMF literature: the Swimmer data [17–19]. The dataset contained 256 images each of  $32 \times 32$  pixels depicting all possible gestures of an artificial swimmer (Fig. S6A and B). For each image, each limb of the swimmer was chosen from one of four gestures for that limb. The true dictionary therefore consisted of  $4 \times 4 = 16$  columns and so the number of all possible combinations of the swimmer gestures was  $4^4 = 256$ .

For this data, our method recovered the correct  $K = 16$  (Fig. S6C). However, Brunet et al.’s method chose  $K = 14$  (Fig. S6D). To elucidate the reason, we noted that each swimmer image had equal contribution from four dictionary columns. Thus under the ground truth dictionary, each image should be assigned to the corresponding four clusters simultaneously. However, Brunet et al.’s approach forced the image to belong to only one cluster. As a result, it would select any one of the four clusters with equal probability. The randomness of an image falling into one of the four

clusters resulted into clustering instability at  $K = 16$ . In contrast, staNMF never assumed any clustering structure and so it also succeeded for a dataset like the Swimmer data.

*Real Data: Drosophila gene expression patterns*

We applied both stability-based criteria to our *Drosophila* spatial gene expression data for  $15 \leq K \leq 30$ . The dictionary learned with  $K < 15$  resulted in PP that were in general too broad, as compared to the pre-organ partitions in the *Drosophila* fate map. These PP also led to poor reconstruction quality when using them to represent the gene expression patterns. Dictionary learned with  $K > 30$  resulted in PP that were too unstable. For  $15 \leq K \leq 30$ , both staNMF and Brunet et al.’s method identified  $K = 21$  as the optimal number of PP (Fig. S7A).

We noticed that our stability criterion,  $\Upsilon(K)$  had similar values for  $K = 21$  and  $K = 22$ . When comparing the  $K = 22$  dictionary with the  $K = 21$  dictionary, we found that three PP from the  $K = 22$  dictionary, PP4, PP5 and PP6, were different from the corresponding PP4 and PP5 of the  $K = 21$  dictionary (Fig. S10). In particular, PP5 (brain) in the  $K = 21$  dictionary was split into PP5 and PP6 (both recognized as brain region) in the  $K = 22$  dictionary. The remaining 19 PP were essentially unchanged. Thus the PP learned using the two different  $K$  were very similar. For simplicity we chose  $K = 21$ .

The learned dictionary for  $K = 21$  can be found in *Dataset S3*.

*Simulation Experiment 3: the denoised Drosophila data*

In Simulation Experiment 1, we demonstrated that dictionary coherence might affect the two stability based model selection criteria. As a sanity check for our real data application, we generated a dataset using the 21 PP learned from the *Drosophila* data and investigated whether staNMF can recover the correct number of PP from this artificial data. Specifically, denote by  $\hat{\mathbf{D}} \in \mathbb{R}_+^{405 \times 21}$  the learned dictionary which contains the 21 PP and  $\hat{\mathbf{A}} \in \mathbb{R}_+^{21 \times 1640}$  the corresponding nonnegative coefficient matrix. We generated the data matrix  $\hat{\mathbf{X}} \in \mathbb{R}_+^{405 \times 1640}$  as the “denoised” version of the original data matrix:  $\hat{\mathbf{X}} = \hat{\mathbf{D}}\hat{\mathbf{A}}$ . For this dataset, both staNMF and Brunet et al.’s method selected  $K = 21$  as the optimal number of PP (Fig. S7B).

### 1.5 Aligning PP on the *Drosophila* fate map

The *Drosophila* fate map [20,21] is a schematic diagram depicting pre-organ regions of a *Drosophila* embryo. To map the computationally derived PP to the fate map, we first identified a few PP which we are confident belong to certain regions of the schematic map and assigned the rest according to their relative positions and shapes. Additionally, we validated our assignments by finding genes with known biological roles using the PP categories (Table S2 and S3) and later stage annotation data [22]. For example, PP1 can be easily mapped to foregut, PP10 to dorsal epidermis, PP16 to ventral mesoderm and PP21 to pole cells as these PP occupy the four corners of the embryo. Next, for the anterior patterns, we identified PP4 and PP5 from their locations and shapes as the brain region of the fate map. These assignments were further substantiated by a number of nervous system genes associated with the two PP (e.g. *numb*, *oc*, *D* and *Doc1*). PP2 is between the brain and the foregut region and we found genes expressed in PP2 associated with either organ, e.g. *hb* and *tll* in brain and *oc* and *hbn* in foregut. Therefore we labeled it as either brain or foregut. PP3 is most likely to be the anterior midgut or anterior ventral mesoderm regions because it is directly beneath PP1 (foregut) and overlaps with PP16 (mesoderm). Genes expressed in PP3 include known midgut genes (e.g. *egg* and *ry*) and known mesoderm genes (e.g. *croc* and *Mes4*). PP6–9

are vertical segmentation patterns that were hinted in the fate map. These vertical patterns are known to be the result of gap, pair-rule, polarity and segmentation genes that eventually establish 14 refined stripes. For the horizontal patterns, PP11 and PP12 are both above the midline of the embryo and hence can be treated as dorsal epidermis region. The below embryo midline PP14 and PP15 can be either the ventral neurogenic region or mesoderm, and there is evidence supporting that parts of the later central nervous system is derived from the mesoderm [23]. PP13 is most often associated with the ventral neurogenic regions (e.g. *SoxN* and *ind*) and the yolk region of the embryo (not part of the fate map) (e.g. *aay* and *llp4*). For the posterior patterns, PP20 is directly to the left of PP21 (pole cells) and so we labeled it as midgut. This mapping is also supported by the fact that many midgut genes are expressed in PP20, including *sc*, *Bgb*, *esg* and *Moe*. PP17 is labeled as hindgut since it is similar in shape and size to the hindgut region of the fate map. PP18 and PP19 are directly above and below PP17 respectively and so they were labeled as hindgut as well. Moreover, we found hindgut genes such as *Abd-B*, *Mkp3* and *D19A* in PP17, *Doc1*, *ebi* and *dm* in PP18 and *byn*, *apt* and *twi* in PP19, further supporting our mapping of PP17–19 to the fate map.

## 1.6 Representing spatial expression patterns by the learned PP

Using nonnegative least squares (NLS), the NMF algorithm gave a nonnegative linear representation matrix of the data under the learned dictionary. With the nonnegativity as an implicit sparsity penalty, NLS can be treated as a method to perform model selection [24]. However, empirically we found that NLS selected more covariates than necessary. To demonstrate this, we generated 1000 data vectors using the model discussed in Simulation Experiment 1 with dictionary coherence  $\text{coh} = 2$  and random linear coefficient sparsity  $\text{p} = 0.2$ . For each of the data vector generated, we applied NLS to estimate the nonnegative linear coefficients with the ground truth dictionary as the covariate matrix. The resulting NLS coefficients contained many more nonzeros than the true nonnegative linear coefficients used to generate the data (Fig. S11A and B). The average support difference between the estimated coefficient and the true coefficient was 7.29, out of a maximum of 40.

To address this issue, we employed the following LASSO+NLS procedure. Let  $\mathbf{x} \in \mathbb{R}_+^d$  be a data vector and  $\mathbf{D} \in \mathbb{R}^{d \times K}$  the dictionary or covariate matrix. We first used the LASSO, or least absolute shrinkage and selection operator [25], with the nonnegative constraints on the linear coefficients:

$$(\hat{\mu}, \hat{\beta}(\lambda)) = \arg \min_{\mu \in \mathbb{R}_+, \beta \in \mathbb{R}_+^K} \|\mathbf{x} - \mathbf{D}\beta - \mu\|_2^2 + \lambda \|\beta\|_1.$$

With a 10-fold cross-validation, the LASSO regularization parameter  $\lambda$  was chosen to be the largest among all parameters whose cross-validation error was within one standard error of the minimum cross-validation error. Denote by  $\hat{\beta}_{\text{lasso}}$  the nonnegative linear coefficient at the selected  $\lambda$ .

Due to the  $l_1$ -penalty term, the LASSO estimator is biased towards zero for finite samples. Similar to [26], in order to reduce the bias, we fitted NLS on the dictionary columns selected by the LASSO. Let  $S = \{k : \hat{\beta}_{\text{lasso}}[k] \neq 0\} \subset \{1, \dots, K\}$  be the nonzero set of coefficients and  $\mathbf{D}[S]$  be the submatrix of  $\mathbf{D}$  with columns indexed by  $S$ . We solved the following NLS problem:

$$(\hat{\nu}, \hat{\gamma}) = \arg \min_{\nu \in \mathbb{R}_+, \gamma \in \mathbb{R}_+^{|S|}} \|\mathbf{x} - \mathbf{D}[S]\gamma - \nu\|_2^2,$$

where  $|S|$  is the size of the set  $S$ . The *sparse PP (sPP) representation* or *sPP coefficient* for the

data vector  $\mathbf{x}$ , denote by the vector  $\alpha \in \mathbb{R}_+^K$ , is such that the entries indexed by  $S$ ,  $\alpha[S] = \hat{\gamma}$  and the entries indexed by the complement of  $S$ ,  $\alpha[S^c] = 0$ .

We applied the LASSO+NLS procedure to our previous simulation example. The distribution of number of nonzero estimated coefficients per observation now matched that of the true coefficients (Fig S11C). The average support difference between the two reduced significantly to 0.3.

We used the R package `glmnet` [27] for the computation.

## 1.7 Predicting annotation terms

We used  $l_1$ -penalized sparse logistic regression models (L1LR) (see, e.g. [28]) to predict the 11 annotation terms. For each term, denote by  $L[i]$  the label of the  $i$ -th image for  $i = 1, \dots, 1640$ :  $L[i] = 1$  if the gene corresponding to the image was labeled as expressed in this term and  $L[i] = 0$  otherwise. To predict the label vector  $L$ , we fitted L1LR using three different covariate sets: (1) the 405 pixels for the pixel-based representation, (2) the 21 sPP coefficients based on the LASSO+NLS procedure, and (3) the 21 sparse Bayesian Factor (BF) coefficients based on [29]. For each annotation term, the observations in each class are weighted by the reciprocal of the corresponding class size so that the two classes are of the same importance. A 10-fold cross-validation was performed and the  $l_1$ -penalization parameter was chosen such that it was the largest among all parameters whose cross-validation Area Under the ROC Curve (AUC) was within one standard error of the maximum AUC. As before, we used the R package `glmnet` [27] for computation. The annotation data is available as *Dataset S2*.

## 1.8 PP associated gene functions: heatmap visualization of sPP coefficients

We created a heatmap visualization of the sPP coefficients for 667 genes that belong to at least one PP category (Fig. 4A). First, for  $k = 1, \dots, 21$ , we defined the  $k$ -th sPP coefficient of a gene to be the maximum  $k$ -th sPP coefficient among all the replicate patterns of the same gene. We then assigned a gene to the  $k$ -th cluster if its  $k$ -th PP has the maximum sPP coefficient. Within each of the 21 clusters, we performed a hierarchical clustering using the correlation of the sPP coefficients as similarity measure. We arranged the genes in each cluster linearly using the order returned by the hierarchical clustering algorithm and assembled all clusters to produce the heatmap.

## 1.9 Constructing spatially local correlation networks (SLCN)

We selected six PP, namely PP6-PP9, PP17 and PP20, which we referred to as *gap-PP* in the main text, to model the gap gene segmentation networks. For each of the six gap-PP, we identified its directly adjacent PP by visual inspection. The PP neighbors are summarized in the following Table S1:

For each of the six gap-PP, we found all transcription factors (TF) expression patterns in the category of the gap-PP, or its directly adjacent PP, with sPP coefficient greater than 0.1. This excluded TF with low or no expression in the gap-PP and its nearby regions and hence reduced the possibility of spurious correlations. Denote this set of patterns by  $T$ . We then computed the weighted correlations for the expression patterns in  $T$  with the  $l_1$ -normalized PP intensity as the weight vector. Specifically, let  $\mathbf{u} \in \mathbb{R}^d$  be a nonnegative vector whose entries sum up to 1 and  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$  represent two data vectors, e.g. two gene expression patterns in  $T$ . The local

	Adjacent PP
PP6	PP4, PP7
PP7	PP6, PP8
PP8	PP7, PP9
PP9	PP8, PP17
PP17	PP9, PP20
PP20	PP17, PP21

**Table S1:** The adjacent PP for the six gap-PP.

correlation between  $\mathbf{x}_1$  and  $\mathbf{x}_2$  with weight  $\mathbf{u}$  is defined as:

$$\text{cor}_{\mathbf{u}}(\mathbf{x}_1, \mathbf{x}_2) = \frac{\text{cov}_{\mathbf{u}}(\mathbf{x}_1, \mathbf{x}_2)}{\text{var}_{\mathbf{u}}(\mathbf{x}_1)^{1/2} \text{var}_{\mathbf{u}}(\mathbf{x}_2)^{1/2}},$$

where

$$\text{var}_{\mathbf{u}}(\mathbf{x}_1) = \sum_{j=1}^d \mathbf{u}[j] (\mathbf{x}_1[j] - \mathbf{x}_1^T \mathbf{u})^2, \text{ and}$$

$$\text{cov}_{\mathbf{u}}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=1}^d \mathbf{u}[j] (\mathbf{x}_1[j] - \mathbf{x}_1^T \mathbf{u})(\mathbf{x}_2[j] - \mathbf{x}_2^T \mathbf{u}).$$

Note that when the  $\mathbf{u}[j] = 1/d$  for all  $1 \leq j \leq d$ , the above correlation is the same as the sample correlation between data vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$ .

As mentioned, many genes had multiple replicate images. For a pair of genes, we defined the local correlation of the two genes to be the local correlation with the maximum magnitude between replicate images of one gene and replicate images of the other. For simplicity, we called this correlation the *maximum correlation*, but we note that it can be the most positive or the most negative correlation. By computing this maximum correlation, we stated that two genes were highly correlated if any of the replicates of the two genes were highly correlated. Spatial expression patterns for some genes changes rapidly within the stage range considered in this paper. For example, significant differences in gene expression were observed for the replicate expression patterns of *hb* and *kni* (Fig. S1). Using maximum local correlation can therefore help to identify those highly variable genes that were likely to interact at some point in the developmental timeline.

For each gap-PP, we computed the local correlation for all pairs of genes in the gene set  $T$  defined earlier. The distribution of the correlations was bimodal, with one peak corresponding to positive correlations and the other to negative correlations (Fig. S21). This is due to the way we defined local correlation of two genes, which excluded image pairs that had low correlations.

To construct the spatially local network for each gap-PP, we set a positive edge between two genes if their local correlation is above the upper 5 percentile of all local correlations for the PP and set a negative edge between two genes if their local correlation is below the lower 5 percentile. See Figs. S22 – S24 for the six spatially local correlation networks for the gap-PP. To validate our network construction, we compared the sub-networks of the six SLCN containing the six gap genes to the schematic gap gene network (main text Fig. 5 and in expanded detail in *SI Appendix* Fig. S25). The six sub-networks are robust for different sPP coefficient thresholds (Fig. S26).

## 1.10 Correlating genes on the whole embryo

We compared our PP-based local network results to those obtained by correlating the expression patterns over the whole embryo, or global correlation analysis. Similar to local correlation, the global correlation of the two TF is defined as the largest correlation between the replicates of the two. Next, we specified a cutoff value for the global correlation in order to form network links. We first combined the six PP-based SLCN into a single network such that two TF share a link in the new network if they share a link in at least one of the six SLCN, regardless of the sign of the link. For fair comparison between the global and local approaches, the cutoff values for the global correlation network was chosen such that (1) the resulting network has the same number of links as in the previous combined network and, (2) the number of positive links is the same as the number of negative links. We converted the original schematic gap gene network to the “global version” without the spatial information accordingly: two gap genes share a link if they share a link in the schematic gap gene network diagram regardless of the location of the interaction (Fig. S27C). Only three links out of nine links in the global version of the gap gene network were recovered (Fig. S27D). An analysis of the relationship between the local and global correlations indicated that, while for some gene-gene interactions global correlation is positively correlated with local correlation, many others have negative correlations (Fig. S27A). For example, *gt* and *hb* are known to be mutual repressors of one another towards the posterior end of the embryo. The global correlation was unable to detect this relationship whereas the local correlation succeeded (Fig. S27B).

