Supplementary materials for "A graphical model method for integrating multiple sources of genome-scale data"

Daniel Dvorkin,[*] Brian Biehs, Katerina Kechris

June 27, 2013

Supplementary Appendices

Supplementary Figures

Supplementary Tables

---
[*]corresponding author: daniel.dvorkin@gmail.com

A   Initialization and termination

We use a "data-driven" method to initialize the marginal models. The starting means for the components are chosen from the quantiles of the data; specifically, the yth mean is the $\{(K - y + 1) - 0.5\}/K$th quantile. For example, when $K = 2$, the starting means are the 0.75th and 0.25th quantiles, while when $K = 3$, the starting means are approximately the 0.833th, 0.5th, and 0.167th quantiles. Implicitly, we are assuming that $p_1 = \cdots = p_K = 1/K$, ordering the components by $\mu_1 > \cdots > \mu_K$, and choosing the median of each component for the starting mean, justified by the fact that both the normal and PVII distributions are symmetrical. In the multivariate case, this procedure is applied to each column of the data. We then calculate the $N \times K$ similarity matrix S as the inverse of Manhattan distances from each sample to each starting mean, and initialize the weight matrix by dividing the similarities for each sample by the sum of its similarities:

$$
\begin{aligned}
s_{n,y} &= 1/\left(\epsilon + \textstyle\sum_d |x_{n,d} - \mu_{y,d}|\right), \\
w_{n,y} &= s_{n,y}/\textstyle\sum_k s_{n,k}
\end{aligned}
\tag{S1}
$$

where ϵ is a very small number, here the R standard of approximately 1.49×10^{-8}. The initial W is then used as input for the first M-step of the algorithm. This method has proven very stable for a wide variety of real and simulated data, and for the current data sets appears to perform at least as well as (and often better than) the use of multiple random restarts for initialization, with much lower computational cost.

For the joint models, we fit marginal models to each data source and use the posterior probability matrices from these fits for the initial $W_1, \ldots, W_Z$. In the layered model, we also fit marginal models with $K = K_0$ to each data source to obtain matrices we may call $W_z^{K_0}$ (observe that $W_z^{K_0} = W_z$ when $K_0 = K_z$) and average these matrices to obtain our initial $U = \frac{1}{Z} \sum_z W_z^{K_0}$. In the chained model, $U = W_1^{K_0}$. Finally, we calculate the initial transition matrices Q_z as the non-negative least-squares solutions (Mullen and van Stokkum, 2012) to $UQ_1 = W_1$ for both the layered and chained models, $UQ_z = W_z$ for layered, and $W_{z-1}Q_z = W_z$ for chained. With these we can calculate the initial V_z's and proceed with the first M-step.

In both marginal and joint models, we set the maximum number of iterations to 1000. We calculate the log-likelihood $\mathcal{L}(\theta^{(i)})$ at each iteration, and terminate when the maximum number of iterations is reached or when changes in log-likelihood between iterations become sufficiently small: $|\{\mathcal{L}(\theta^{(i)})/\mathcal{L}(\theta^{(i-1)})\} - 1| \le 10^{-6}$. This criterion usually leads to termation after 100-300 iterations.

B   The joint model E-step

For compactness in the discussion of the E-step, let all parameter notation refer to the previous iteration's parameter estimates. For example, $p_{z,y_z} = \mathrm{f}(y_z|\theta) = \Pr(\mathcal{Y}_z = y_z|\theta^{(i-1)})$, with parameters $\theta = \theta^{(i-1)}$. We will also write density functions without explicit reference to the parameters, such as $g_{z,n,y_z} = \mathrm{f}_{y_z}(\boldsymbol{x}_{z,n})$ rather than $g_{z,n,y_z} = \mathrm{f}_{y_z}(\boldsymbol{x}_{z,n}|\theta)$.

Let $\alpha_{0,n,y_0} = p_{0,y_0}$ for all $n = 1, \ldots, N$. For the layered model, we also define the intermediate terms

$$
\begin{aligned}
g'_{z,n,y_0} &= \mathrm{f}(\boldsymbol{x}_{z,n}|y_0) &&= \textstyle\sum_{k_z} g_{z,n,k_z} q_{y_0,k_z}, \\
g^\star_{n,y_0} &= \mathrm{f}(\boldsymbol{x}_{\cdot,n}|y_0) &&= \textstyle\prod_z g'_{z,n,y_0}, \\
\beta_{0,n,y_0} &= \mathrm{f}(\boldsymbol{x}_{\cdot,n}, y_0) &&= g^\star_{n,y_0} \alpha_{0,n,y_0}, \\
\gamma_n &= \mathrm{f}(\boldsymbol{x}_{\cdot,n}) &&= \textstyle\sum_{k_0} \beta_{0,n,k_0}.
\end{aligned}
\tag{S2}
$$

Then the posterior probabilities are calculated as

$$
\begin{aligned}
u_{n,y_0} &= \beta_{0,n,y_0}/\gamma_n, \\
v_{z,y_0,n,y_z} &= u_{n,y_0} g_{z,n,y_z} q_{z,y_0,y_z}/g'_{z,n,y_0}, \\
w_{z,n,y_z} &= \textstyle\sum_{k_0} v_{z,k_0,n,y_z}.
\end{aligned}
\tag{S3}
$$

What may seem excessive formalism in the terms defined above is to facilitate efficient matrix computation. Let G_z be the $N \times K_z$ matrix of which the (n, y_z)th element is g_{z,n,y_z}, and similarly define the matrices α_0, G'_z, $G^\star$, and β_0, and the vector γ. Also let Q_z be the $K_0 \times K_z$ matrix of which the (y_0, y_z)th element is $q_{z,y_0,y_z} = \mathrm{f}(y_z|y_0)$.