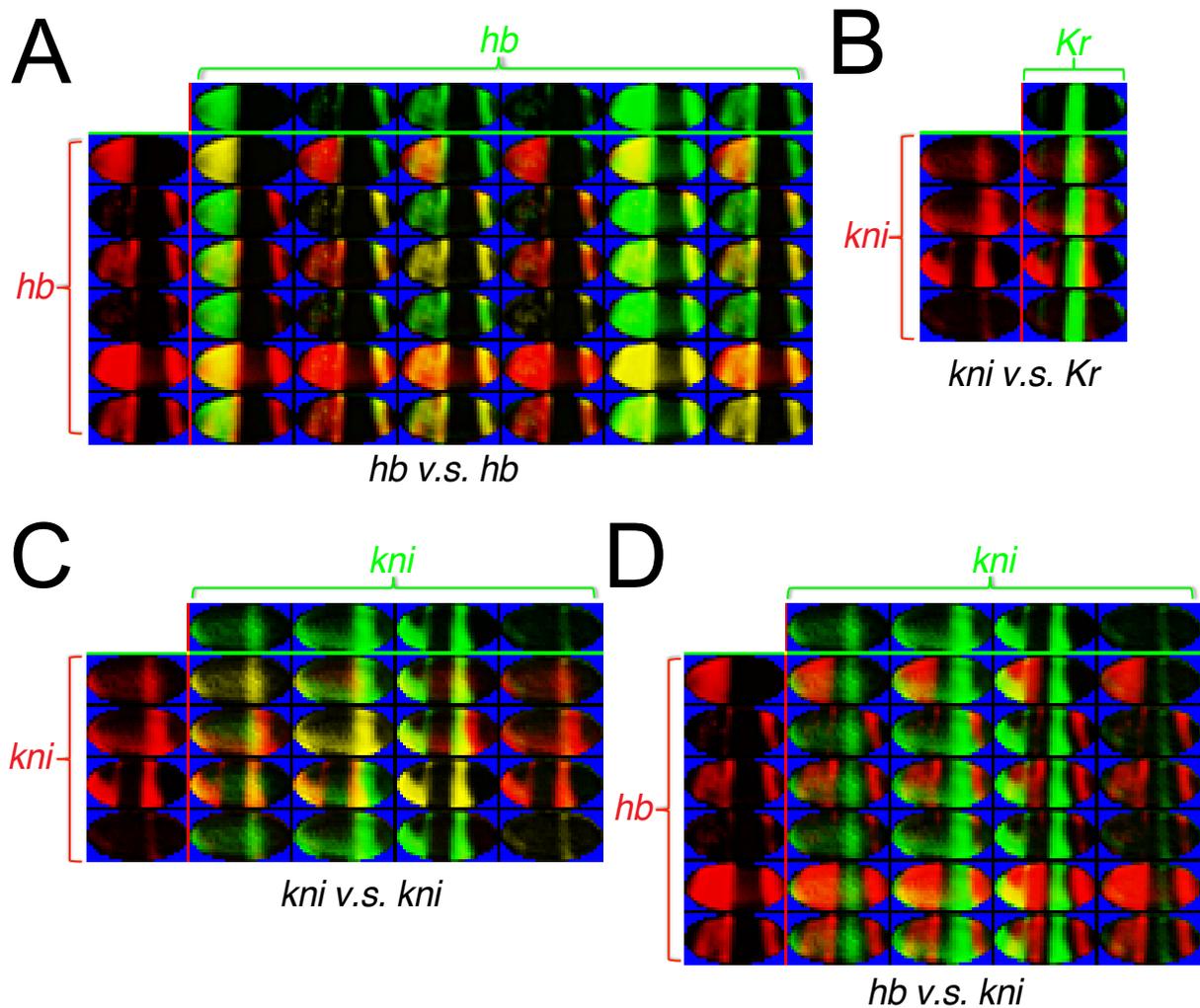
## 2 List of Datasets

The following datasets are available as *SI Datasets*:

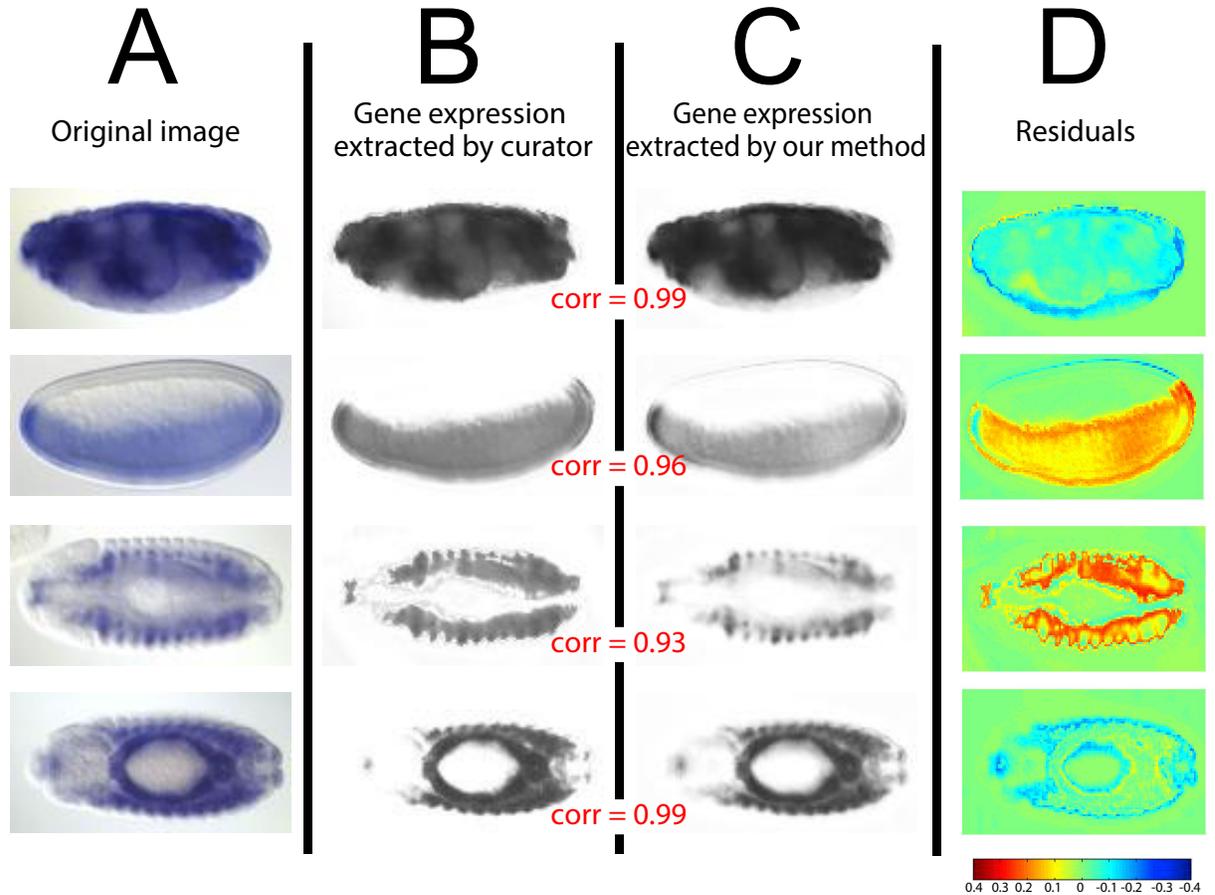
1. *Dataset S1*: the 1640 preprocessed expression patterns, corresponding to 701 unique genes. The gene symbols associated with the patterns are in the first row of the dataset. Each gene expression pattern is stored as a numeric column vector with 405 entries.
2. *Dataset S2*: the manual annotation data for the 1640 expression patterns. Each column contains the 11 annotation terms for one expression pattern. The 1640 patterns are arranged in the same order as in *Dataset S1*. The names of the 11 annotation terms are in the first column of the file. For each expression pattern and annotation term, we indicated by “1” if the pattern was labeled by the annotation term and “0” otherwise.
3. *Dataset S3*: the 21 principal patterns (PP) learned using staNMF. The first row of the dataset are names of the PP. Each PP is stored as a numeric column vector with 405 entries.
4. *Dataset S4*: the sPP coefficients for the 1640 expression patterns. The first row of the dataset are the gene symbols. Each column stores the 21 sPP coefficients and the intercept term.
5. *Dataset S5*: the sPP coefficients for the 701 genes. The first row of the dataset stores the gene symbols. As described in *SI Appendix Methods*, we defined the  $k$ -th sPP coefficient of a gene to be the maximum  $k$ -th sPP coefficient among all the replicate patterns of the same gene.
6. *Dataset S6*: this file indicates which expression patterns correspond to transcription factors (TF). The first row of the dataset stores the gene symbols. TF are indicated as “1” and the remaining genes are as “0”.

The above datasets are also available at <http://insitu.fruitfly.org/downloads>.

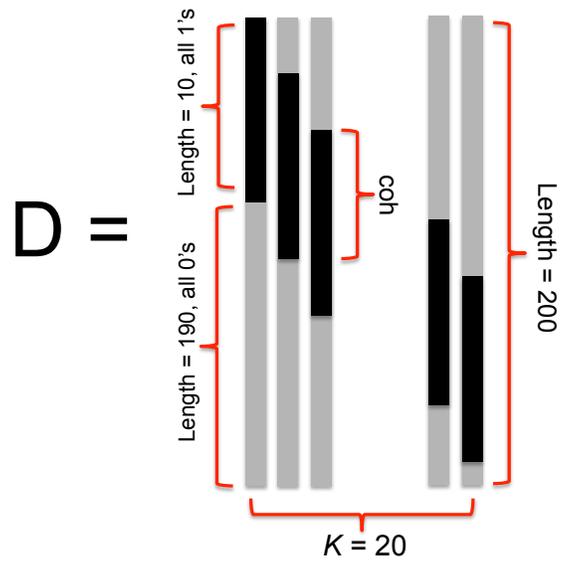
### 3 Supporting Figures and Tables



**Fig. S1:** Visual evaluation of embryo registration. By overlaying two gene expression patterns with different colors, in this case, red and green, virtual double staining was performed between the replicates of the same gene (i.e. (A) *hb* v.s. *hb* and (C) *kni* v.s. *kni*), and between the replicates of genes one of which is known to be repressor of the other (i.e. (B) *kni* v.s. *Kr* and (D) *hb* v.s. *kni*). In both cases, the boundaries of the genes match, indicating that our registration approach performed reasonably well in transforming a *Drosophila* embryo into a common frame of reference.



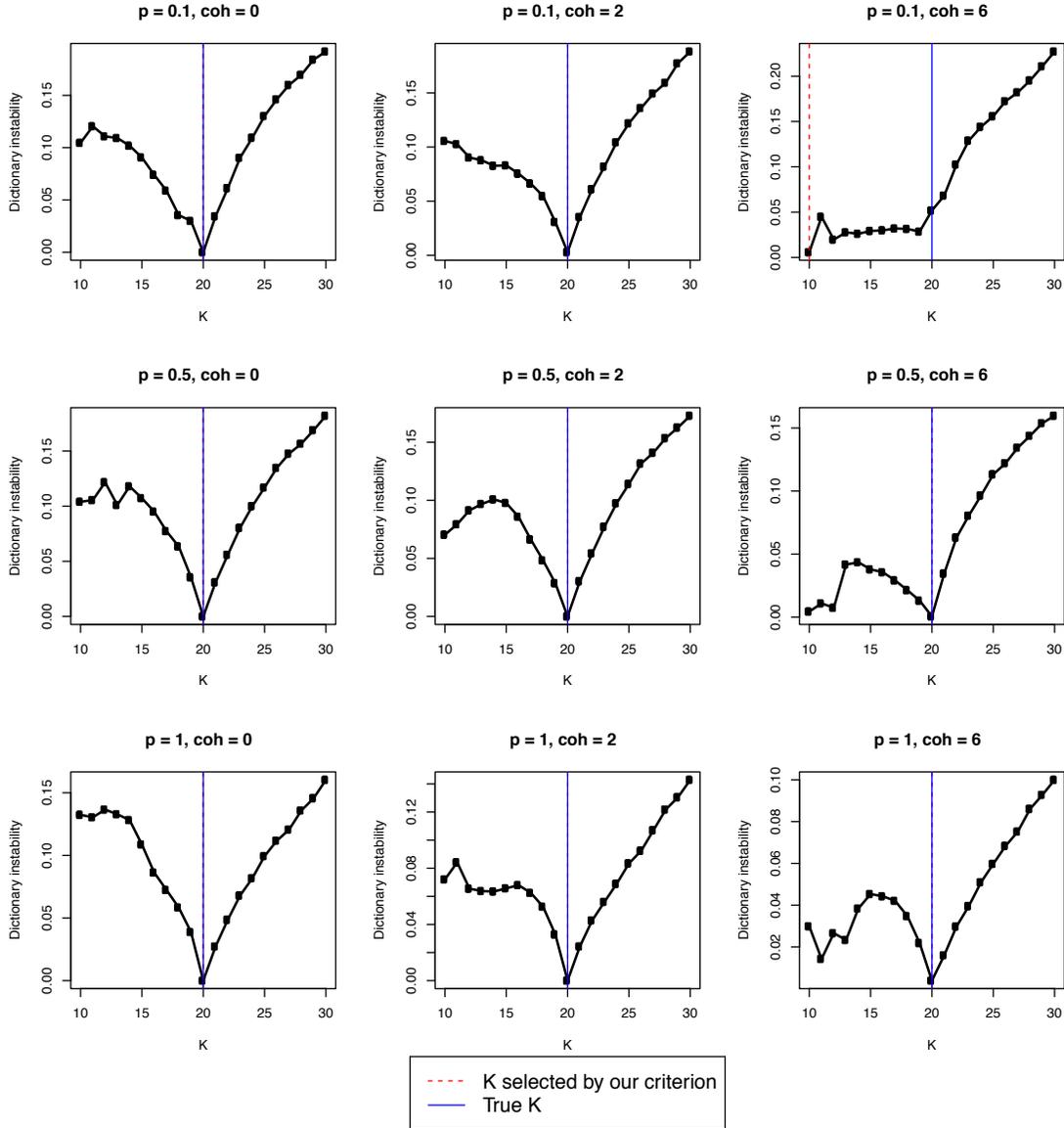
**Fig. S2:** Extracting gene expression patterns from images obtained through differential interference contrast (DIC) microscopy. (A) The original image was standardized to an image of  $64 \times 128$  pixels. (B) A curator used the selection tool in Adobe<sup>®</sup> Photoshop<sup>®</sup> to extract regions of the *Drosophila* embryo deemed as having the blue dye. Since we averaged the three color channels to yield a proxy for gene expression intensity, this resulting image is displayed in gray scale, with white being the region of low expression, and black the region of high expression. (C) We extracted gene expression using a linear combination of the RGB features from the original image (see *SI Appendix Methods*). For each example, the correlation indicated in red is the correlation between the gene expression extracted by curator and that extracted by our method. (D) The difference between the predicted pattern and the pattern extracted by the human curator is shown in the residual plot.



**Fig. S3:** Construction of the dictionary in Simulation Experiment 1. In the above illustration, each vertical bar represents a column of the dictionary. The black region of a bar indicates the entries that are ones and gray region the entries that are zeros. The parameter  $\text{coh} \in \{0, 1, \dots, 10\}$  is the number of common entries that are ones between two consecutive dictionary columns. It measures also the coherence of the dictionary. From the  $i$ -th bar to the  $(i + 1)$ -th bar, the black region is shifted down by the constant amount  $10 - \text{coh}$ .

# Simulation Experiment 1

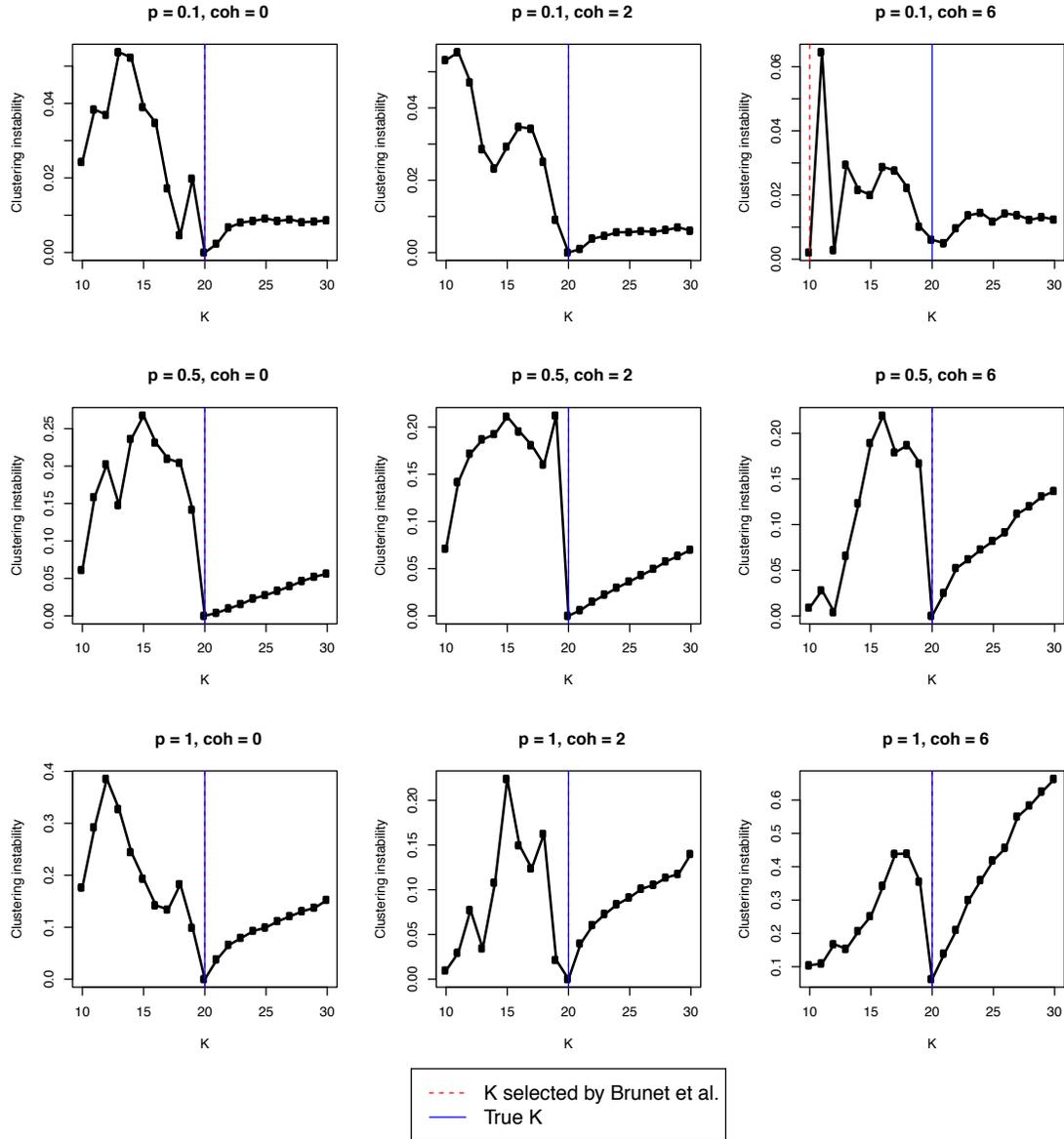
## staNMF



**Fig. S4:** NMF model selection using staNMF: Simulation Experiment 1. For each  $(p, \text{coh}) \in \{0.1, 0.5, 1\} \times \{0, 2, 6\}$ , we generated the data matrix  $\mathbf{X}$  from the model described in Simulation Experiment 1. For each parameter configuration, we ran NMF  $B = 100$  times for every  $10 \leq K \leq 30$  and then applied our stability criterion. For each plot, the vertical axis represents the dictionary instability as measured by  $\Upsilon(K)$  defined in *SI Appendix Methods*. The lower the value, the more stable the dictionaries with respect to random initial values.

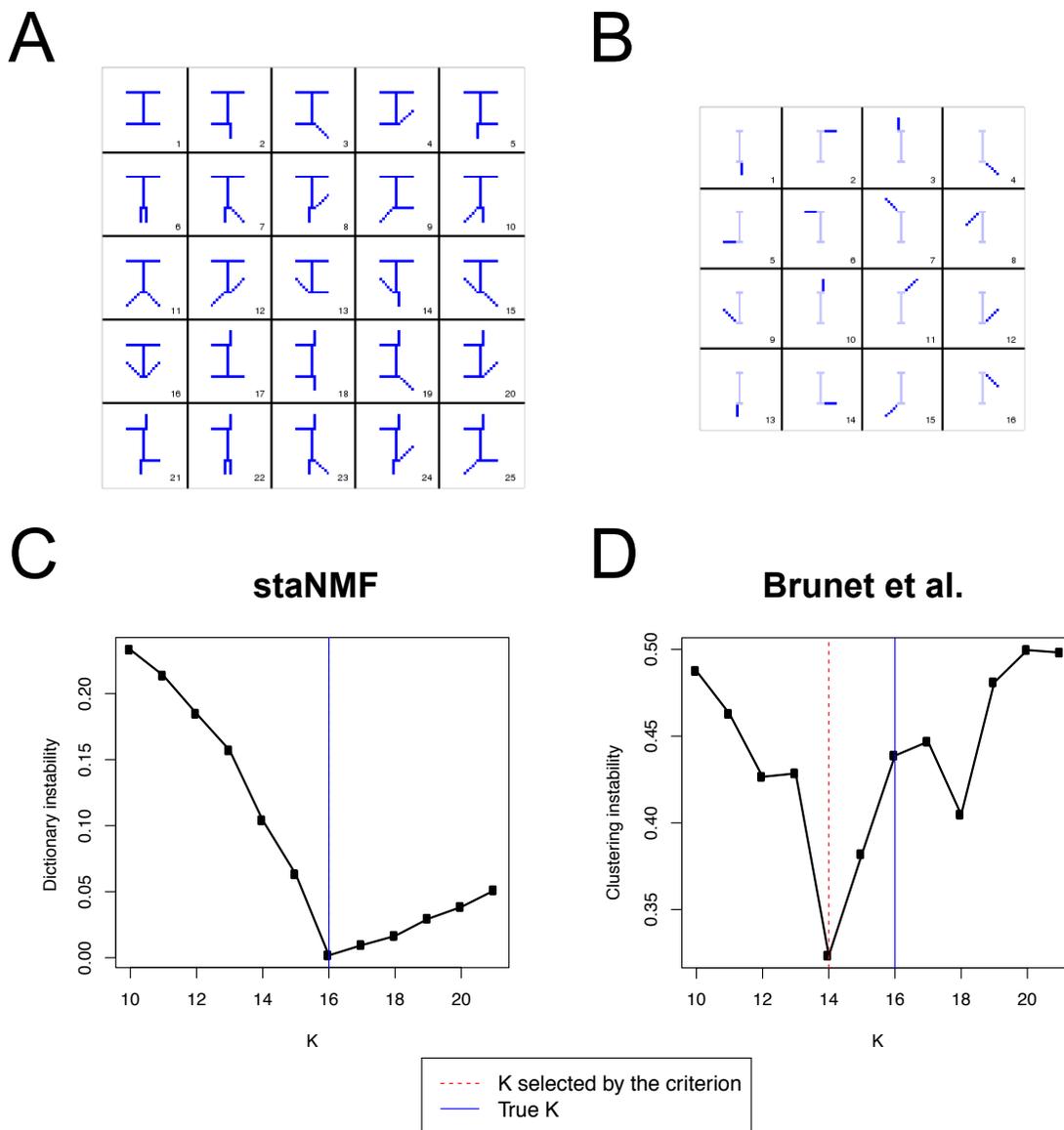
# Simulation Experiment 1

## Brunet et al.

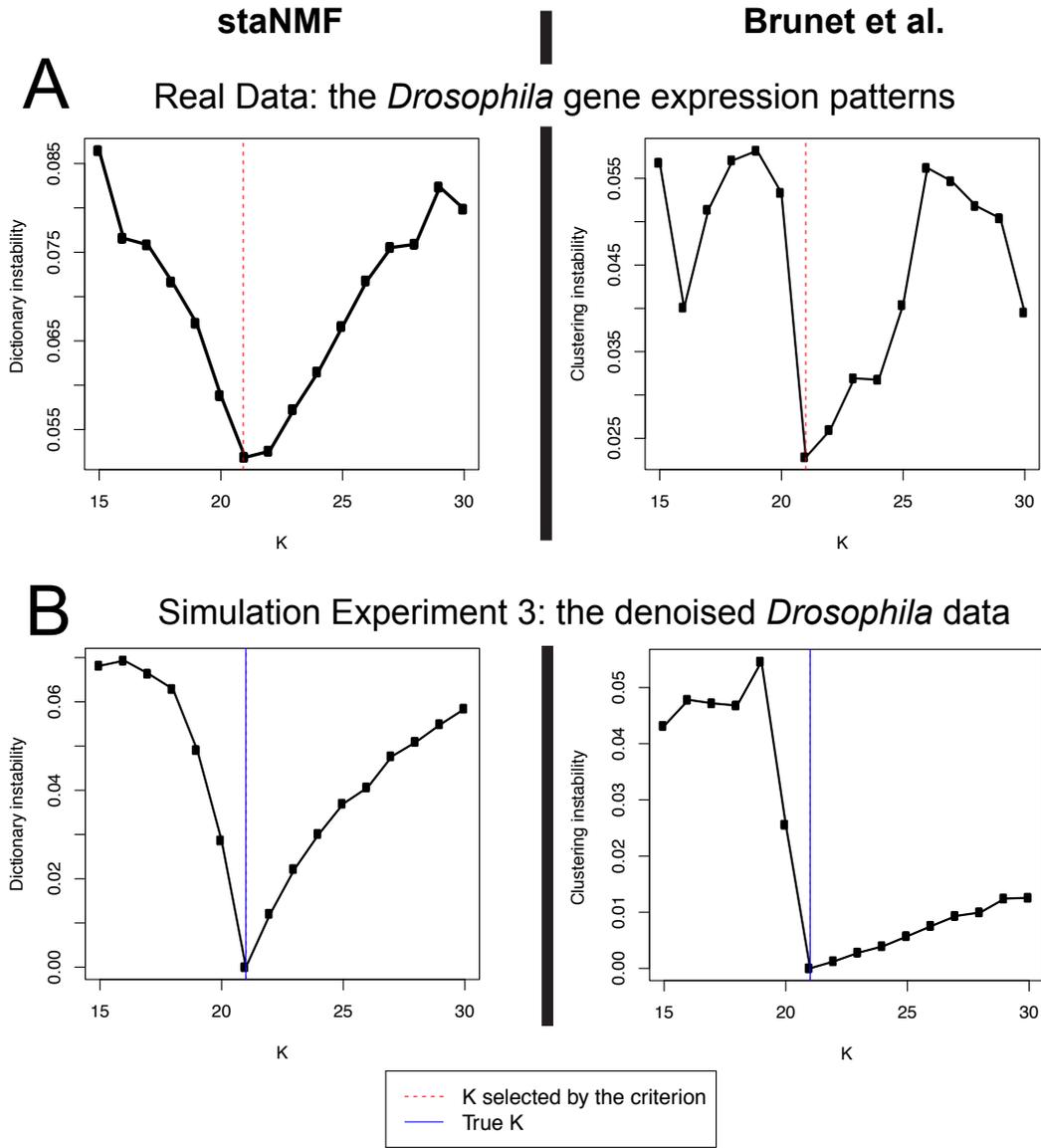


**Fig. S5:** NMF model selection using Brunet et al.’s clustering instability criterion [1]: Simulation Experiment 1. The synthetic data are exactly the same as those in Fig S4. For each plot, the vertical axis represents the clustering instability as measured by one minus the cophenetic correlation coefficient of the NMF cluster consensus matrix (*SI Appendix* Methods). The lower the value, the more stable the cluster assignment with respect to random initial values.

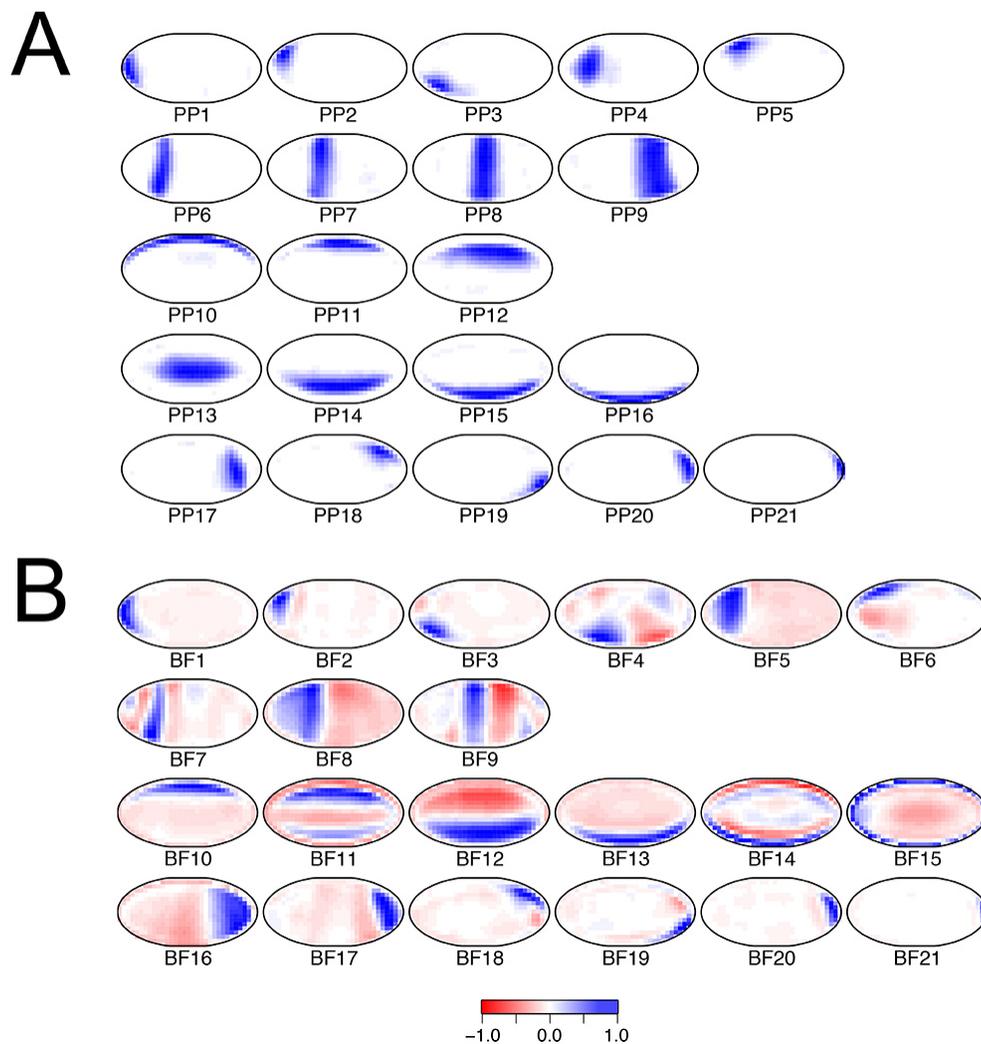
## Simulation Experiment 2: the Swimmer Data



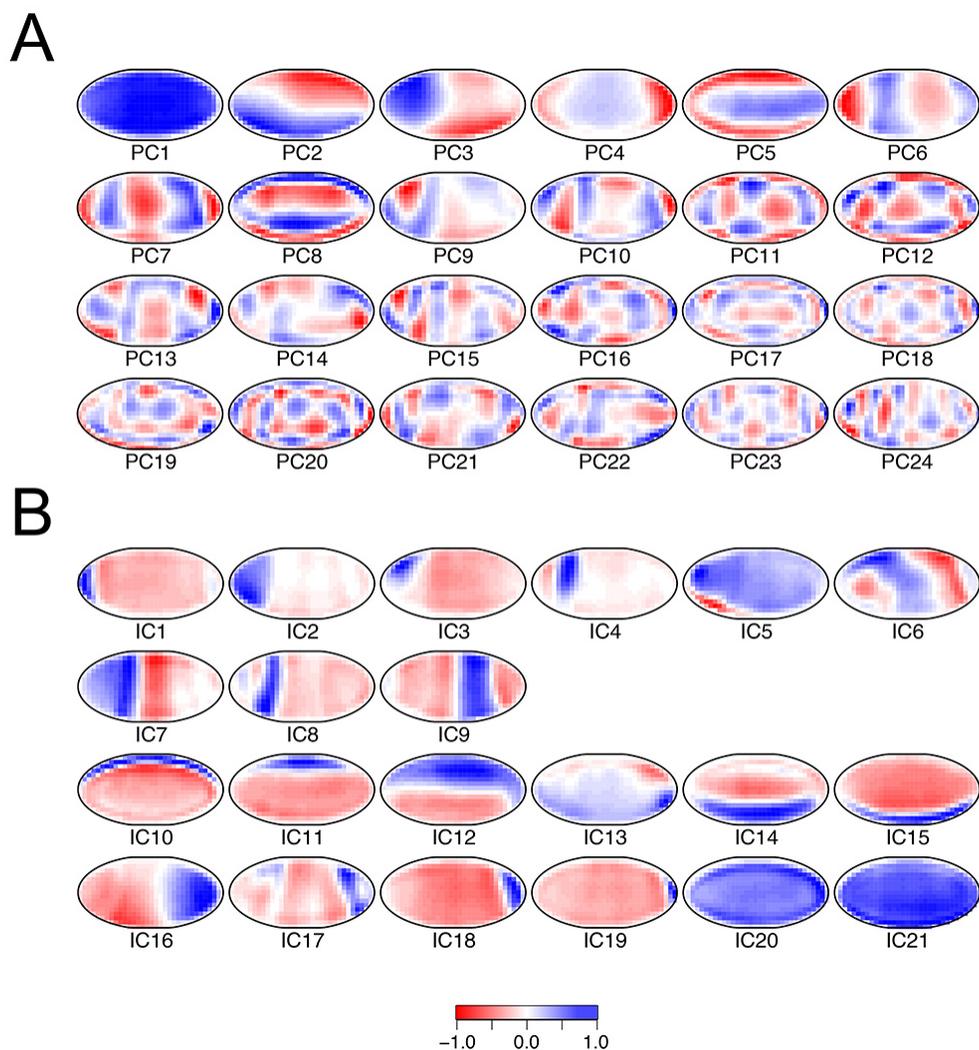
**Fig. S6:** NMF model selection: Simulation Experiment 2 – the Swimmer dataset [17]. **(A)** A sample of 25 images containing the artificial swimmers. **(B)** The 16 dictionary columns recovered by NMF. The dark blue region of each basis image corresponds to a limb of the artificial swimmer, whereas the light blue region indicates the torso of the swimmer. **(C)** StaNMF identified correctly  $K = 16$ . **(D)** Brunet et al.’s method selected  $K = 14$ .