Now for a vector $\boldsymbol{a}$ of length J, and matrices B and C of dimension $J \times K$, let "$\circ$" denote the Hadamard (element-wise) product of the matrices, that is, $B \circ C$ is the $J \times K$ matrix such that $(B \circ C)_{j,k} = b_{j,k}c_{j,k}$. Define the row-wise vector-matrix product as $(\boldsymbol{a} \circ B)_{j,k} = a_j b_{j,k}$, and similarly define row-wise division as $(B/\boldsymbol{a})_{j,k} = b_{j,k}/a_j$. Finally,

define row-wise matrix summation so that $\sum_k(\boldsymbol{B})$ is the J-length vector of which the jth element is $\sum_k b_{j,k}$. Then Equations (S2) and (S3) become

$$
\begin{aligned}
\boldsymbol{G}'_z &= \boldsymbol{G}_z \boldsymbol{Q}_z^{\mathrm{T}}, \\
\boldsymbol{G}^\star &= \textstyle\prod_z \boldsymbol{G}'_z, \\
\boldsymbol{\beta}_0 &= \boldsymbol{\alpha}_0 \circ \boldsymbol{G}^\star, \\
\boldsymbol{\gamma} &= \textstyle\sum_{k_0}(\boldsymbol{\beta}_0), \\
\boldsymbol{U} &= \boldsymbol{\beta}_0 / \boldsymbol{\gamma}, \\
\boldsymbol{V}_{z,y_0} &= \boldsymbol{u}_{\cdot,y_0} \circ (\boldsymbol{q}_{z,y_0,\cdot} \circ \boldsymbol{G}_z^{\mathrm{T}})^{\mathrm{T}} / \boldsymbol{g}'_{z,\cdot,y_0}, \\
\boldsymbol{W}_z &= \textstyle\sum_{k_0} \boldsymbol{V}_{z,k_0}
\end{aligned}
\tag{S4}
$$

where $\boldsymbol{B}^{\mathrm{T}}$ denotes the transpose of $\boldsymbol{B}$. Note that in the final line of Equation (S4), the operation is not row-wise matrix summation, but summation over a list of matrices; that is, $\boldsymbol{W}_z = \boldsymbol{V}_{z,1} + \cdots + \boldsymbol{V}_{z,K_0}$ where each $\boldsymbol{V}_{z,k_0}$ is an $N \times K_z$ matrix.

For the chained model, let $\boldsymbol{x}_{z-,n} = (\boldsymbol{x}_{1,n}, \ldots, \boldsymbol{x}_{z,n})$, and similarly $\boldsymbol{x}_{z+,n} = (\boldsymbol{x}_{z,n}, \ldots, \boldsymbol{x}_{Z,n})$. Also let $r_{z,y_z,y_{z-1}} = \mathrm{f}(y_{z-1}|y_z) = p_{z-1,y_{z-1}} q_{z,y_{z-1},y_z} / p_{z,y_z}$, where $q_{z,y_{z-1},y_z} = \mathrm{f}(y_z|y_{z-1})$. Then the intermediate terms are

$$
\begin{aligned}
\alpha_{z,n,y_z} &= \mathrm{f}(\boldsymbol{x}_{z-,n}, y_z) &=& \textstyle\sum_{k_{z-1}} \alpha_{z-1,n,k_{z-1}} q_{z,k_{z-1},y_z} g_{z,n,y_z}, \\
\beta_{Z,n,y_Z} &= \mathrm{f}(\boldsymbol{x}_{Z,n}, y_Z) &=& p_{Z,y_Z} g_{Z,n,y_Z}, \\
\beta_{z,n,y_z} &= \mathrm{f}(\boldsymbol{x}_{z+,n}, y_z) &=& \textstyle\sum_{k_{z+1}} \beta_{z+1,n,k_{z+1}} r_{z+1,k_{z+1},y_z} g_{z,n,y_z} \text{ for } z < Z, \\
\beta_{0,n,y_0} &= \mathrm{f}(\boldsymbol{x}_{\cdot,n}, y_0) &=& \textstyle\sum_{k_1} \beta_{1,n,k_1} r_{1,k_1,y_0}
\end{aligned}
\tag{S5}
$$

while γ_n is as in Equation (S2). Then u_{n,y_0} is as in Equation (S3), and the other posterior probabilities are

$$
\begin{aligned}
v_{z,y_{z-1},n,y_z} &= \alpha_{z-1,n,y_{z-1}} \beta_{z,n,y_z} q_{z,y_{z-1},y_z} / (p_{z,y_z} \gamma_n), \\
w_{z,n,y_z} &= \textstyle\sum_{k_{z-1}} v_{z,k_{z-1},n,y_z}.
\end{aligned}
\tag{S6}
$$

In matrix form, $\boldsymbol{R}_z = (\boldsymbol{p}_{z-1} \circ \boldsymbol{Q}_z)^{\mathrm{T}} / \boldsymbol{p}_z$, and Equations (S5) and (S6) become

$$
\begin{aligned}
\boldsymbol{\alpha}_z &= (\boldsymbol{\alpha}_{z-1} \boldsymbol{Q}_z) \circ \boldsymbol{G}_z, \\
\boldsymbol{\beta}_Z &= (\boldsymbol{p}_Z \circ \boldsymbol{G}_Z^{\mathrm{T}})^{\mathrm{T}}, \\
\boldsymbol{\beta}_z &= (\boldsymbol{\beta}_{z+1} \boldsymbol{R}_{z+1}) \circ \boldsymbol{G}_z \text{ for } z < Z, \\
\boldsymbol{\beta}_0 &= \boldsymbol{\beta}_1 \boldsymbol{R}_1, \\
\boldsymbol{V}_{z,y_{z-1}} &= \{(\boldsymbol{\alpha}_{z-1,\cdot,y_{z-1}} \circ \boldsymbol{\beta}_z)^{\mathrm{T}} \circ \boldsymbol{q}_{z,y_{z-1},\cdot} / \boldsymbol{p}_z\}^{\mathrm{T}} / \boldsymbol{\gamma}, \\
\boldsymbol{W}_z &= \textstyle\sum_{k_{z-1}} \boldsymbol{V}_{z,k_{z-1}}
\end{aligned}
\tag{S7}
$$

with $\boldsymbol{\gamma}$ and $\boldsymbol{U}$ as in Equation (S4).