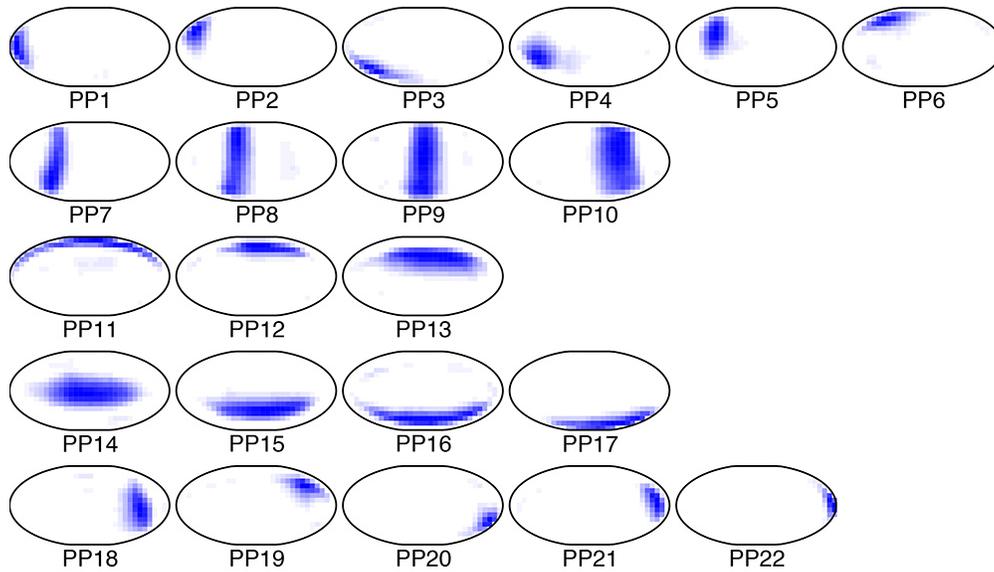
**Fig. S7:** Using StaNMF and Brunet et al.’s stability criterion on (A) the *Drosophila* spatial gene expression data and (B) the corresponding denoised data. The two methods agreed on  $K = 21$  in both examples. Note that for the real data we do not know the true number of PP and so only the red dash lines were drawn for the two plots in (A). The denoised data was constructed as  $\hat{\mathbf{X}} = \hat{\mathbf{D}}\hat{\mathbf{A}}$ , where  $(\hat{\mathbf{D}}, \hat{\mathbf{A}})$  is the output dictionary and the nonnegative coefficient matrix from the NMF. See *SI Appendix Methods* for more details.



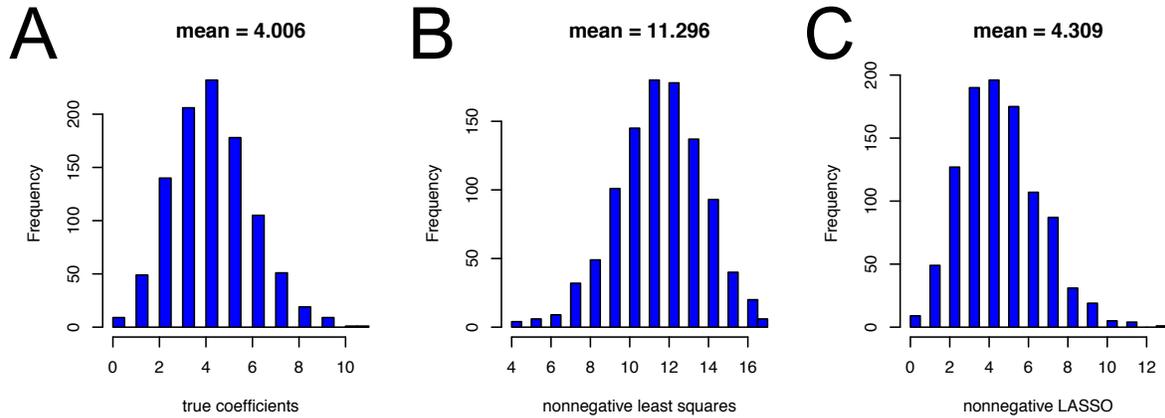
**Fig. S8:** A comparison between the 21 principal patterns (PP) and the 21 sparse Bayesian Factors (BF) [29]. **(A)** The 21 learned PP. This is the dictionary that achieved the minimum mean square loss in the NMF optimization among the 100 runs of NMF with  $K = 21$ . Every PP was normalized to have maximum intensity equal to 1. **(B)** The 21 learned BF. The BF are arranged in a way such that they correspond to the layout of PP in (A). Blue intensity indicates positive value and red indicates negative values. Every BF was normalized to have maximum absolute intensity equal to 1. For some of the BF, e.g. BF1, 2, 3, 10, 13, 20 and 21, the intensity in the negative region is rather uniform and the BF can be associated with the corresponding PP, e.g. PP1, 2, 3, 11, 16, 20 and 21. However, the biological meaning of the remaining BF is not immediately clear. For example, BF4 splits the mesoderm region into a positive half and a negative half, with other positive and negative regions scattering around the embryo. BF9 seems to be made up from the positive PP8 and the negative PP9. By allowing negative values in the sparse linear coefficients, some of the BF also appeared to be much broader than the PP, e.g. BF5, 8 and 16. We used the MATLAB code from [29] to obtain the BF. Since their program involved MCMC computation, we did not perform model selection. Instead, we simply set the number of BF to be  $K = 21$  as in NMF for direct comparison.



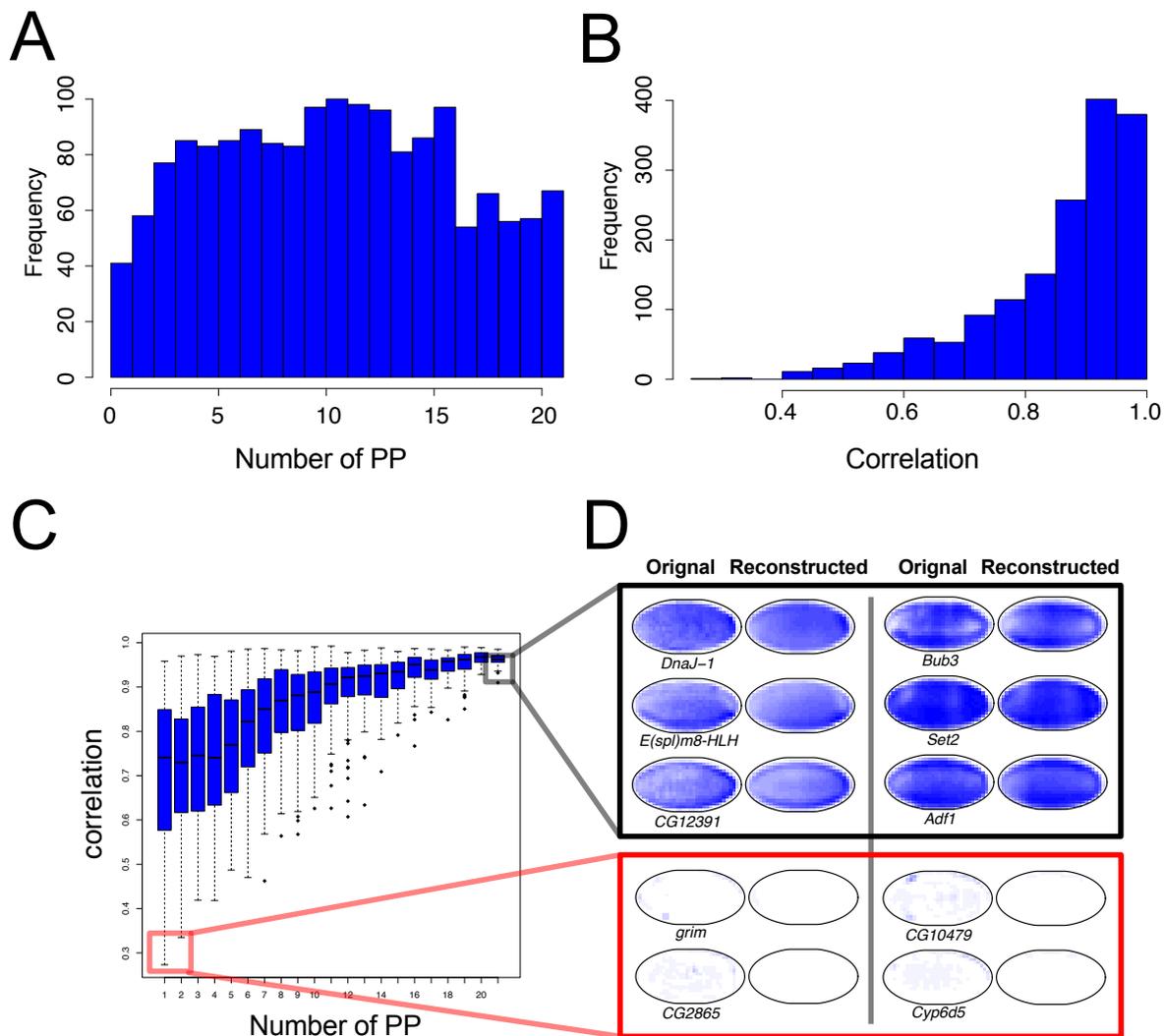
**Fig. S9:** Principal component analysis (PCA) and independent component analysis (ICA) for the *Drosophila* gene expression data. The components are not as biologically meaningful as the PP. **(A)** The top 24 principal components as ranked by the corresponding eigenvalues. As a consequence of the orthogonality of the PC, the derived components show oscillating patterns and are extremely difficult to interpret. **(B)** The 21 independent components. Similar to BF, the negative components in the IC make them less interpretable. For the computation of ICA, we used the R package `FastICA` [30].



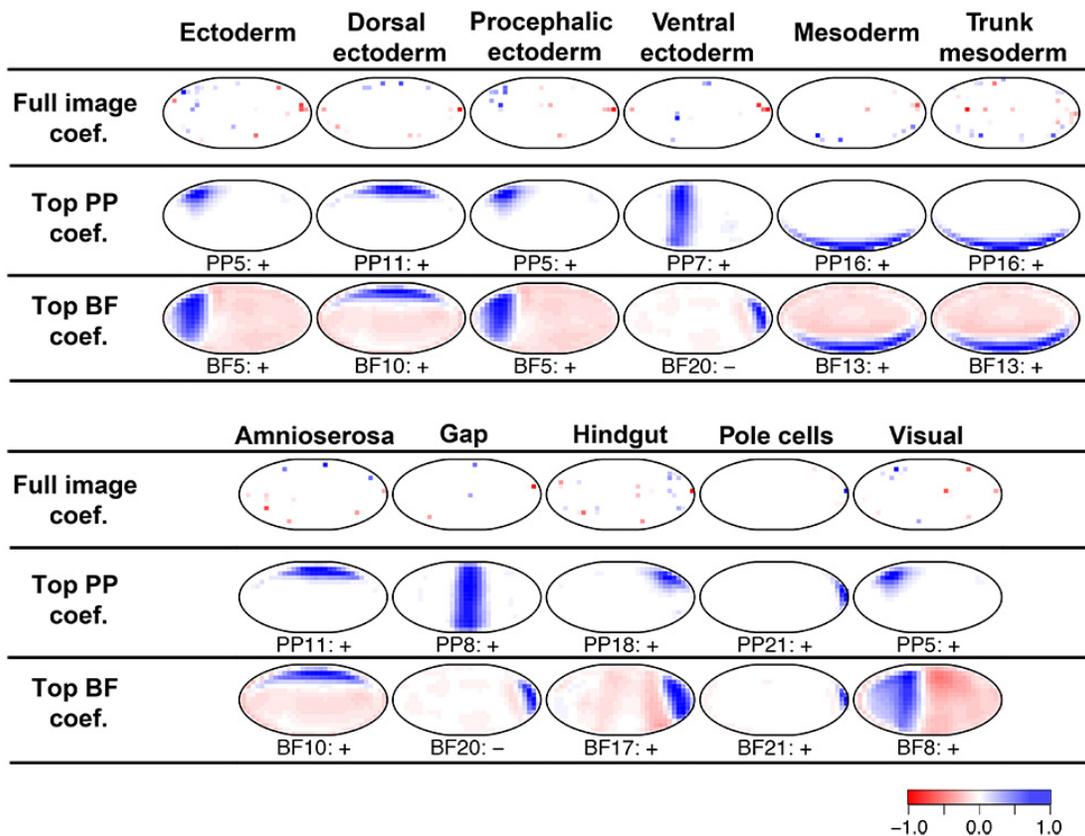
**Fig. S10:** The PP learned with parameter  $K = 22$  instead of the optimal  $K = 21$ . This is the dictionary that achieved the minimum mean square loss in the NMF optimization among the 100 runs of NMF with  $K = 22$ . The learned PP are similar to those with  $K = 21$ . The only major differences from the  $K = 21$  dictionary are PP4, PP5 and PP6 in the  $K = 22$  dictionary. PP4 for  $K = 22$  appears to be smaller than the PP4 in the  $K = 21$  dictionary. The PP5 in the  $K = 21$  dictionary splits into PP5 and PP6 in the  $K = 22$  dictionary. Other PP are highly similar for both cases.



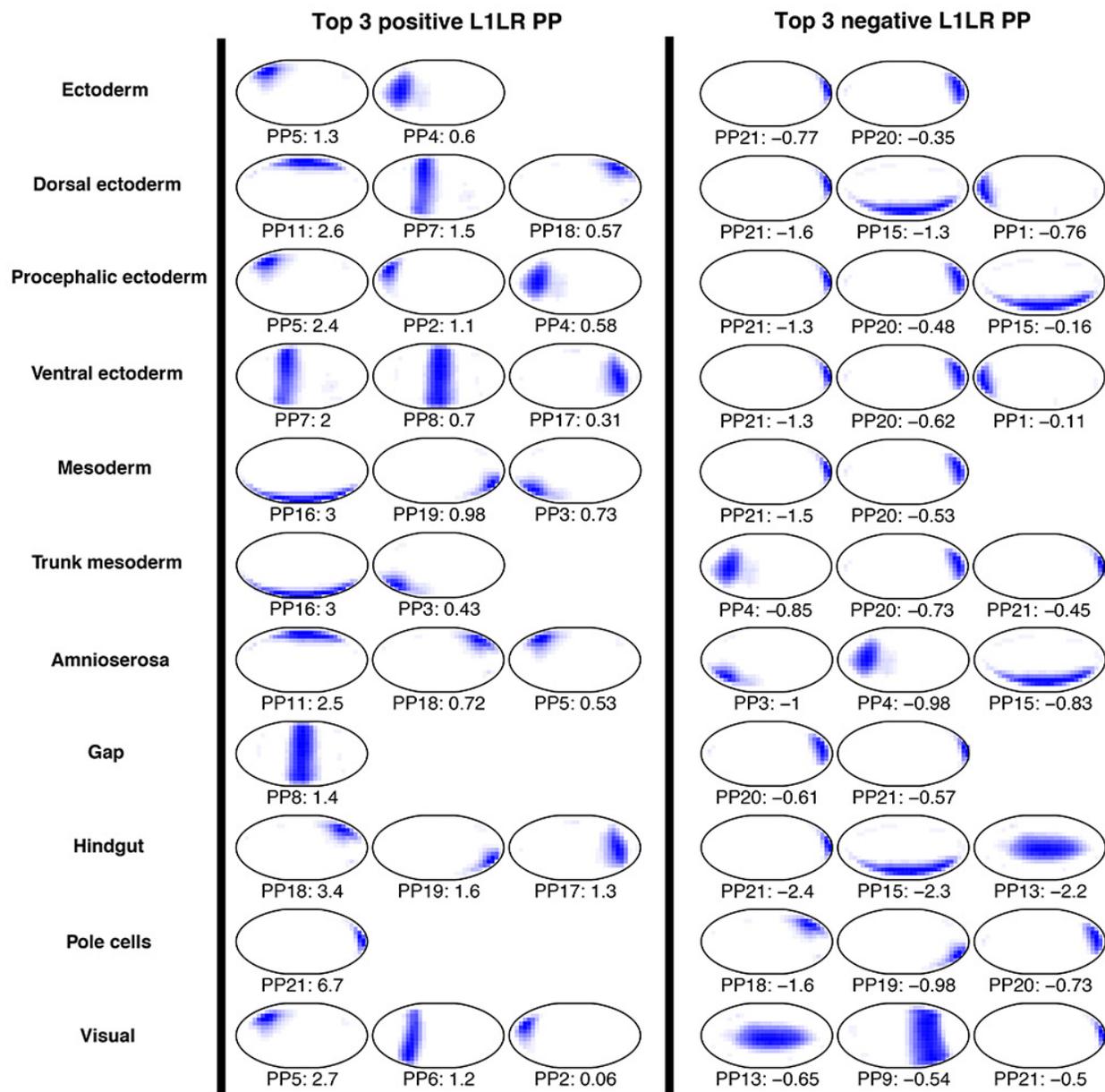
**Fig. S11:** Effectiveness of the LASSO+NLS model selection and fitting procedure. We generated 1000 data vectors according to the model described in Simulation Experiment 1, with dictionary parameter  $\text{coh} = 2$  and random linear coefficient sparsity  $p = 0.2$ . For each data vector, using the ground truth dictionary as the covariate matrix, both nonnegative least squares (NLS) and LASSO+NLS were applied to estimate the linear coefficients. Shown are the histograms of the number of nonzeros in the linear coefficients of (A) the true model, (B) the NLS estimates and (C) the LASSO+NLS estimates. If model selection is performed properly, the resulting distribution of number of nonzeros should match with that of the true coefficients. Here, we can see that the distribution of the number of nonzeros for NLS shifted significantly to the right (B), indicating that NLS tended to overselect covariates. The LASSO+NLS fitting procedure, on the other hand, produced number of nonzeros distribution almost identical to histogram for the true coefficients (C).