The calculations in the chained model are almost the same as those in the Baum-Welch algorithm (Baum et al., 1970; McLachlan and Krishnan, 2008, pp. 290-293) for hidden Markov models, but the goal differs: because here the $\mathcal{Y}_z$'s are heterogeneous, we wish to estimate the individual transition matrices $\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_Z$ rather than a single transition matrix for the entire model. In the special case of $K_0 = K_1 = \cdots = K_Z$, the Baum-Welch estimator for the transition matrix in the standard hidden Markov model can be seen as an average of the estimators for the $\boldsymbol{Q}_z$'s in the M-step of our method.
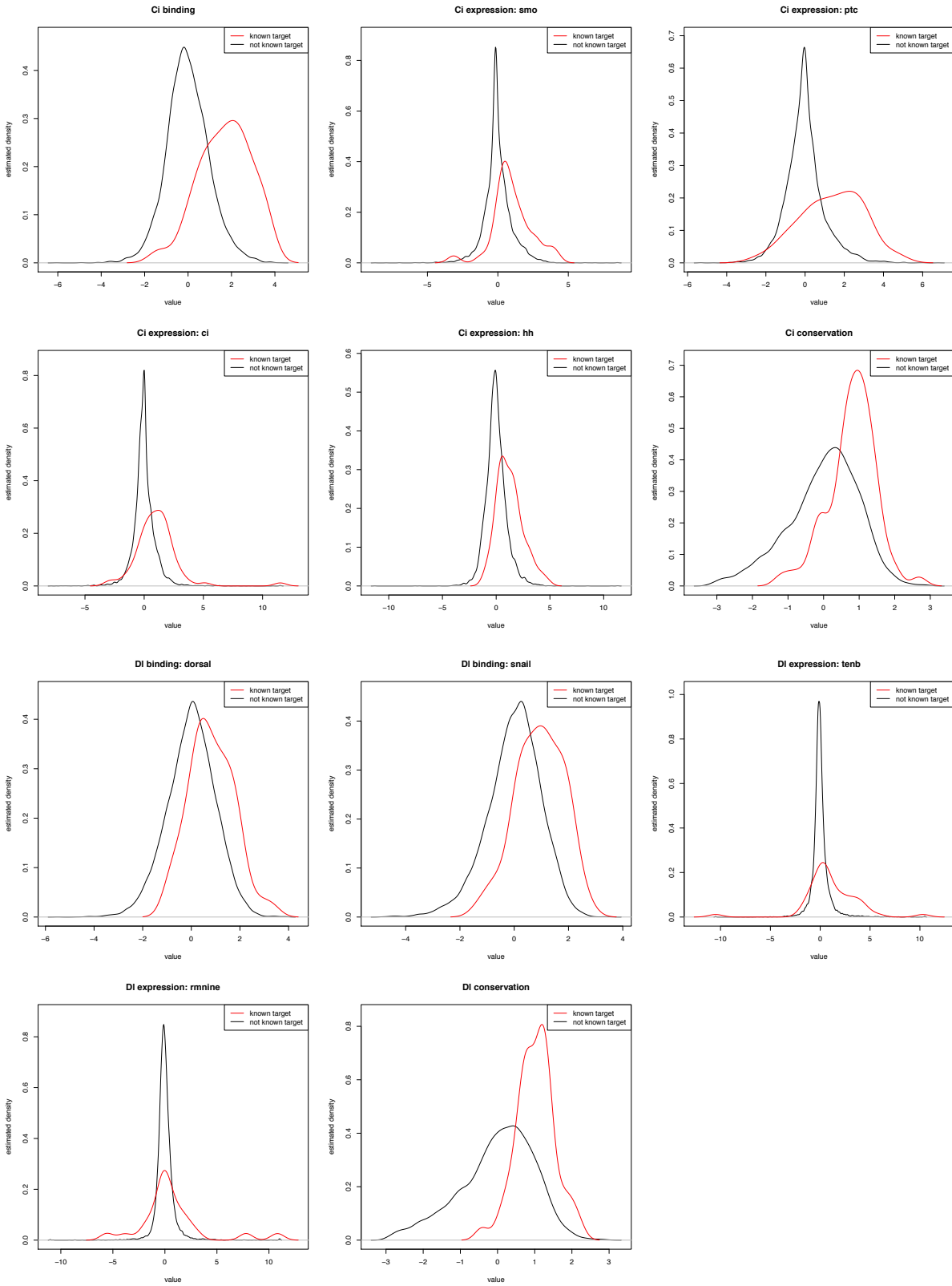
Supplementary Figure 1: Nonparametric density estimates for all Ci and Dl data sources, for known target genes and genes of unknown target status. In Dl expression, "tenb" refers to $pipe^-/pipe^-$ vs. $Toll^{10B}$, while "rmnine" refers to $pipe^-/pipe^-$ vs. $Toll^{rm9}/Toll^{rm10}$.

Supplementary Table 1: Marginal data simulation parameters.

type	dimension	number of components	component probabilities	component means	component variances
binding	1	2	$\begin{pmatrix} 0.03 \\ 0.97 \end{pmatrix}$	$\begin{pmatrix} 0.825 \\ -0.825 \end{pmatrix}$	$\begin{pmatrix} 1.50 \\ 0.50 \end{pmatrix}$
expression	3	3	$\begin{pmatrix} 0.05 \\ 0.92 \\ 0.03 \end{pmatrix}$	$\begin{pmatrix} 1.50 & 1.50 & 1.50 \\ 0.00 & 0.00 & 0.00 \\ -1.50 & -1.50 & -1.50 \end{pmatrix}$	$\begin{pmatrix} \begin{pmatrix} 3.00 & 1.00 & -0.50 \\ 1.00 & 3.00 & 0.00 \\ -0.50 & 0.00 & 3.00 \end{pmatrix} \\ \begin{pmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{pmatrix} \\ \begin{pmatrix} 2.00 & -0.20 & 0.10 \\ -0.20 & 2.00 & 0.10 \\ 0.10 & 0.10 & 2.00 \end{pmatrix} \end{pmatrix}$
conservation	1	2	$\begin{pmatrix} 0.25 \\ 0.75 \end{pmatrix}$	$\begin{pmatrix} 1.25 \\ -1.25 \end{pmatrix}$	$\begin{pmatrix} 1.50 \\ 0.50 \end{pmatrix}$

Supplementary Table 2: ICL-BIC and ROC AUC results for joint model fits to simulated data. Compare to Table 1 in the main paper, particularly subtable (c). ICL-BIC is strongly biased toward the chained model and is much less likely than BIC to select the model with the highest ROC AUC.

(a) Mean (and standard error) of ICL-BIC for layered and chained fits to data generated from layered and chained topologies. Higher (less negative) values are preferred. The "difference" row shows the difference between ICL-BICs for correct vs. incorrect fit topologies.

	layered gen.	chained gen.
layered fit	-170030.2(187.5)	-155617.0(135.7)
chained fit	-166421.1(132.5)	-154650.1(87.1)
difference	-3609.0(94.5)	966.9(107.3)

(b) Mean (and standard error) of ROC AUC for layered and chained fits, joint fit selected by ICL-BIC, and marginal fits to data generated from layered and chained topologies.

	layered gen.	chained gen.
layered fit	0.881(0.0007)	0.854(0.0009)
chained fit	0.838(0.0009)	0.877(0.0008)
selected fit	0.839(0.0010)	0.877(0.0008)

(c) Proportions of correct choices for layered and chained fits to data generated from the corresponding topologies (ICL-BIC, ROC AUC) and of the fit selected by ICL-BIC being best by ROC AUC ("conditional").

	ICL-BIC	ROC AUC	conditional
layered fit	0.017	1.000	0.600
chained fit	0.983	0.993	0.498

Supplementary Table 3: Results of the marginal model selection procedure by ICL-BIC for Ci and Dl data. "K_{sel}" denotes the selected number of components while "family" denotes the selected distribution family.

data	normal, $K = 2$	normal, $K = 3$	K_{sel}	PVII, $K = K_{sel}$	family
Ci binding	-41125.58	-42961.09	2	-40367.84	PVII
Ci expression	-102701.05	-99883.38	3	-102939.13	normal
Ci conservation	-38834.12	-45025.59	2	-34740.74	PVII
Dl binding	-62318.74	-71849.10	2	-65808.63	normal
Dl expression	-55081.15	-60665.64	2	-62649.57	normal
Dl conservation	-49795.43	-52554.07	2	-43768.86	PVII

Supplementary Table 4: Results of the joint model selection procedure by BIC for Ci and Dl data, with the marginal model specifications selected by ICL-BIC as shown in Supplementary Table 3.

data	layered	chained	selected
Ci	-152248.5	-152243.3	chained
Dl	-144180.1	-144282.1	layered

Supplementary Table 5: Fitted model parameters for data types modeled as marginally PVII, from chained model for Ci and layered model for Dl.

data	dimension	number of components	component probabilities	component means	component scales	component shapes
Ci binding	1	2	$\begin{pmatrix} 0.303 \\ 0.697 \end{pmatrix}$	$\begin{pmatrix} 0.650 \\ -0.280 \end{pmatrix}$	$\begin{pmatrix} 2.412 \\ 1.591 \end{pmatrix}$	$\begin{pmatrix} 3.447 \\ 3.062 \end{pmatrix}$
Ci conservation	1	2	$\begin{pmatrix} 0.654 \\ 0.346 \end{pmatrix}$	$\begin{pmatrix} 0.458 \\ -0.859 \end{pmatrix}$	$\begin{pmatrix} 1.854 \\ 4.022 \end{pmatrix}$	$\begin{pmatrix} 4.740 \\ 5.562 \end{pmatrix}$
Dl conservation	1	2	$\begin{pmatrix} 0.636 \\ 0.364 \end{pmatrix}$	$\begin{pmatrix} 0.490 \\ -0.851 \end{pmatrix}$	$\begin{pmatrix} 1.908 \\ 4.551 \end{pmatrix}$	$\begin{pmatrix} 5.167 \\ 6.289 \end{pmatrix}$

Supplementary Table 6: Fitted model parameters for data types modeled as marginally normal, from chained model for Ci and layered model for Dl.

data	dimension	number of components	component probabilities	component means	component variances
Ci expression	4	3	$\begin{pmatrix} 0.136 \\ 0.621 \\ 0.243 \end{pmatrix}$	$\begin{array}{cccc} smo & ptc & ci & hh \\ -0.114 & 0.047 & 0.196 & 0.377 \\ -0.113 & 0.025 & 0.006 & -0.080 \\ 0.353 & -0.089 & -0.125 & -0.008 \end{array}$	$\begin{array}{c|cccc} & smo & ptc & ci & hh \\ smo & 3.346 & 0.198 & -0.581 & 1.033 \\ ptc & 0.198 & 2.427 & 1.255 & 0.493 \\ ci & -0.581 & 1.255 & 4.542 & 0.355 \\ hh & 1.033 & 0.493 & 0.355 & 2.957 \end{array}$ $\begin{array}{c|cccc} & smo & ptc & ci & hh \\ smo & 0.397 & 0.133 & -0.005 & 0.216 \\ ptc & 0.133 & 1.015 & 0.395 & 0.151 \\ ci & -0.005 & 0.395 & 0.577 & 0.163 \\ hh & 0.216 & 0.151 & 0.163 & 0.848 \end{array}$ $\begin{array}{c|cccc} & smo & ptc & ci & hh \\ smo & 1.061 & 0.102 & -0.216 & 0.393 \\ ptc & 0.102 & 0.152 & 0.025 & 0.044 \\ ci & -0.216 & 0.025 & 0.062 & -0.073 \\ hh & 0.393 & 0.044 & -0.073 & 0.198 \end{array}$
Dl binding	2	2	$\begin{pmatrix} 0.673 \\ 0.327 \end{pmatrix}$	$\begin{array}{cc} dorsal & snail \\ 0.294 & 0.319 \\ -0.606 & -0.657 \end{array}$	$\begin{array}{c|cc} & dorsal & snail \\ dorsal & 0.670 & 0.583 \\ snail & 0.583 & 0.637 \end{array}$ $\begin{array}{c|cc} & dorsal & snail \\ dorsal & 1.133 & 0.879 \\ snail & 0.879 & 1.104 \end{array}$
Dl expression	2	2	$\begin{pmatrix} 0.162 \\ 0.838 \end{pmatrix}$	$\begin{array}{cc} tenb & rmine \\ 0.391 & 0.361 \\ -0.076 & -0.070 \end{array}$	$\begin{array}{c|cc} & tenb & rmine \\ tenb & 5.031 & 2.221 \\ rmine & 2.221 & 4.910 \end{array}$ $\begin{array}{c|cc} & tenb & rmine \\ tenb & 0.185 & 0.070 \\ rmine & 0.070 & 0.214 \end{array}$

Supplementary Table 7: Posterior probability cutoff (and corresponding fdr) for top 200 genes identified by joint and marginal models for each data set.

	joint	binding	expression	conservation
Ci	0.808(0.192)	0.858(0.142)	1.000(0.000)	0.974(0.026)
Dl	0.985(0.015)	0.949(0.051)	1.000(0.000)	0.991(0.009)

Supplementary Table 8: Model ranks for all KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2012) Hedgehog pathway genes, for Ci data set. Boldface indicates ranks ≤ 200. Some genes, such as *wntd*, *ci*, *pka-c3*, and *su(fu)*, are highly ranked by expression, but not by the joint model. These are genes whose expression appears to regulated by one or more of Smo, Ptc, Ci, and Hh, but which are neither direct Ci binding targets nor highly conserved. The expression-regulation pathways for these genes may involve several intermediate steps, which cannot be elucidated from the current data sets.

gene symbol	joint	binding	expression	conservation
hh	**12**	**79**	**111**	2461
wg	**40**	407	**99**	754
smo	**71**	570	**182**	1450
wntd	**127**	848	339	3287
ckiα	228	437	1420	402
ptc	468	517	6145	219
wnt6	631	683	2573	4018
cos	746	765	6515	4459
gbb	2155	2227	5225	3528
ci	2326	3917	**23**	3076
pka-c1	2782	3354	1444	1723
rab23	3200	3384	4163	1225
sgg	3668	3848	5322	1387
gish	4021	4621	1589	426
fu	4214	4398	6051	1139
pka-c2	4596	2932	7910	5134
dco	5375	5666	3486	886
pka-c3	5425	9170	359	5312
su(fu)	5695	9813	225	5843
dpp	6972	10244	1122	3224
slmb	7403	8057	3126	802
wnt5	8129	9260	2922	2971
wnt10	9710	8951	9125	1399

Supplementary Table 9: Top 50 GO FAT BP (Huang et al., 2009a,b) terms from the top 200 genes from the joint model for all Ci data. The "count" column gives the number of genes annotated with the term. Bold IDs indicate "overenriched" terms, that is, terms which are not present in the marginal lists, or are more strongly enriched in the joint lists than in the marginal lists.