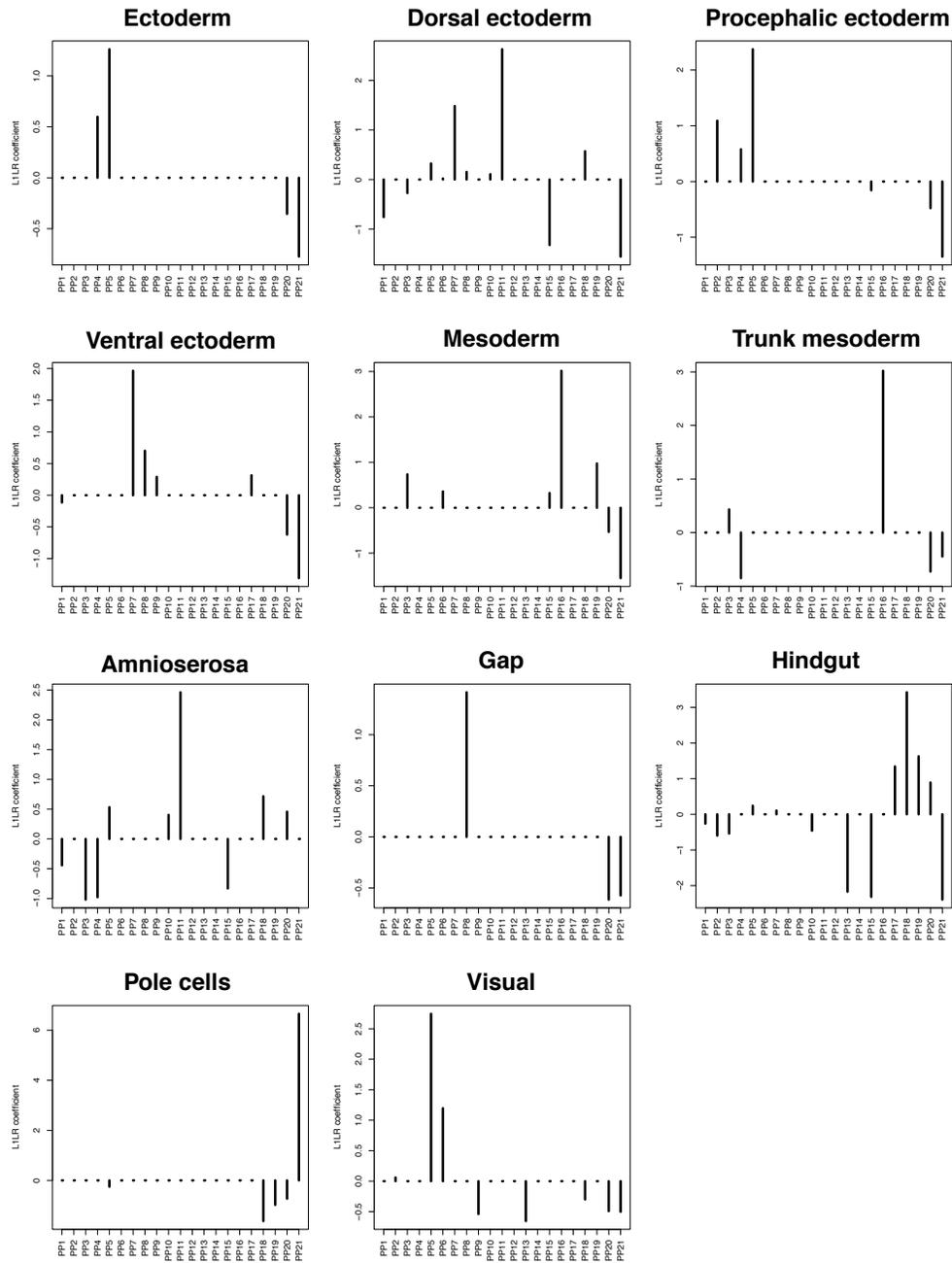
**Fig. S12:** Spatial gene expression reconstruction quality of the sparse PP (sPP) representation. (A) Histogram of the number of selected PP per expression pattern. (B) Histogram of the correlation between a gene expression pattern and the reconstructed pattern (linear combination of the 21 learned PP using the sPP representation as coefficients). (C) The relationship between the number of selected PP and the correlation. (D) A sample of expression patterns represented with more than 20 PP (black box) and a sample of expression patterns with poor reconstruction quality, i.e. correlation less than 0.35 (red box).



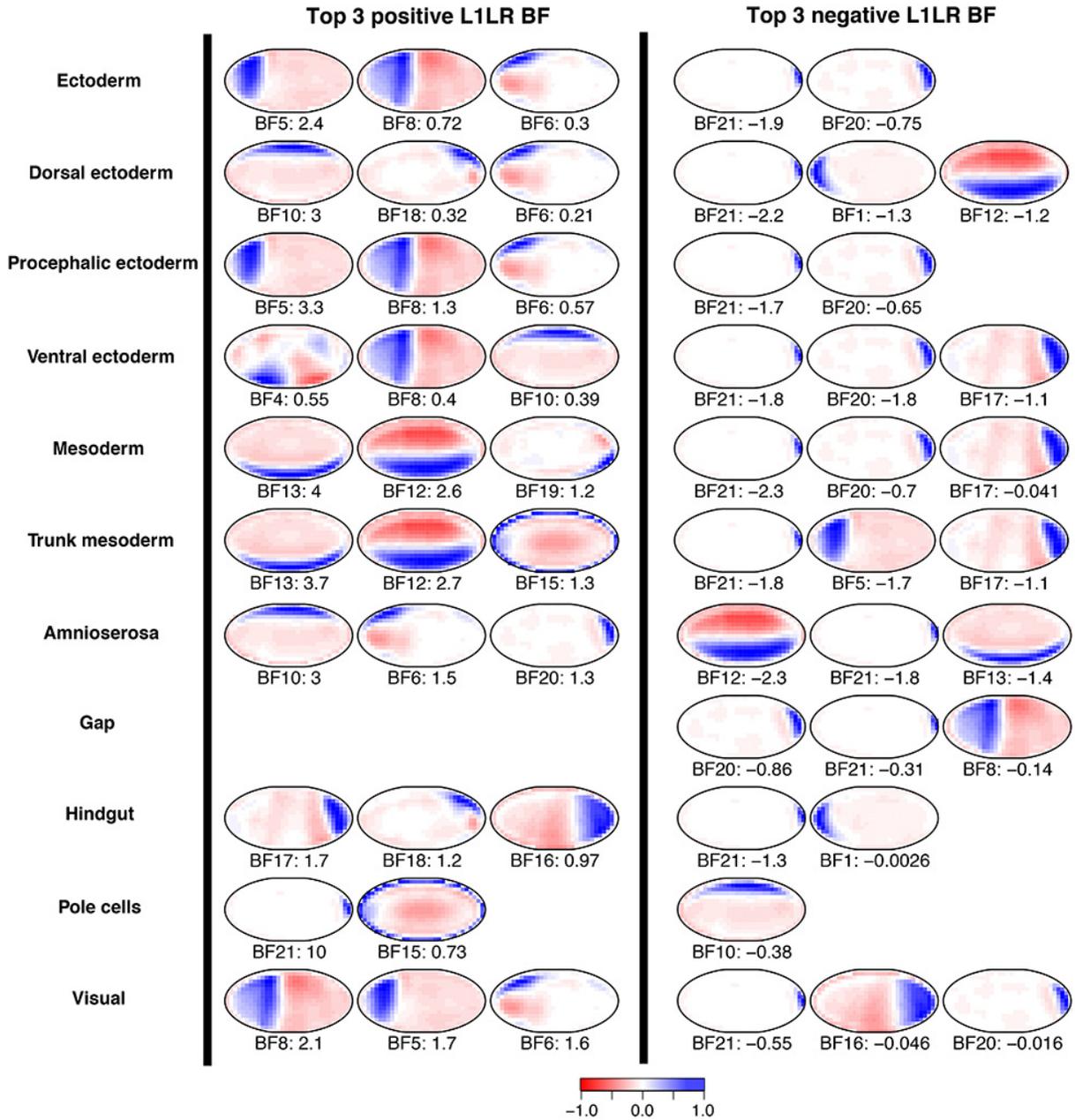
**Fig. S13:** Interpretability of the L1LR under the pixel-based, the sPP and the BF representations. Shown is the complete version of Fig. 3C. The pixel-based full image representation: all L1LR coefficients are shown as pixel values within the embryo; the sPP-based and the BF-based representations: only the top L1LR PP or BF that corresponds to the largest L1LR coefficient is shown. The scale goes from -1 to 1 and is color coded respectively from red to blue. For the PP and BF, “+” indicates that the largest L1LR coefficient is positive, and “-” indicates that the largest L1LR coefficient is negative. For some annotation terms, the positive predictors for the pixel based model, the top L1LR PP and the positive part of the top L1LR BF overlapped with the regions in the embryo described by the controlled vocabulary terms. However, only the top PP consistently showed the annotation term exactly as a curator would annotate the gene expression. For example, for the annotation terms such as “dorsal ectoderm”, “mesoderm”, “trunk mesoderm” and “pole cells”, the top PP corresponds to the areas of the embryo that can be easily recognized as those anlagen. On the other hand, the selected predictors for the pixel-based representation are predominantly isolated pixels at locations associated with the specific annotation term. Of the 11 annotation terms, all of the top L1LR PP but only nine of top L1LR BF-based components have positive fitted L1LR coefficients. For some of the nine terms with positive association for both PP and BF (e.g. “dorsal ectoderm”, “mesoderm” and “pole cells”), the top L1LR PP and the positive part of the top L1LR BF have similar shapes and sizes. For other terms such as “hindgut” and “visual”, the positive part of the top L1LR BF pattern appears to be much broader than the top L1LR PP. We refer readers to Fig. S14 and S16 for more PP and BF with high L1LR coefficients.



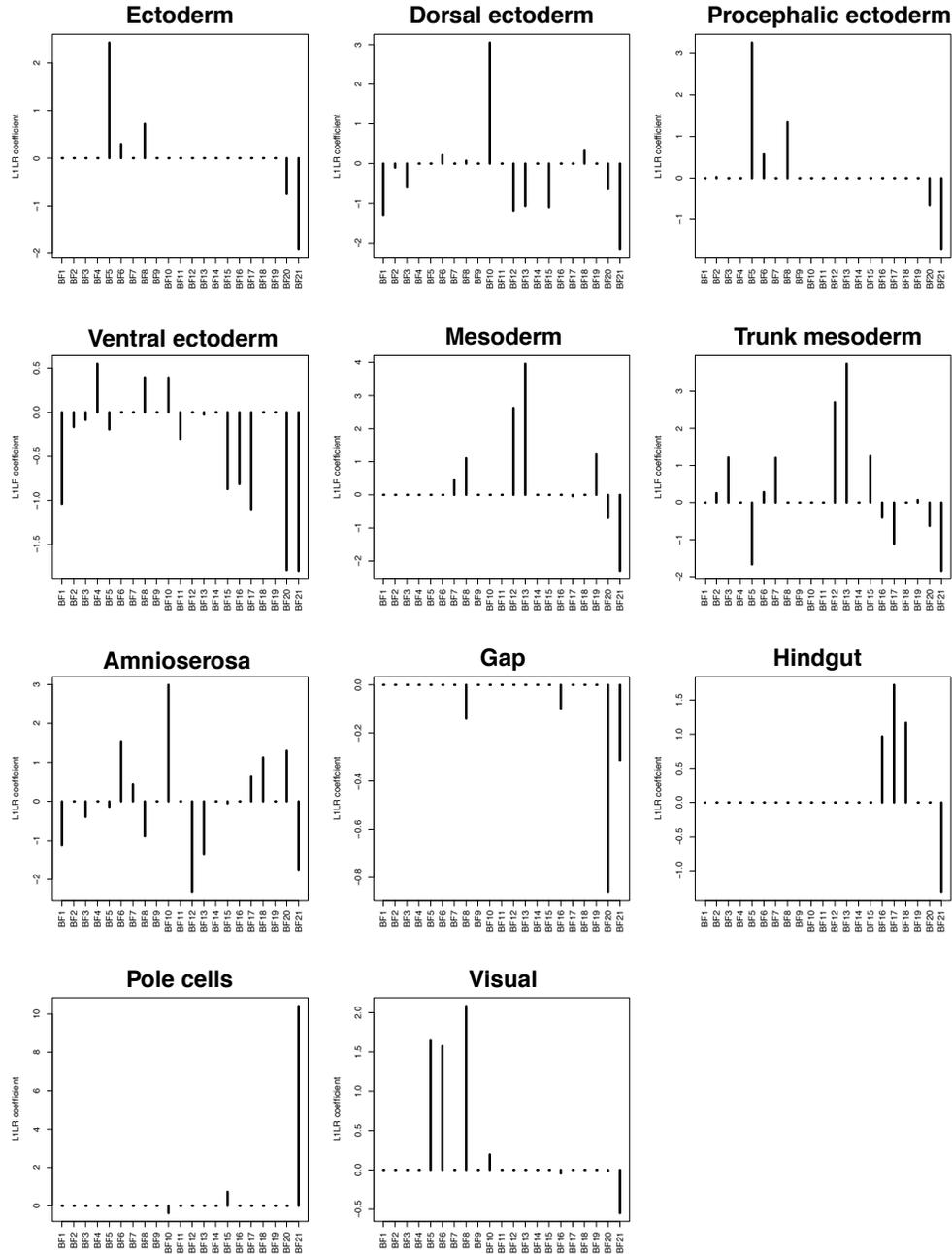
**Fig. S14:** Top L1LR PP for all 11 annotation terms. For each term, the PP with the three most positive and three most negative L1LR coefficients are shown. The PP name and the corresponding L1LR coefficient are under each PP. For each annotation term, the top positive L1LR PP are likely to be PP that are directly adjacent to each other. For example, for the term “hindgut”, the three top PP, PP17–19, correspond to the three adjacent parts of the hindgut pre-organ system. As for the three most negative L1LR PP, we note that the pole cells pattern PP21 show up in nine out of the 11 terms. This indicates that if a gene is expressed in the pole cells, it is unlikely that it will be annotated with any other terms. In fact, a large number of genes exhibit their strongest expression in pole cells and have limited expression in other PP (main text Section “PP associated gene functions and relationships between spatial regions”). See Fig. S15 for the full PP L1LR coefficients.



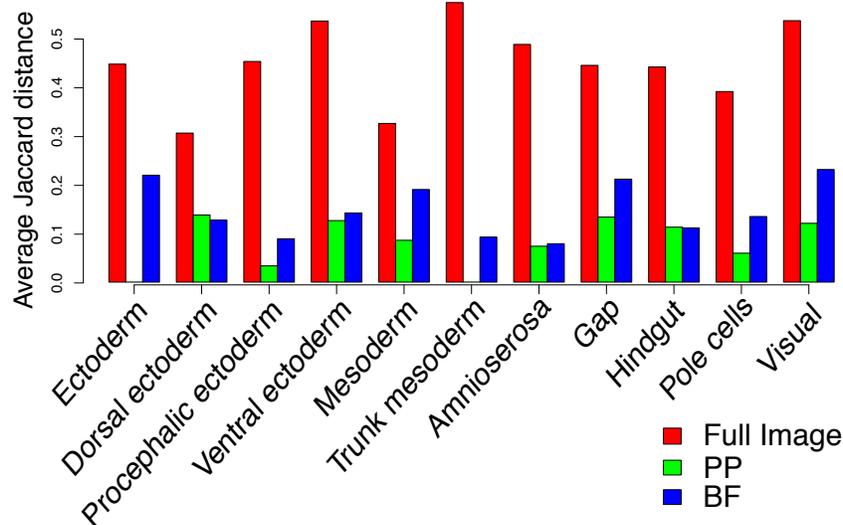
**Fig. S15:** The full L1LR coefficients for annotation prediction using the sPP representation. For all 11 terms, the L1LR coefficients are sparse and the largest L1LR coefficients are always positive. Furthermore, the top L1LR coefficient is much larger than the second largest L1LR coefficient in magnitude. These facts indicate that the top L1LR PP is the dominating factor in determining whether a gene expression pattern is labeled with an annotation term. This provides evidence why we can consider only the top L1LR PP when associating PP with annotation terms.