ID	name	*p*-value	count
GO:0001708	cell fate specification	4.063e-12	15
GO:0045165	cell fate commitment	4.451e-12	25
GO:0007389	pattern specification process	5.315e-12	35
GO:0001709	cell fate determination	1.764e-11	19
GO:0003002	regionalization	2.913e-11	33
GO:0045449	regulation of transcription	4.620e-11	44
GO:0006355	regulation of transcription, DNA-dependent	5.632e-11	38
GO:0051252	regulation of RNA metabolic process	1.058e-09	38
GO:0007423	sensory organ development	5.877e-09	28
GO:0035218	leg disc development	2.332e-08	11
GO:0007365	periodic partitioning	4.165e-08	11
GO:0007560	imaginal disc morphogenesis	4.951e-08	22
GO:0048563	post-embryonic organ morphogenesis	4.951e-08	22
GO:0007422	peripheral nervous system development	6.373e-08	13
GO:0006357	regulation of transcription from RNA polymerase II . . .	7.833e-08	18
GO:0048707	instar larval or pupal morphogenesis	8.164e-08	24
GO:0048569	post-embryonic organ development	1.070e-07	22
GO:0009886	post-embryonic morphogenesis	1.117e-07	24
GO:0007424	open tracheal system development	1.255e-07	16
GO:0060541	respiratory system development	1.255e-07	16
GO:0007552	metamorphosis	1.765e-07	24
GO:0007444	imaginal disc development	4.618e-07	25
GO:0045596	negative regulation of cell differentiation	1.035e-06	11
GO:0048859	formation of anatomical boundary	1.154e-06	9
GO:0045941	positive regulation of transcription	1.369e-06	13
GO:0007447	imaginal disc pattern formation	1.384e-06	12
GO:0010628	positive regulation of gene expression	1.501e-06	13
GO:0007449	proximal/distal pattern formation, imaginal disc	1.572e-06	7
GO:0035114	imaginal disc-derived appendage morphogenesis	1.785e-06	18
GO:0035107	appendage morphogenesis	2.120e-06	18
GO:0051173	positive regulation of nitrogen compound metabolic . . .	2.152e-06	13
GO:0045935	positive regulation of nucleobase, nucleoside, . . .	2.152e-06	13
GO:0002165	instar larval or pupal development	2.218e-06	24
GO:0048737	imaginal disc-derived appendage development	2.243e-06	18
GO:0048736	appendage development	2.654e-06	18
GO:0009954	proximal/distal pattern formation	2.837e-06	7
GO:0045944	positive regulation of transcription from RNA . . .	3.181e-06	9
GO:0007155	cell adhesion	3.699e-06	15
GO:0009791	post-embryonic development	4.191e-06	24
GO:0010557	positive regulation of macromolecule biosynthetic . . .	4.236e-06	13
GO:0031328	positive regulation of cellular biosynthetic . . .	4.587e-06	14
GO:0009891	positive regulation of biosynthetic process	4.587e-06	14
GO:0045893	positive regulation of transcription, . . .	4.590e-06	11
GO:0007400	neuroblast fate determination	4.827e-06	7
GO:0010160	formation of organ boundary	5.055e-06	8
GO:0009880	embryonic pattern specification	5.177e-06	17
GO:0051301	cell division	5.328e-06	16
GO:0035282	segmentation	5.620e-06	18
GO:0051254	positive regulation of RNA metabolic process	5.636e-06	11
GO:0048645	organ formation	7.240e-06	8

Supplementary Table 10: Top 50 GO FAT BP terms from the top 200 genes from the marginal model for Ci binding data.

ID	name	*p*-value	count
GO:0006355	regulation of transcription, DNA-dependent	1.149e-23	51
GO:0045449	regulation of transcription	1.680e-23	57
GO:0051252	regulation of RNA metabolic process	1.162e-21	51
GO:0045165	cell fate commitment	2.073e-17	29
GO:0048598	embryonic morphogenesis	1.170e-14	26
GO:0006350	transcription	1.195e-13	34
GO:0001709	cell fate determination	1.322e-13	20
GO:0030182	neuron differentiation	1.889e-13	32
GO:0006357	regulation of transcription from RNA polymerase II . . .	2.668e-13	23
GO:0007498	mesoderm development	7.603e-13	17
GO:0007366	periodic partitioning by pair rule gene	4.418e-12	9
GO:0048666	neuron development	4.899e-12	28
GO:0048729	tissue morphogenesis	5.969e-12	24
GO:0048812	neuron projection morphogenesis	1.828e-11	25
GO:0031175	neuron projection development	1.969e-11	25
GO:0048667	cell morphogenesis involved in neuron . . .	2.120e-11	25
GO:0048858	cell projection morphogenesis	3.142e-11	26
GO:0000902	cell morphogenesis	4.762e-11	30
GO:0007507	heart development	4.848e-11	14
GO:0000904	cell morphogenesis involved in differentiation	6.198e-11	25
GO:0032990	cell part morphogenesis	6.553e-11	26
GO:0007389	pattern specification process	6.920e-11	31
GO:0006928	cell motion	7.116e-11	25
GO:0007369	gastrulation	8.538e-11	14
GO:0001708	cell fate specification	2.005e-10	13
GO:0003002	regionalization	4.632e-10	29
GO:0030030	cell projection organization	5.301e-10	26
GO:0007365	periodic partitioning	6.617e-10	12
GO:0035289	posterior head segmentation	1.391e-09	8
GO:0001710	mesodermal cell fate commitment	1.391e-09	8
GO:0048333	mesodermal cell differentiation	1.391e-09	8
GO:0032989	cellular component morphogenesis	2.066e-09	30
GO:0016477	cell migration	2.319e-09	18
GO:0000122	negative regulation of transcription from RNA . . .	2.627e-09	12
GO:0009880	embryonic pattern specification	4.100e-09	20
GO:0001707	mesoderm formation	4.498e-09	9
GO:0048732	gland development	6.126e-09	17
GO:0048332	mesoderm morphogenesis	6.305e-09	9
GO:0048870	cell motility	6.425e-09	18
GO:0051674	localization of cell	1.311e-08	18
GO:0045893	positive regulation of transcription, . . .	1.338e-08	13
GO:0001704	formation of primary germ layer	1.596e-08	9
GO:0051254	positive regulation of RNA metabolic process	1.736e-08	13
GO:0035287	head segmentation	2.319e-08	8
GO:0007444	imaginal disc development	3.046e-08	25
GO:0045941	positive regulation of transcription	3.640e-08	14
GO:0048813	dendrite morphogenesis	4.039e-08	14
GO:0010628	positive regulation of gene expression	4.039e-08	14
GO:0016358	dendrite development	4.039e-08	14
GO:0035295	tube development	5.182e-08	13

Supplementary Table 11: Top 50 GO FAT BP terms from the top 200 genes from the marginal model for Ci expression data.

ID	name	*p*-value	count
GO:0035161	imaginal disc lineage restriction	5.934e-06	5
GO:0007367	segment polarity determination	1.123e-05	7
GO:0048100	wing disc anterior/posterior pattern formation	1.959e-05	5
GO:0007447	imaginal disc pattern formation	3.892e-05	9
GO:0007365	periodic partitioning	7.558e-05	7
GO:0035224	genital disc anterior/posterior pattern formation	1.235e-04	4
GO:0035221	genital disc pattern formation	1.235e-04	4
GO:0007448	anterior/posterior pattern formation, imaginal ...	1.235e-04	5
GO:0009880	embryonic pattern specification	2.437e-04	12
GO:0035222	wing disc pattern formation	2.754e-04	7
GO:0007350	blastoderm segmentation	6.334e-04	11
GO:0035282	segmentation	6.667e-04	12
GO:0007224	smoothened signaling pathway	9.129e-04	5
GO:0003002	regionalization	9.608e-04	16
GO:0035217	labial disc development	1.042e-03	3
GO:0048099	anterior/posterior lineage restriction, imaginal ...	1.042e-03	3
GO:0007389	pattern specification process	1.681e-03	16
GO:0035277	spiracle morphogenesis, open tracheal system	1.920e-03	4
GO:0016055	Wnt receptor signaling pathway	2.502e-03	6
GO:0007449	proximal/distal pattern formation, imaginal disc	2.585e-03	4
GO:0007446	imaginal disc growth	2.964e-03	4
GO:0035265	organ growth	3.376e-03	4
GO:0009954	proximal/distal pattern formation	3.376e-03	4
GO:0007473	wing disc proximal/distal pattern formation	3.552e-03	3
GO:0007424	open tracheal system development	4.272e-03	8
GO:0060541	respiratory system development	4.272e-03	8
GO:0045168	cell-cell signaling involved in cell fate ...	4.694e-03	3
GO:0035288	anterior head segmentation	4.694e-03	3
GO:0007418	ventral midline development	4.694e-03	3
GO:0000902	cell morphogenesis	5.782e-03	14
GO:0030707	ovarian follicle cell development	5.814e-03	9
GO:0035215	genital disc development	5.951e-03	4
GO:0003006	reproductive developmental process	7.138e-03	15
GO:0009952	anterior/posterior pattern formation	7.538e-03	8
GO:0055114	oxidation reduction	7.810e-03	17
GO:0007444	imaginal disc development	8.586e-03	13
GO:0007166	cell surface receptor linked signal transduction	8.793e-03	17
GO:0030030	cell projection organization	9.245e-03	12
GO:0001708	cell fate specification	9.263e-03	5
GO:0007411	axon guidance	9.544e-03	7
GO:0048858	cell projection morphogenesis	1.019e-02	11
GO:0045165	cell fate commitment	1.219e-02	9
GO:0044271	nitrogen compound biosynthetic process	1.249e-02	9
GO:0007090	regulation of S phase of mitotic cell cycle	1.252e-02	3
GO:0007440	foregut morphogenesis	1.252e-02	3
GO:0032990	cell part morphogenesis	1.272e-02	11
GO:0002164	larval development	1.273e-02	5
GO:0001654	eye development	1.297e-02	11
GO:0048812	neuron projection morphogenesis	1.399e-02	10
GO:0031175	neuron projection development	1.429e-02	10

11

Supplementary Table 12: Top 50 GO FAT BP terms from the top 200 genes from the marginal model for Ci conservation data.

ID	name	*p*-value	count
GO:0030182	neuron differentiation	1.963e-06	25
GO:0046530	photoreceptor cell differentiation	9.391e-06	13
GO:0048666	neuron development	1.988e-05	21
GO:0002165	instar larval or pupal development	3.055e-05	23
GO:0009791	post-embryonic development	5.401e-05	23
GO:0048707	instar larval or pupal morphogenesis	6.557e-05	20
GO:0000902	cell morphogenesis	7.096e-05	23
GO:0009886	post-embryonic morphogenesis	8.253e-05	20
GO:0032989	cellular component morphogenesis	1.005e-04	25
GO:0001754	eye photoreceptor cell differentiation	1.089e-04	11
GO:0007552	metamorphosis	1.153e-04	20
GO:0000904	cell morphogenesis involved in differentiation	1.210e-04	18
GO:0006350	transcription	1.236e-04	23
GO:0007560	imaginal disc morphogenesis	1.844e-04	17
GO:0048563	post-embryonic organ morphogenesis	1.844e-04	17
GO:0007423	sensory organ development	2.039e-04	21
GO:0048667	cell morphogenesis involved in neuron ...	2.164e-04	17
GO:0000278	mitotic cell cycle	2.995e-04	19
GO:0048569	post-embryonic organ development	3.066e-04	17
GO:0048592	eye morphogenesis	3.295e-04	16
GO:0042063	gliogenesis	3.722e-04	7
GO:0001751	compound eye photoreceptor cell differentiation	3.832e-04	10
GO:0007010	cytoskeleton organization	4.016e-04	22
GO:0006796	phosphate metabolic process	6.396e-04	24
GO:0006793	phosphorus metabolic process	6.396e-04	24
GO:0016310	phosphorylation	8.710e-04	20
GO:0000022	mitotic spindle elongation	9.439e-04	8
GO:0051231	spindle elongation	1.018e-03	8
GO:0030030	cell projection organization	1.052e-03	18
GO:0060284	regulation of cell development	1.273e-03	10
GO:0007049	cell cycle	1.277e-03	25
GO:0007017	microtubule-based process	1.315e-03	19
GO:0001745	compound eye morphogenesis	1.570e-03	14
GO:0042461	photoreceptor cell development	1.608e-03	7
GO:0048812	neuron projection morphogenesis	1.847e-03	15
GO:0031175	neuron projection development	1.909e-03	15
GO:0048598	embryonic morphogenesis	1.991e-03	13
GO:0006468	protein amino acid phosphorylation	2.151e-03	14
GO:0051252	regulation of RNA metabolic process	2.166e-03	26
GO:0048813	dendrite morphogenesis	2.432e-03	9
GO:0016358	dendrite development	2.432e-03	9
GO:0000226	microtubule cytoskeleton organization	2.701e-03	15
GO:0001654	eye development	2.789e-03	16
GO:0007444	imaginal disc development	3.794e-03	18
GO:0003002	regionalization	4.430e-03	19
GO:0048858	cell projection morphogenesis	5.106e-03	15
GO:0007051	spindle organization	5.688e-03	12
GO:0007052	mitotic spindle organization	5.747e-03	11
GO:0032990	cell part morphogenesis	6.838e-03	15
GO:0035272	exocrine system development	6.927e-03	9