**Fig. S16:** Top L1LR BF for all 11 annotation terms. For each term, the BF with the three most positive and three most negative L1LR coefficients are shown. The BF name and the corresponding L1LR coefficient are under each BF. Similar to the PP case, the top positive BF are usually the BF with adjacent positive regions, but the negative regions make them harder to interpret. For the term “gap”, all L1LR coefficients are non-positive, indicating that no BF is positively associated with the term. The three most negative L1LR BF are rather consistent with the L1LR PP, with BF21, the BF version of pole cells, appearing frequently in the top list. The scale goes from -1 to 1 and is color coded respectively from red to blue. See Fig. S17 for the full BF L1LR coefficients.



**Fig. S17:** The full L1LR coefficients for annotation prediction using the BF representation. The L1LR coefficients are sparse. But unlike the PP case in Fig. S15, the largest L1LR coefficients are not always positive (“ventral ectoderm” and “gap”). In addition, some of the largest L1LR coefficients are much closer to the second largest L1LR coefficients than in the PP case: e.g. the terms “mesoderm” and “visual”. This is because the positive regions of some BF have a significant amount of overlap, e.g. BF12 versus BF13, and BF5 versus BF6. The overlap between PP is much smaller.



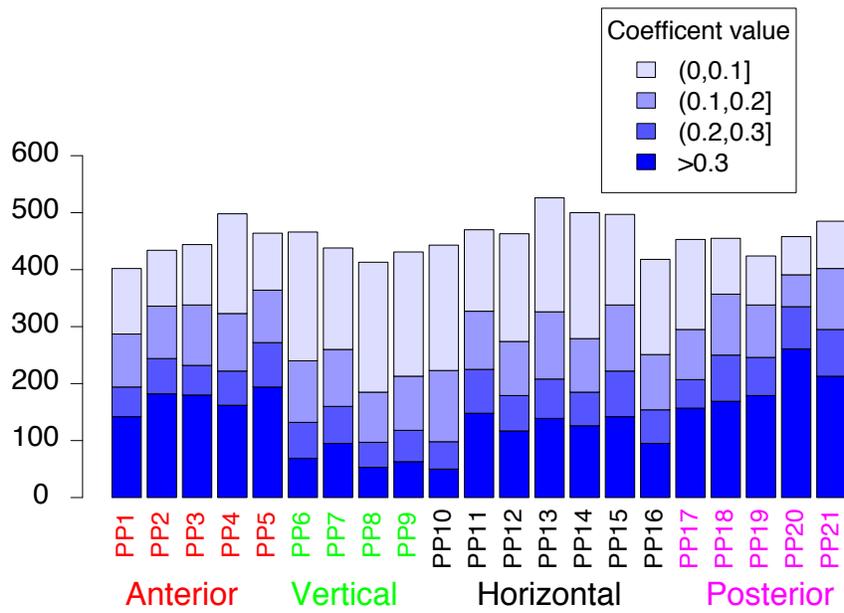
**Fig. S18:** Stability analysis of the set of selected L1LR predictors for three representations. As described in the text, for each annotation term and representation, the  $l_1$ -penalization parameter was chosen such that it was the largest among all parameters whose cross-validation AUC was within one standard error of the maximum AUC. For this plot, instability is measured by the Jaccard distance between the supports of two L1LR coefficients, averaged over all 45 coefficient pairs in the 10-fold cross-validation. The higher the Jaccard distance, the more unstable the support of the L1LR coefficients. The plot indicates that the selected L1LR model for the sPP representation is most stable among the three representations, except for two terms: “dorsal ectoderm” and “hindgut”, for which the BF approach is slightly better. The pixel-based approach selects highly unstable predictor sets.

PP category	Gene symbols
PP1	<i>croc ImpE2 kni CG9514 fd102C knr1 hkb twi toc CG1663 CG8289 Mdr49 fd19B CG14427 pnt Pka-C3 sc Adgf-A rdx Ndae1 Sep5 Cad74A Btk29A CycB h Alh dm CG2915 CG12581 ird5 CG31909 CG12420 retn CG16815 drm Mkp3 Mitf rib CG11652 prb Zasp52 rad50 Xpc Akap200 CG10365 Hsp83 Optix Blimp-1 Acn noc roX1 dpm oc hbn bai chif milt egg nerfin-1 lrp4 RnrS Caf1 CG2941 CG11208 ry Neu2 Set2 Moe CG2469 Aatf sala wg Best2 fu2 srp CG8788 wisp CG8388 CG31038 term Prx5 ftz CG14814 CG12702 pros GH06606 CG33116 ind Sec61beta Tao Ocho CG5888 Mink BicD Asph RhoGEF3 velo 18w dap Pcl</i>
PP2	<i>hbn danr Blimp-1 nuf path pcs Bili Mitf dan Ppa CG3097 Nek2 seq h CG8289 fd19B CG16815 aos Alh pdm2 CG12420 CG1663 CG14427 apt CG2915 retn p115 zen Six4 fj CG11652 CG5059 CG14711 gukh bbg Hsp83 Akap200 numb Optix CG31909 CG5888 oc Adgf-A Ocho CG42342 rad50 ASPP tld CG11208 exex RhoGEF3 dm rdx tkv Btk29A Sec61beta egr CG14805 Xpc Su(H) CG43394 dpm pnt Ndae1 jigr1 Acn Neu2 ird5 Asph pros Sep5 18w CG10479 Cad74A tll term LManII hb CG7800 Bsg25D Prx5 Elba2 noc Cyp6v1 wisp chif cnc bai Tao dnk Set2 CG13653 Caf1 CycB velo CD98hc CG10338 CG9514 Aatf CG11417</i>
PP3	<i>CG5346 Mdr49 CG5888 CG31909 kni sprt Ndae1 vnd knr1 CG14427 rdx 26-29-p Zasp52 ry egg edl CG11208 wisp slp2 Elba2 Cyp310a1 Mes2 Pcl NetA fd19B hb twi Mitf oc Pka-C3 E5 Mes4 sna dm neur CG12177 slp1 term srp numb ems htl Sdc gukh dan roX1 gcm CG10479 Bsg25D pros siz Ppa Xpc CycB Sep5 Coop stumps croc CaMKII Asph CG11652 Akap200 bou Cad74A milt CG8788 T-cp1 sog bowl CG2941 Lk6 CG33099 lrp4 ind Dg seq Sec61beta Hsp83 Poc1 noc danr lama Acn RhoU CG32982 CG5059 CHOp24 CG9514 chif zfh1 Bub3 CG6654 bai btd vls hrg Prx5 Tao CG2469 CG43394</i>
PP4	<i>numb SoxN D siz oc phyl hb CG14427 tld danr dan CG5346 fd19B Elba2 slp2 yellow-e3 pros Akap200 ems E5 Sec61beta Ppa jigr1 CG12581 Tao esg CG14814 Bsg25D CG14805 CG16815 Nek2 18w CG5888 fj Acn CG5059 lok term Miro Aatf p115 path CG13894 CG11652 slp1 chif tll CG42666 rdx Prx5 CG17786 CG12702 Blimp-1 bai sala toc Poc1 CG2915 Caf1 CG11208 dnk velo gt Vha68-2 Btk29A CG42516 Cad74A Ocho Su(H) ASPP Meltrin Set2 dbo Alh hrg Egm retn Xpc ftz Aldh-III salm milt dm vls CD98hc E2f1 janA ind sog neur seq Neu2 Asph Hsp83 Usp5 Cyp1 bwa CG14915 pnt ifc</i>
PP5	<i>Pepck Doc1 Doc2 lov CG7800 aos Ance Gbs-70E oc ill Dl Coop numb fd19B hb Sec61beta zen tkv otk slp1 toc slp2 chn CG5888 jigr1 wg CG5346 seq SoxN CG13894 egr dm CG17786 bai CG14427 dan grn MESR3 siz jing CG42516 rdx T-cp1 yellow-e3 Hsp83 Blimp-1 Nek2 Dll Elba2 18w tmod bmm danr shn CG11208 nuf Sema-5c E2f1 Su(H) Atet salm pros E5 CG12420 path CG13868 CG5532 bowl esg tok Ppa lok ind milt Xpc Gsc ems corto Prx5 link Akap200 p115 CG11652 brat sala CG4199 Srp68 CG3838 CG14814 tld chif vfl term nAChRalpha2 Cyp6v1 retn GH06606 Eip71CD CG14711 CG14805</i>
PP6	<i>btd Blimp-1 prd Dfd CG10479 kn CG13894 path slp1 fj bmm danr slp2 prb knr1 ems rdx toc hb tmod salm h Dl CG10979 CG30431 CG10176 Akap200 jbug chn dan Coop CaMKII dpm Ppa CG14814 CG43394 eve Pepck CG5888 18w D Nek2 CG2469 bai GH06606 chif tutt link rpr Aatf jigr1 nuf CG14805 siz Traf4 Xpc CG11696 Set2 Puf3 Prx5 Poc1 CG42666 yem kni dnk CG5059 CG11652 MESR3 Elba2 Cys CHOp24 gt Zasp52 Acn Vha68-2 Miro CG5346 E5 wisp Adf1 CG14657 term Aldh-III MFS14 Moe CG7800 CycT ZnT63C run htl Sema-5c Caf1 milt velo srp retn CG14915 CG42516 CG2915 p115</i>
PP7	<i>odd salm hb dan rpr CG5888 sala path danr lok 18w Sema-5c CG13894 CenB1A jigr1 bmm Cad74A prd prb bai sob CG14814 ftz fj corto Akap200 Ppa opa D lama CG8147 run Dfd Alh chif CG14657 MESR3 fz2 Dll CG11696 eve wg Cys gt yem Eip71CD Elba2 Btk29A CG11652 CG14805 Moe tex Set2 Nek2 jbug CG2915 CG12420 Poc1 Prx5 tmod CG10176 janA Aatf Coop CG7800 CHOp24 Miro velo rdx CG45186 exex CG5059 CG10979 Sep5 chn comm Cyt-b5 toc Meltrin CG5346 Aldh-III h Caf1 link siz pdm2 hrg Atg18a Mkp3 CG14915 DnaJ-1 wisp E5 milt CycT nuf CG10479 CG9505 sad aay</i>
PP8	<i>Kr tsh Blimp-1 lok Sema-5c Btk29A dpm cad prb D Glu44EF mtd E2f1 Ppa Alh CG1146 CG14814 CG8147 jigr1 Sodh-1 dbo lov Bsg25D Elba2 danr Eip71CD yem Cad74A opa prd Atg18a pdm2 sob corto chif Akap200 stwl gk CG14805 Ect4 mas ftz CG11696 CG7800 Moe nuf h GH06606 Set2 Cys jbug RnrS Poc1 vfl CD98hc fz2 MFS14 CG5888 CG5059 hb Miro Prx5 dan DnaJ-1 Acn run janA lama CG14657 Aatf sprt CG12702 Sep5 ara CG45186 tex Atg1 velo pros Pimo term Caf1 step Su(H) Aldh-III ird5 phm CG12581 Gbs-70E vls Mem3 Gdi gukh trx CaBP1 Fbel7 esg CHOp24 slp1 Adf1 nub pdm2 tsh gk Glu44EF kni Atg1 CG8147 D cad prb Btk29A danr Ppa h CG14814 opa Alh gt dpm</i>
PP9	<i>jigr1 exex Moe prd dbo nuf vfl GH06606 mas sc whd lov phm CG14657 Akap200 ftz Bsg25D chif CG14805 CaBP1 CG11696 esg fz2 ird5 corto Atg18a trx CG1146 CG11652 18w Prx5 run Gdi Poc1 janA toc DnaJ-1 CG11134 Acn Elba2 velo htl eve RhoBTB CG3838 CG10631 CG10176 pcs CG5059 CG8668 hrg CHOp24 CycT Fbel7 chn CG1910 Nek2 rad50 odd Imp bai nerfin-1 Caf1 E2f1 Set2 path CD98hc CG12391 milt step 26-29-p Miro Sdc CG7800 Mnt Cys D19A ude E(spl)m8-HLH CG14915</i>
PP10	<i>CG13653 ASPP link zen CG14427 bun shn Doc1 dpm egr esg toc lov CG12581 CG10479 pan Cen CG1146 ftz fz2 corto Caf1 D19A GH06606 CG14915 Dok Cys slmb CycB chn trx Doc2 CG16815 RapGAP1 net gcm2 Not10 Adf1 Eip71CD Aldh-III ush rho pcs Dl Set2 Bsg25D eroc inv Ect4 milt hth bel Prx5 zen2 Fbel7 CycA Sodh-1 Pcl Xpc 18w tmod C15 CG10176 Patj Atg18a eve Alh mtd Sema-5c mew D CG4702 dap mnt Pino h lok tutt tld velo lama vul bou Sep5 CG8147 CG5498 CG5532 CG42516 larpr CG5346 CG34383 Cad74A tex l(1)G0255 danr dbo Puf3 Hmx CG30431 CG4232</i>

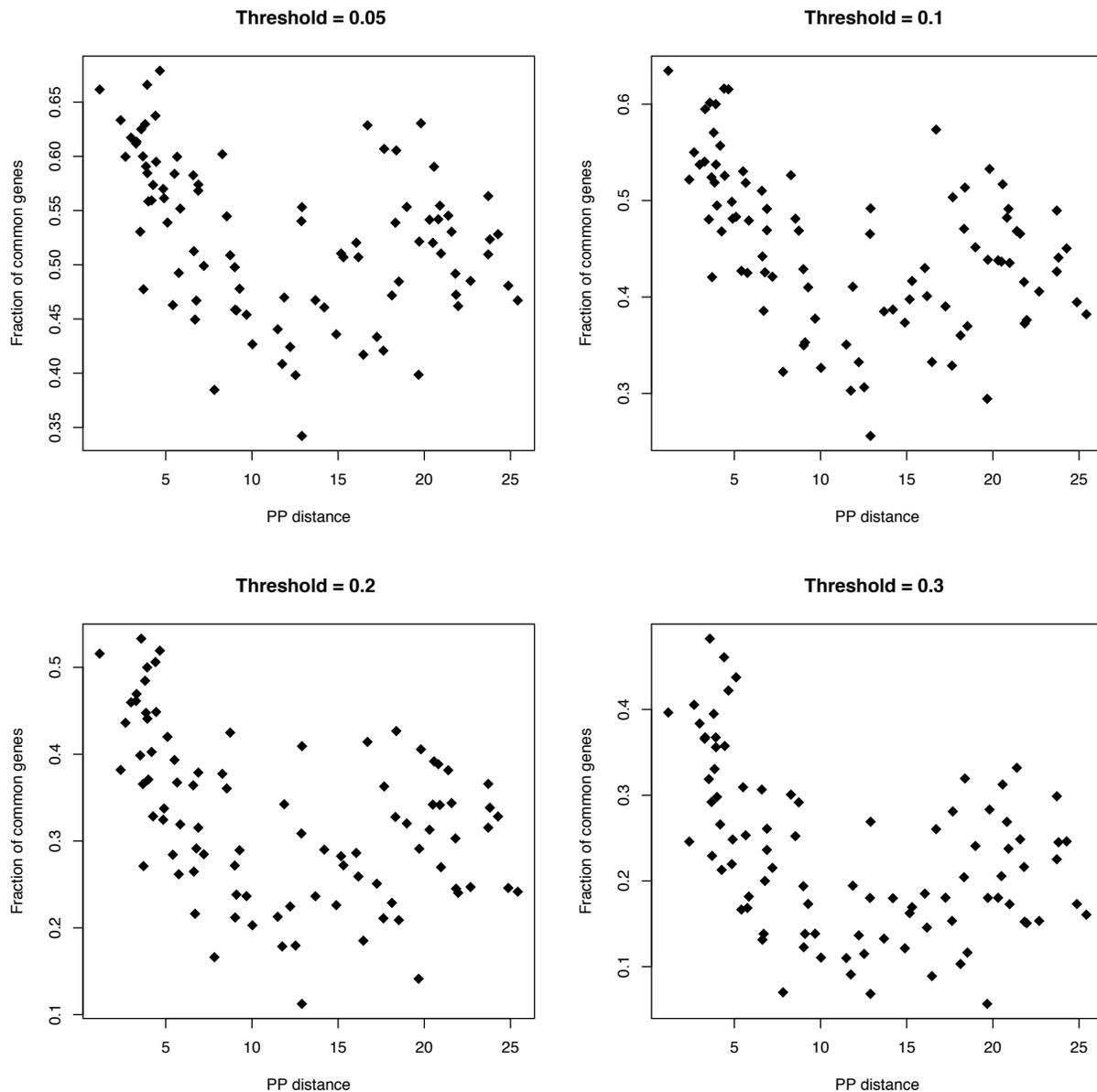
**Table S2:** Genes in PP categories 1-10. For the  $k$ -th PP category,  $k = 1, \dots, 21$ , genes are sorted in decreasing order according to their  $k$ -th sPP coefficients. Recall that for a gene, its  $k$ -th sPP coefficient is defined as the maximum  $k$ -th sPP coefficients across all the replicate expression pattern of the same gene. Only the top 100 genes are shown for each PP category. The entire list can be derived from *Dataset S5*.

PP category	Gene symbols
PP11	<i>ush net bun Ama Dl egr Doc1 zen C15 CG13653 Doc2 D CG10479 CG8147 toc rho esg CG45186 CG9505 Ect4 pan dbo scat CG2941 CG11076 Nrt Eip71CD tld tmod bel zen2 CG6885 CG2162 gk Dys Egrf nuf mew CycE CG42516 gwI CG7800 hth E2f1 CG14427 Sema-5c Fbxl7 CG11151 shn CG17698 Adf1 numb Acn CG11791 CG10176 jing MFS14 slmb pyd3 tutl brat CG5498 CG5888 Piezo Cys mas Vha68-2 mtd CG14915 ara CG34383 sprt lov CG11417 Srp68 ASPP dap chn CG32982 Bsg25D CG13784 CG4455 CG5059 Doc3 hyx CG8668 tex CG4080 Sep5 CG12420 Pino pcs GH06606 CG3838 vls Bap55 fz2 dpn neur CG11652</i>
PP12	<i>zen toc Doc3 link bun ASPP mor Z600 CG6885 Bub3 wisp tld CG42666 shn milt CG16815 CG10176 E5 CG14915 Tao Mlf CG14427 Su(H) Prx5 Eip71CD CG42516 bou Cys CG12581 Ama dpn Poc1 BEAF-32 Srp68 CG5059 CG2162 Acn CD98hc aay net CG11791 CG14657 CG12420 fz2 CycE Caf1 T-cp1 Cyp1 Btk29A Dl GH06606 hrg exer lov GILTI tmod RhoGAP1A Pino Ocho h Akap200 ftz fu2 dnk Fbxl7 CG3838 Miro gk Ect4 janA CaBP1 pcs ife CG8668 chif Ah Set2 CG11418 Mfap1 numb CG12702 jbug CG11652 esg 18w egg Oaz CG14712 pros MESR3 cmb Aldh-III Bsg25D CenB1A CG11007 Adf1 CG34383 nuf CG7182 danr SoxN CG42666 CG16815 CG14427 Meltrin CG34383 aay ASPP CG18446 ind zen E5 neur MESR3 tld milt Tao CG14657 CG42516 srp Dl MFS14 CG12702 dpn CG14915 CG11417 hrg CG10176 slp1 Pino CG2941 Cys CG12581 Ama CG5059 Btk29A dnk mor Acn Mf CycB eIF-4E janA fz2 Su(H) pros Usp5 h BEAF-32 CG10082 ifc wisp Neu2 Miro Bub3 fu2 CG11791 Lsd-2 Bsg25D Ilp4 lov RnrS ftz Mink Poc1 toc CD98hc Cad74A D19A Cyp1 Aldh-III comm2 Set2 Imp chif CG10075 Caf1 Adf1 BicD Prx5 CG14805 tutl CG11652 Mnt eve CG5888 larp jbug mnd Mkp3 CG42232 Ah GH06606 danr Traf4 CG14814 SCOT Ocho CG5027 yem rho sog NetA Ilp4 ASPP Ama trx Dg Pka-C3 CG42666 CG2941 CycB gukh neur Egm Mes4 E2f1 Asph dm bou</i>
PP13	<i>aay milt Cys Adf1 Bub3 MFS14 toc Tao vnd twi Btk29A CG14657 Acn CG32982 sala Prx5 Usp5 RnrS dpn CG5027 CG5059 CG11070 CG2469 Aldh-III CG14427 phm fz2 mor hrg Set2 Meltrin CaMKII Elba2 Sdc CG8001 ftz CHO24 CG11652 phyl Poc1 Caf1 CG14805 Cct5 CG8312 E5 sna MESR3 janA Bsg25D wntD h CG12702 CG42516 Cyp310a1 tld CG10176 htl Traf4 Su(H) eve Ah Lk6 Neu2 path Mnt Mem3 Cad74A E(spl)m8-HLH egg chif CG11417 CG6983 Ocho term Sep5 pnt slp1 Mlf rdx ind</i>
PP14	<i>htl Ama NetA Cyp310a1 neur CG9005 CycE Mes2 CG12177 mor Ndae1 egg CG32982 sna Mlf Fbxl7 hyx CG8312 Lk6 Pka-C3 Sdc gukh T-cp1 dm CG5027 milt sala Tao CG8066 CG14427 Vha68-2 Ilp4 CG2941 shn CG11791 wntD RnrS Bub3 trx flr Mdr49 CG42666 bou twi Aatf Bap55 CG1910 CG11652 Mes4 Gapdh1 Asph CG10075 dpn Mfap1 CG11076 Dg roX1 CD98hc cmb CG8788 E2f1 CG12702 CG12391 CG8230 toc CG10413 Mcm3 brat phm CG8001 phyl Egm ttm50 wisp Prx5 CycA CG14805 CG2469 CG11444 CG10984 Piezo Adf1 Dl CycB Cct5 CG11418 tmod BEAF-32 ASPP CG14657 Cad74A ry CG42516 E(spl)m8-HLH CG11417 CG2915 vnd aay Acn Poc1</i>
PP15	<i>NetA Mes2 Traf4 Pka-C3 CG12177 sna CG32982 stumps Cyp310a1 wntD twi gukh Fbxl7 CycB Ilp4 Ndae1 milt CG3036 CG11247 Asph Mdr49 sala CG14427 Shroom step trx GEFmeso phm neur CG1416 Sep5 Cys rad50 Btk29A CycA bou Ah htl Caf1 Acn CG8312 CG11652 prd opa CG9005 brat eve danr ftz tkv fj fz2 run CG11696 ry CG5346 rdx CG42232 CycT emc Bsg25D chif net mtd RnrS CG10631 18w CG10075 term h Mnt D Mink Ama bai pan mor Poc1 dpn CG31909 shn Cen phyl CG11007 Cad74A Adf1 E2f1 CG11076 Lk6 CG2941 roX1 BEAF-32 osk wde nerfin-1 CG8001 fu2 CG8319 CycE CG33116</i>
PP16	<i>Blimp-1 hb cad fj Ppa Ah CG5888 Mkp3 pcs Atg1 D CG8147 lov ftz Nek2 CG43394 bai salm toc Moe run CG14814 dpn Imp Set2 CycE CG14805 Btk29A CG11652 ken Usp5 milt CG30431 sala Acn tld CG14915 ird5 E5 chif hrg Akap200 gk Bsg25D Pepck sprt siz Cys p115 Abd-B CG14427 CG11696 eve whd Ama wde CG12702 Vinc jigr1 sc D19A dbo 26-29-p Mnt ifc ASPP CG11417 CG5059 mas gukh Sep5 Caf1 danr exer Z600 DnaJ-1 CG32982 comm2 fz2 18w h janA cmb Mink RhoBTE vfl CG12391 Gdi Sdc Atet Poc1 Ptx1 Cct5 grn rau CG16815 Prx5 CG11791 CG14712 phm rau scw MESR3 tipE CG3036 Atet CG8147 Doc2 Pepck hb sprt Doc1 mRps34 mas Atg1 Blimp-1 CG32982 cad CycA Nek2</i>
PP17	<i>net dm CG2941 CG2469 exer ebi mnd CG4080 CG14427 CG1910 Patj sc CG10479 Mkp3 CG30118 disco tutl CG13653 tup CG5888 salm veil vfl Pino pyd3 18w jigr1 CG11151 CG7182 CG10082 CD98hc Bsg25D CG34383 CG3838 htl CG2162 Abd-B whd T-cp1 mspo p115 rpr phm shn CG8654 dpn link GH06606 slmb Eip71CD egr run aay pcs CG14814 Srp68 gk Poc1 ttm50 CG42666 Prx5 siz CG10341 Ptx1 ara nst CG11076 Vinc mor larp CG4455 lama zen2 E2f1 bel CG10924 CG11652 rib sob nAChRalpha2</i>
PP18	<i>GEFmeso ry CG31909 Ndae1 MESR3 twi yem emc sala Fbxl7 Mif apt GH06606 trx tll RH48922 Mkp3 CG1146 byn bel Pepck Lk6 Pka-C3 Zasp52 salm step disco cad hb CycB Cyp310a1 Ama HLH54F fkh CHO24 Poc1 wde Mes4 whd CG2941 E(spl)m8-HLH Eip71CD CG11247 zen bou CG10924 CycT gukh pcs bmm roX1 CG11007 CG8147 CG9005 Mnt CG10365 CG14427 Bsg25D CG42232 htl CG5888 Cys CycA CG1416 Sep5 CG11652 sim CG12723 bai rad50 mRps34 CG14814 E(spl)m7-HLH osk larp milt Acn zfh2 phm tmod mor CycE RnrS Traf4 CG12177 BicD CG12581 Sdc CG11418 Atg1 Elba3 E5 CG11696 Asph Acf1 Adf1 Set2 Nek2 Dip-C Mdr49</i>
PP19	<i>bbg exer MFS14 apt tll dm Pepck zen Pino Mkp3 srp zen2 janA mRps34 Sema-5c sprt CG16815 CG11417 milt Doc3 mor Acn byn bai CG11652 CG12702 aay CG32982 Elba3 Eip71CD Usp5 CG14805 CG8147 Cys RhoBTE CG10365 CG10924 Cct5 Moe DnaJ-1 cmb CG1146 hkb ifc hrg Ptp4E ird5 sala Ptx1 Gdi Atg1 Vinc tld CG30118 CG10338 scw CG11151 Set2 ASPP CG14814 Nek2 CG14712 MESR3 CG12420 26-29-p whd vfl Dip-C CG14915 CG12391 cad Aldh-III Srp68 Argk CG2941 rib GH06606 E5 CG1416 CG1434 bmm CG11696 wisp p115 mnd CG10082 Mink Btk29A velo Caf1 Cyp1 yip2 CG8654 esg Sep5 CG11791 disco Mat89Ba Akap200 dnk lok RhoGAP1A 26-29-p rdx yem Hsp83 dap bel Pi3K21B cup CycB CG14814 exer CG8915 gwI pgc PpD3 DnaJ-1 wisp Dok Cyp1 CG14712 Hsp27 Cen dnk Cctgamma DyakGE19022 Rpn8 Tif-IA c(3)G Gapdh1 CG32473 CBP Tao Moe Sema-5c Atg1 Mitf CG10462 Wee1 l(1)G0255 gcl fu2 ird5 srp Mnt mtd DNaseII CG9925 Cct5 ebi CG10979 rl CG42232 NK7.1 jigr1 Dip-B MFS14 bun osk fkh aay cib Bsg25D ifc CG14427 Vha68-2 CG10365 zen2 shu flr CG10924 CG5292 whd CG33158 Z600 msk Hsp26 Acf1 Elba3 zpg tll eIF-4E MESR3 UGP Doc1 Pepck ovo CG10631 BicD CG12581 CG8036 lqfR dm CG2941 GH06606 Lsd-2 Gdi CHO24 yip2</i>
PP20	
PP21	

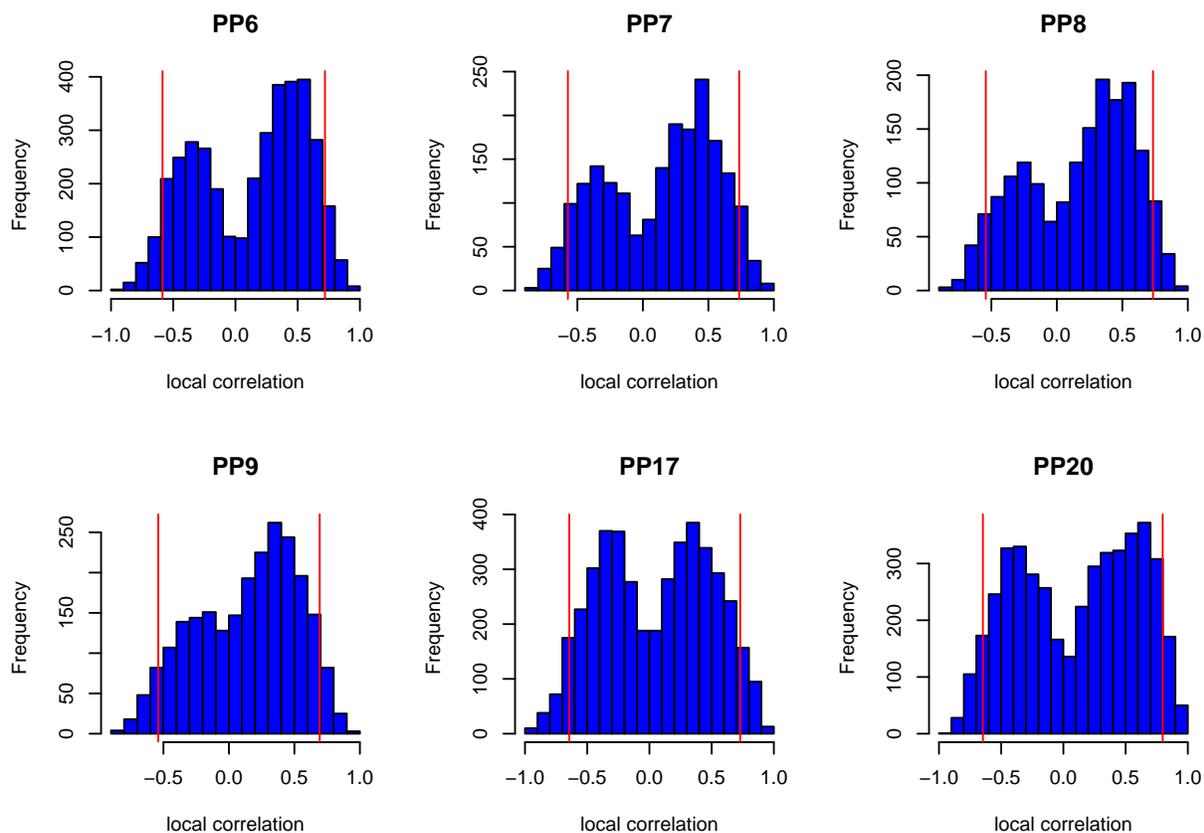
**Table S3:** Genes in PP categories 11-21. See the caption of Table S2 for details. The entire list can be derived from *Dataset S5*.



**Fig. S19:** Number of genes expressed in each of the 21 PP, with colors corresponding to different values of the sPP coefficients.



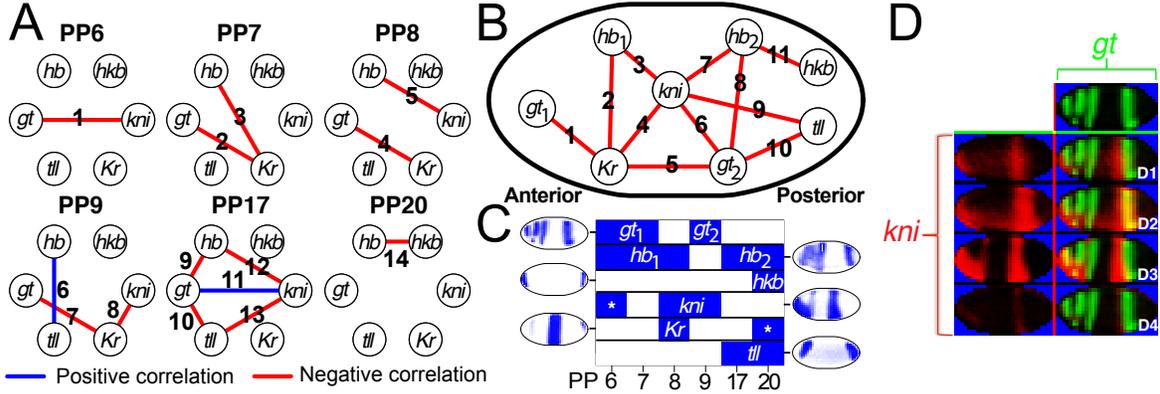
**Fig. S20:** The relationship between the fraction of common genes in a pair of PP categories and the centroid distance of the two PP. The fraction of common genes in a pair of PP categories is defined as the Jaccard distance between the two categories. Shown are the scatter plots of the fraction and the PP distance for different sPP coefficient threshold values in  $\{0.05, 0.1, 0.2, 0.3\}$ . As described in the main text, we thresholded the sPP coefficients to assign genes to PP categories. The plots show that although the absolute fraction changes as we increase the threshold, the down-up trend remains the same.



**Fig. S21:** Histograms of local correlations for the six gap-PP. The two red vertical lines in each histogram indicate the lower and upper five percentiles of the local correlations, respectively. The distribution of the correlations is bimodal, with one peak corresponding to positive correlations and the other to negative correlations. This is a direct consequence of our definition of local correlation: we used the maximum correlation of all pairwise local correlations between the replicate images of one gene and the replicate images of another gene (*SI Appendix Methods*). This procedure excluded image pairs that had low correlations.