Supplementary Table 13: Top 50 GO FAT BP terms from genes with fdr ≤ 0.2 from the joint model for all Ci data. 234 genes are represented.

ID	name	*p*-value	count
GO:0045165	cell fate commitment	5.096e-14	29
GO:0001708	cell fate specification	1.122e-13	17
GO:0007389	pattern specification process	1.795e-13	40
GO:0003002	regionalization	7.776e-13	38
GO:0006355	regulation of transcription, DNA-dependent	5.265e-12	43
GO:0045449	regulation of transcription	1.370e-11	49
GO:0051252	regulation of RNA metabolic process	1.448e-10	43
GO:0001709	cell fate determination	2.060e-10	19
GO:0006357	regulation of transcription from RNA polymerase II . . .	3.866e-09	21
GO:0048563	post-embryonic organ morphogenesis	6.688e-09	25
GO:0007560	imaginal disc morphogenesis	6.688e-09	25
GO:0007423	sensory organ development	9.243e-09	30
GO:0007365	periodic partitioning	1.358e-08	12
GO:0048569	post-embryonic organ development	1.615e-08	25
GO:0048707	instar larval or pupal morphogenesis	1.845e-08	27
GO:0009886	post-embryonic morphogenesis	2.627e-08	27
GO:0007422	peripheral nervous system development	3.754e-08	14
GO:0007552	metamorphosis	4.395e-08	27
GO:0035218	leg disc development	9.337e-08	11
GO:0007447	imaginal disc pattern formation	1.088e-07	14
GO:0007424	open tracheal system development	1.453e-07	17
GO:0060541	respiratory system development	1.453e-07	17
GO:0007444	imaginal disc development	1.566e-07	28
GO:0035114	imaginal disc-derived appendage morphogenesis	1.593e-07	21
GO:0035107	appendage morphogenesis	1.954e-07	21
GO:0048737	imaginal disc-derived appendage development	2.090e-07	21
GO:0048736	appendage development	2.553e-07	21
GO:0048859	formation of anatomical boundary	2.859e-07	10
GO:0045596	negative regulation of cell differentiation	4.707e-07	12
GO:0002165	instar larval or pupal development	7.537e-07	27
GO:0007155	cell adhesion	8.519e-07	17
GO:0045944	positive regulation of transcription from RNA . . .	9.048e-07	10
GO:0045941	positive regulation of transcription	1.022e-06	14
GO:0010160	formation of organ boundary	1.062e-06	9
GO:0010628	positive regulation of gene expression	1.129e-06	14
GO:0009791	post-embryonic development	1.538e-06	27
GO:0048645	organ formation	1.611e-06	9
GO:0051173	positive regulation of nitrogen compound metabolic . . .	1.661e-06	14
GO:0045935	positive regulation of nucleobase, nucleoside, . . .	1.661e-06	14
GO:0048732	gland development	1.888e-06	16
GO:0009880	embryonic pattern specification	1.961e-06	19
GO:0022610	biological adhesion	2.392e-06	17
GO:0045893	positive regulation of transcription, . . .	2.418e-06	12
GO:0035282	segmentation	2.622e-06	20
GO:0051254	positive regulation of RNA metabolic process	3.028e-06	12
GO:0010557	positive regulation of macromolecule biosynthetic . . .	3.436e-06	14
GO:0007449	proximal/distal pattern formation, imaginal disc	3.673e-06	7
GO:0031328	positive regulation of cellular biosynthetic . . .	4.515e-06	15
GO:0009891	positive regulation of biosynthetic process	4.515e-06	15
GO:0035214	eye-antennal disc development	5.877e-06	10

Supplementary Table 14: KEGG pathways from selected genes from Ci gene lists.

(a) From top 200 genes from the joint model for all data.

ID	name	*p*-value	count
dme04340	Hedgehog signaling pathway	4.629e-03	4
dme00600	Sphingolipid metabolism	5.026e-02	3
dme04320	Dorso-ventral axis formation	5.396e-02	3

(b) From top 200 genes from the marginal model for binding data.

ID	name	*p*-value	count
dme04320	Dorso-ventral axis formation	3.048e-02	3

(c) From top 200 genes from the marginal model for expression data.

ID	name	*p*-value	count
dme04340	Hedgehog signaling pathway	1.194e-02	4
dme00260	Glycine, serine and threonine metabolism	9.108e-02	3
dme00903	Limonene and pinene degradation	9.588e-02	5

(d) From top 200 genes from the marginal model for conservation data.

ID	name	*p*-value	count
dme03010	Ribosome	7.142e-03	8

(e) From genes with fdr ≤ 0.2 from the joint model for all data.

ID	name	*p*-value	count
dme04340	Hedgehog signaling pathway	8.124e-04	5
dme04320	Dorso-ventral axis formation	1.133e-02	4
dme00600	Sphingolipid metabolism	7.658e-02	3
dme00250	Alanine, aspartate and glutamate metabolism	9.324e-02	3

Supplementary Table 15: Top 50 GO FAT BP terms from the top 200 genes from the joint model for all Dl data. Note that all terms in the top 50 are overenriched compared to the marginal lists.