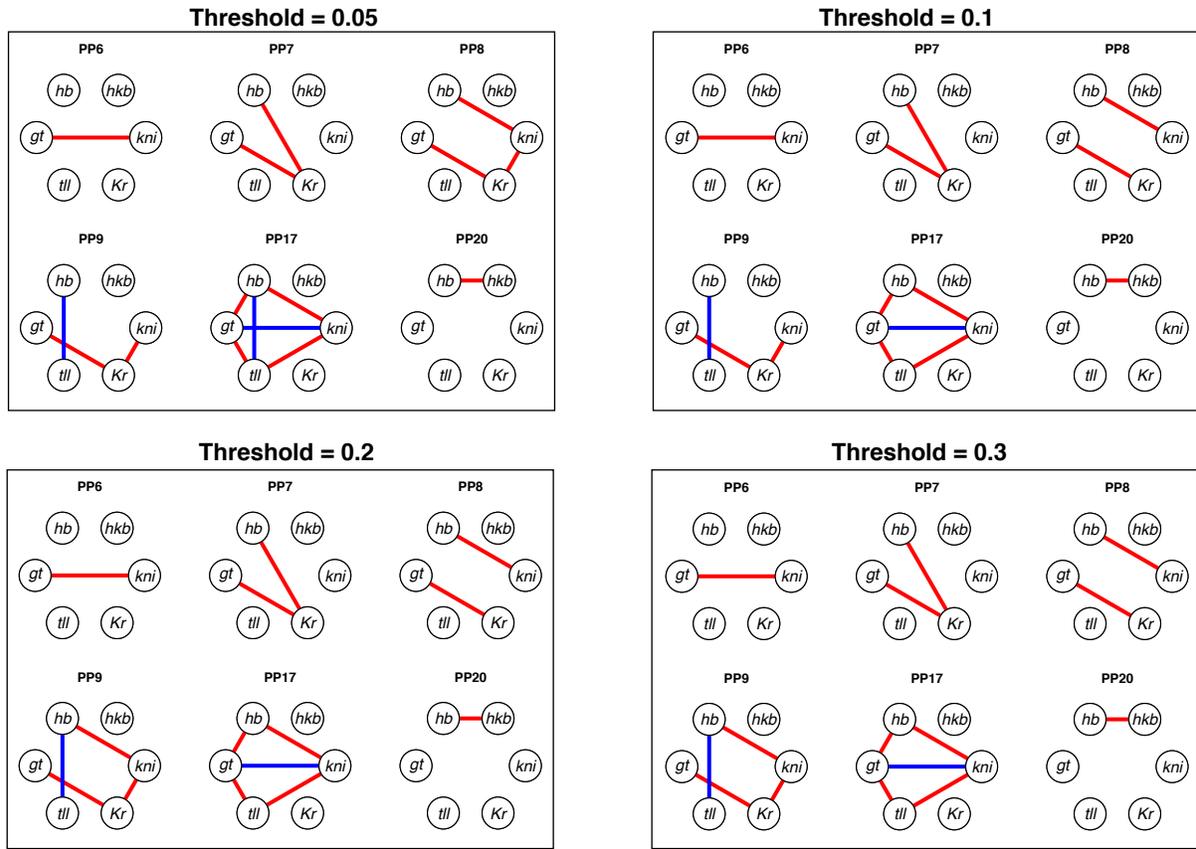




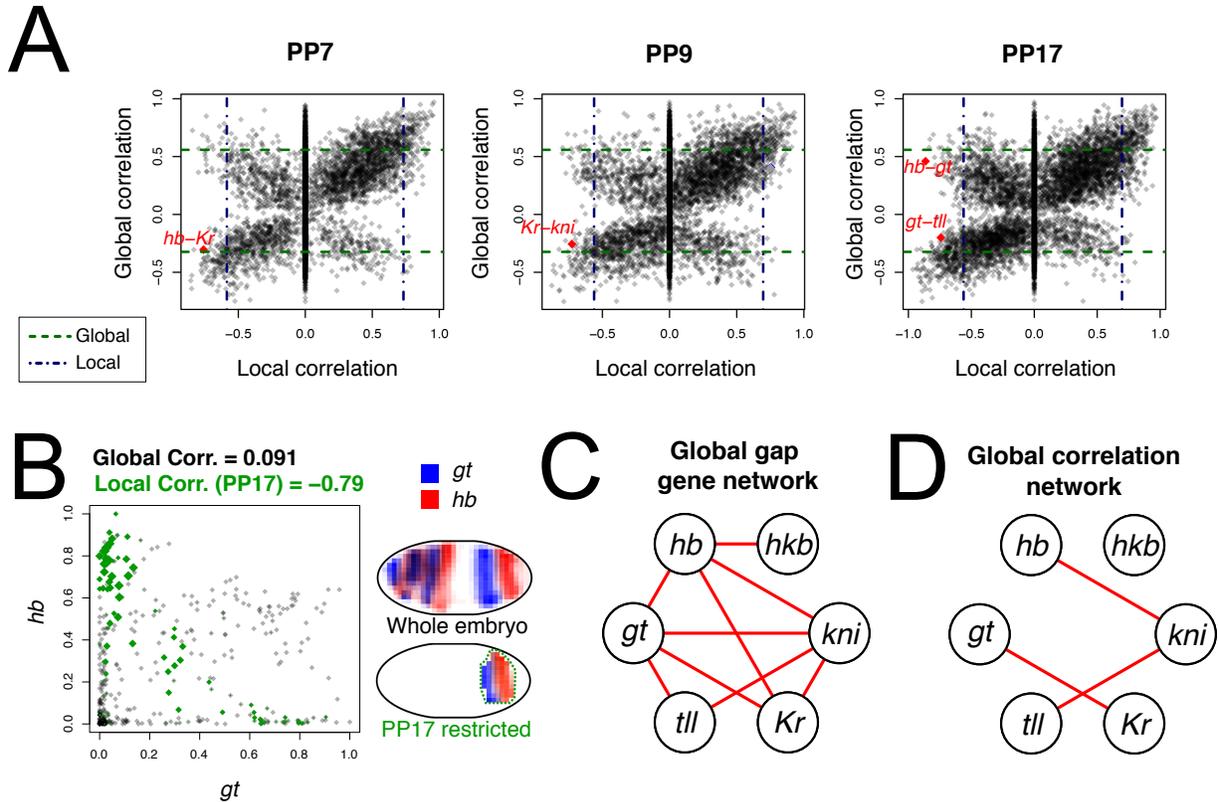
**Fig. S25:** Validating the SLCN with the gap gene network. Sub-figures A-C are the same as Fig 5A-C in the main text. (A) The six spatially local correlation networks for the six gap genes with each link labeled from 1 to 14. (B) The gap gene network depicting repressive interactions of six genes, as originally described in [31]. Links are numbered from 1 to 11 and multiple occurrence of the same gene are subscripted by numbers (e.g.  $hb_1$  and  $hb_2$ ). The directions of the interactions are not indicated. (C) Schema showing the composite mapping of gap gene expression to the six PP arranged in a linearly ordered way from anterior to posterior. For each of the six gap genes, the regions diagrammed in blue are the PP with sPP coefficient greater than or equal to 0.1 for at least one of the replicate images, while the regions diagrammed in white are the PP with a coefficient less than 0.1 for all replicate images. To evaluate the prediction performance of the SLCN in (A), we first mapped each node in (B) to a connected PP group in (C). According to (C),  $gt$  has two major components, the anterior part which has expression in PP6 and PP7, and the posterior part that has expression in PP9. The anterior  $gt_1$  and the posterior  $gt_2$  symbols in (B) can be mapped to these two components respectively.  $hb$  also has two major connected PP components, the anterior part which has expression in PP6-8 that corresponds to  $hb_1$  in (B), and the posterior part that has expression in PP17 and PP20 that corresponds to  $hb_2$  in (B). For  $hkb$ , the only expression in PP20 corresponds to the  $hkb$  gene symbol in (B).  $kni$  has two components. The first one in PP6 does not correspond to any node in (B) (the \* symbol indicates a region of gene expression with no match in (B)), whereas the second one in PP8 and PP9 corresponds to the  $kni$  symbol in (B). Similarly, the first component of  $Kr$  in PP8 corresponds to the symbol  $Kr$  in (B), whereas the posterior part in PP20 does not appear in (B). Finally, the only component of  $tll$  in PP17 and PP20 correspond to the only  $tll$  symbol in (B). We considered an interaction between two gene nodes G1 and G2 in the schematic network as successfully identified by our method, if the same interaction exists in the any SLCN associated with the overlapping PP in the connected PP groups corresponding to G1 and G2. We examined each link in (B) and the results are summarized in Table S4. Our SLCN recovered 10 out of 11 interactions in the gap gene network and discovered three interactions not described by the network. (D) Overlaying gene expressions for  $gt$  and  $kni$ . There are four replicate images for  $kni$  and one for  $gt$ . A clear complementary patterns in the anterior region (PP6) between the two genes can be seen in D2 and D3. Moreover in D2, the two genes have a significant amount of overlap in the posterior region (PP17).

Link in gap gene network	G1	G1 PP	G2	G2 PP	Overlapping PP	Link(s) in SLCN
1	<i>gt<sub>1</sub></i>	PP6,7	<i>Kr</i>	PP8	PP7,8	2,4
2	<i>hb<sub>1</sub></i>	PP6-8	<i>Kr</i>	PP8	PP7,8	3
3	<i>hb<sub>1</sub></i>	PP6-8	<i>kni</i>	PP8,9	PP7-9	5
4	<i>kni</i>	PP8,9	<i>Kr</i>	PP8	PP8,9	8
5	<i>gt<sub>2</sub></i>	PP9	<i>Kr</i>	PP8	PP8,9	4,7
6	<i>gt<sub>2</sub></i>	PP9	<i>kni</i>	PP8,9	PP8,9	No link
7	<i>hb<sub>2</sub></i>	PP17,20	<i>kni</i>	PP8,9	PP9,17	12
8	<i>gt<sub>2</sub></i>	PP9	<i>hb<sub>2</sub></i>	PP17,20	PP9,17	9
9	<i>kni</i>	PP8,9	<i>tll</i>	PP17,20	PP9,17	13
10	<i>gt<sub>2</sub></i>	PP9	<i>tll</i>	PP17,20	PP9,17	10
11	<i>hb<sub>2</sub></i>	PP17,20	<i>hkb</i>	PP20	PP17,20	14

**Table S4:** Validating the SLCN with the gap gene network. Link in gap gene network: link number in the schematic gap gene network (Fig. S25B). G1 and G2: gene nodes in the schematic gap gene network. G1 PP and G2 PP: the connected PP group in the linearly ordered PP representation that correspond to G1 and G2 respectively (Fig. S25C). Overlapping PP: the overlapping PP of G1 PP and G2 PP. Link(s) in the SLCN: the link(s) in the predicted SLCN (Fig. S25A) that correspond to a link in the schematic gap gene network. Out of 11 links in the schematic gap gene network, there is one (i.e. Link 6) that has no corresponding link in the SLCN. There are three links out of 14 in the SLCN that have no corresponding links in the gap gene network diagram.



**Fig. S26:** Sub-networks of the six SLCN that contain the six gap genes, for sPP coefficient threshold in  $\{0.05, 0.1, 0.2, 0.3\}$ . As described in the main text and in *SI Appendix Methods*, we thresholded the sPP coefficients to filter TF for SLCN construction. The reconstructed gap gene networks are nearly identical for these threshold values.



**Fig. S27:** Correlating transcription factors (TF) over the whole embryo (global correlation). (A) Scatter plots of the global correlation vs. the local correlations for PP7, PP9 and PP17. The dashed lines correspond to the lower and upper cutoffs for the local correlations (vertical lines) and global correlations (horizontal lines). See *SI Appendix Methods* for how we determined the cutoffs for the global correlations. The four branches of each plot are due to the fact that the distributions of both global correlations and local correlations are bimodal. Highlighted in the scatter plots are the gap-gene links correctly identified by the local networks but missed by the global network. (B) The PP-based correlation approach detected locally complementary patterns whereas the global correlation approach failed. The scatterplot showed the pixel-wise intensity relationship between a pair of expression images of *gt* and *hb*. The green dots corresponded to the pixels in the region defined by PP17, with dot size proportional to the pixel intensity of PP17. We observed a clear negative association between the two TF in PP17. However, this association disappears when we consider the scatterplot of all 405 pixels of the embryo. (C) The gap gene network without the spatial information. Here, two gap genes share a link if they share a link in the schematic gap gene network (Fig. S25B) regardless of the location of the interaction. (D) The gap gene constructed based on correlation measurements over the whole embryo identified only three out of nine links of the global version of gap gene network in (C).

## References

- [1] Brunet JP, Tamayo P, Golub TR, Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences* 101:4164–4169.
- [2] (2013) Berkeley drosophila genome project in situ homepage. (<http://insitu.fruitfly.org/cgi-bin/ex/insitu.pl>) Accessed: 2013-03.
- [3] Frise E, Hammonds AS, Celniker SE (2010) Systematic image-driven analysis of the spatial drosophila embryonic expression landscape. *Molecular systems biology* 6.
- [4] Puniyani K, Faloutsos C, Xing EP (2010) Spex2: automated concise extraction of spatial gene expression patterns from fly embryo ish images. *Bioinformatics* 26:i47–i56.
- [5] Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401:788–791.
- [6] Wang Y, Jia Y, Hu C, Turk M (2005) Non-negative matrix factorization framework for face recognition. *International Journal of Pattern Recognition and Artificial Intelligence* 19:495–511.
- [7] Xu W, Liu X, Gong Y (2003) *Document Clustering Based on Non-negative Matrix Factorization*, SIGIR '03 (ACM, New York, NY, USA), pp 267–273.
- [8] Virtanen T (2007) Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *IEEE Transactions on Audio, Speech, and Language Processing* 15:1066–1074.
- [9] Gillis N (2014) The why and how of nonnegative matrix factorization. *Regularization, Optimization, Kernels, and Support Vector Machines* 12:257.
- [10] Liu J, Liu J, Wonka P, Ye J (2012) Sparse non-negative tensor factorization using columnwise coordinate descent. *Pattern Recognition* 45:649–656.
- [11] Marial J, Bach F, Ponce J, Sapiro G (2010) Learning the parts of objects by non-negative matrix factorization. *Journal of Machine Learning Research* 11:19–60.
- [12] Amari S, Cichocki A, Yang HH (1996) A new learning algorithm for blind signal separation. *Advances in neural information processing systems* pp 757–763.
- [13] Gribonval R, Schnass K (2010) Dictionary identification – sparse matrix-factorisation via  $l_1$ -minimisation. *IEEE Transactions on Information Theory* 56(7):3523–3539.
- [14] Geng Q, Wang H, Wright J (2011) On the local correctness of  $l_1$  minimization for dictionary learning. *Technical report, preprint arXiv:1102.1249.2011*.
- [15] Gribonval R, Jenatton R, Bach F (2014) Sparse and spurious: dictionary learning with noise and outliers. *Technical report, preprint arXiv:1407.5155v3*.
- [16] Wu S, Yu B (2015) Local identifiability of  $l_1$ -minimization dictionary learning: a sufficient and almost necessary condition. *arXiv preprint arXiv:1505.04363*.

- [17] Donoho D, Stodden V (2004) in *Advances in Neural Information Processing Systems 16*, eds Thrun S, Saul LK, Schölkopf B (MIT Press), pp 1141–1148.
- [18] Tan VY, Févotte C (2013) Automatic relevance determination in nonnegative matrix factorization with the  $\beta$ -divergence. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35:1592–1605.
- [19] Sun M, Zhang X (2015) A stable approach for model order selection in nonnegative matrix factorization. *Pattern Recognition Letters* 54:97–102.
- [20] Lohs-Schardin M, Cremer C, Nüsslein-Volhard C (1979) A fate map for the larval epidermis of *Drosophila melanogaster*: localized cuticle defects following irradiation of the blastoderm with an ultraviolet laser microbeam. *Developmental biology* 73:239–255.
- [21] Hartenstein V (1993) *Atlas of Drosophila Development* (Plainview: Cold Spring Harbor Laboratory Press).
- [22] Hammonds AS, et al. (2013) Spatial expression of transcription factors in *Drosophila* embryonic organ development. *Genome Biology* 14:R140.
- [23] Ashraf SI, Hu X, Roote J, Ip YT (1999) The mesoderm determinant snail collaborates with related zinc-finger proteins to control *Drosophila* neurogenesis. *The EMBO journal* 18:6426–6438.
- [24] Slawski M, Hein M (2013) Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization. *Electronic Journal of Statistics* 7:3004–3056.
- [25] Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* pp 267–288.
- [26] Liu H, Yu B (2013) Asymptotic properties of lasso+mls and lasso+ridge in sparse high-dimensional linear regression. *Electronic Journal of Statistics* 7:3124–3169.
- [27] Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33:1–22.
- [28] Hastie T, Tibshirani R, Friedman J (2001) *The Elements of Statistical Learning*, Springer Series in Statistics (Springer New York Inc., New York, NY, USA).
- [29] Pruteanu-Malinici I, Mace DL, Ohler U (2011) Automatic annotation of spatial expression patterns via sparse bayesian factor models. *PLoS computational biology* 7:e1002098.
- [30] Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks* 10:626–634.
- [31] Jaeger J (2011) The gap gene network. *Cellular and Molecular Life Sciences* 68:243–274.