ID	name	*p*-value	count
GO:0007389	pattern specification process	3.604e-31	59
GO:0003002	regionalization	1.603e-27	54
GO:0007444	imaginal disc development	8.535e-24	48
GO:0045165	cell fate commitment	1.051e-22	37
GO:0007424	open tracheal system development	5.193e-22	31
GO:0060541	respiratory system development	5.193e-22	31
GO:0048729	tissue morphogenesis	9.462e-22	37
GO:0032989	cellular component morphogenesis	2.171e-20	49
GO:0048598	embryonic morphogenesis	5.630e-20	34
GO:0007423	sensory organ development	2.153e-19	43
GO:0000902	cell morphogenesis	5.311e-19	44
GO:0048563	post-embryonic organ morphogenesis	1.002e-18	36
GO:0007560	imaginal disc morphogenesis	1.002e-18	36
GO:0030182	neuron differentiation	1.407e-18	42
GO:0060429	epithelium development	1.885e-18	33
GO:0009886	post-embryonic morphogenesis	2.865e-18	39
GO:0002165	instar larval or pupal development	3.120e-18	42
GO:0048569	post-embryonic organ development	4.287e-18	36
GO:0002009	morphogenesis of an epithelium	4.880e-18	32
GO:0048707	instar larval or pupal morphogenesis	1.189e-17	38
GO:0009791	post-embryonic development	1.221e-17	42
GO:0045449	regulation of transcription	1.811e-17	56
GO:0007350	blastoderm segmentation	1.857e-17	31
GO:0001654	eye development	2.047e-17	37
GO:0007552	metamorphosis	4.862e-17	38
GO:0035220	wing disc development	4.954e-17	34
GO:0051252	regulation of RNA metabolic process	5.152e-17	51
GO:0006357	regulation of transcription from RNA polymerase II …	5.234e-17	29
GO:0009880	embryonic pattern specification	9.591e-17	31
GO:0006355	regulation of transcription, DNA-dependent	1.218e-16	48
GO:0035282	segmentation	3.251e-16	32
GO:0001709	cell fate determination	4.460e-16	24
GO:0048736	appendage development	6.680e-16	31
GO:0035120	post-embryonic appendage morphogenesis	7.674e-16	30
GO:0048732	gland development	9.690e-16	26
GO:0035272	exocrine system development	2.665e-15	24
GO:0007431	salivary gland development	2.665e-15	24
GO:0035107	appendage morphogenesis	3.338e-15	30
GO:0048737	imaginal disc-derived appendage development	3.675e-15	30
GO:0000904	cell morphogenesis involved in differentiation	4.005e-15	33
GO:0048666	neuron development	4.653e-15	35
GO:0006928	cell motion	4.832e-15	33
GO:0048667	cell morphogenesis involved in neuron …	6.721e-15	32
GO:0035114	imaginal disc-derived appendage morphogenesis	1.879e-14	29
GO:0030030	cell projection organization	2.129e-14	35
GO:0048812	neuron projection morphogenesis	3.946e-14	31
GO:0031175	neuron projection development	4.334e-14	31
GO:0007365	periodic partitioning	7.617e-14	16
GO:0035152	regulation of tube architecture, open tracheal …	1.795e-13	14
GO:0032990	cell part morphogenesis	3.135e-13	32

Supplementary Table 16: Top 50 GO FAT BP terms from the top 200 genes from the marginal model for Dl binding data.

ID	name	*p*-value	count
GO:0045165	cell fate commitment	9.753e-05	13
GO:0007444	imaginal disc development	6.347e-04	16
GO:0007460	R8 cell fate commitment	8.890e-04	4
GO:0007389	pattern specification process	1.075e-03	17
GO:0006357	regulation of transcription from RNA polymerase II . . .	1.511e-03	10
GO:0045465	R8 cell differentiation	1.610e-03	4
GO:0001709	cell fate determination	2.243e-03	8
GO:0010557	positive regulation of macromolecule biosynthetic . . .	2.243e-03	8
GO:0007447	imaginal disc pattern formation	2.285e-03	7
GO:0001700	embryonic development via the syncytial blastoderm	3.164e-03	10
GO:0033554	cellular response to stress	3.340e-03	9
GO:0007616	long-term memory	3.465e-03	4
GO:0010604	positive regulation of macromolecule metabolic . . .	3.842e-03	8
GO:0009792	embryonic development ending in birth or egg hatching . . .	4.280e-03	10
GO:0003002	regionalization	4.543e-03	15
GO:0048190	wing disc dorsal/ventral pattern formation	5.253e-03	5
GO:0009891	positive regulation of biosynthetic process	5.773e-03	8
GO:0060541	respiratory system development	5.773e-03	8
GO:0007424	open tracheal system development	5.773e-03	8
GO:0031328	positive regulation of cellular biosynthetic . . .	5.773e-03	8
GO:0060429	epithelium development	5.841e-03	10
GO:0051252	regulation of RNA metabolic process	6.384e-03	19
GO:0048729	tissue morphogenesis	8.012e-03	10
GO:0048732	gland development	8.088e-03	8
GO:0007450	dorsal/ventral pattern formation, imaginal disc	8.358e-03	5
GO:0042127	regulation of cell proliferation	8.358e-03	5
GO:0048477	oogenesis	8.463e-03	16
GO:0051726	regulation of cell cycle	8.901e-03	8
GO:0000902	cell morphogenesis	9.132e-03	14
GO:0007419	ventral cord development	9.225e-03	4
GO:0007292	female gamete generation	9.478e-03	16
GO:0010629	negative regulation of gene expression	1.001e-02	10
GO:0060284	regulation of cell development	1.037e-02	7
GO:0001736	establishment of planar polarity	1.058e-02	5
GO:0006928	cell motion	1.074e-02	11
GO:0007391	dorsal closure	1.084e-02	6
GO:0001708	cell fate specification	1.119e-02	5
GO:0007164	establishment of tissue polarity	1.119e-02	5
GO:0006355	regulation of transcription, DNA-dependent	1.163e-02	17
GO:0007346	regulation of mitotic cell cycle	1.229e-02	6
GO:0006096	glycolysis	1.290e-02	4
GO:0010605	negative regulation of macromolecule metabolic . . .	1.381e-02	11
GO:0007298	border follicle cell migration	1.385e-02	5
GO:0032989	cellular component morphogenesis	1.385e-02	15
GO:0007366	periodic partitioning by pair rule gene	1.391e-02	3
GO:0007613	memory	1.393e-02	4
GO:0046843	dorsal appendage formation	1.393e-02	4
GO:0002009	morphogenesis of an epithelium	1.401e-02	9
GO:0048598	embryonic morphogenesis	1.436e-02	9
GO:0002165	instar larval or pupal development	1.450e-02	13

Supplementary Table 17: Top 50 GO FAT BP terms from the top 200 genes from the marginal model for Dl expression data.

ID	name	*p*-value	count
GO:0048729	tissue morphogenesis	5.251e-21	34
GO:0007507	heart development	4.074e-17	19
GO:0001708	cell fate specification	3.137e-15	17
GO:0002009	morphogenesis of an epithelium	5.149e-15	27
GO:0007498	mesoderm development	8.450e-15	19
GO:0060429	epithelium development	1.484e-14	27
GO:0007389	pattern specification process	8.555e-14	36
GO:0045165	cell fate commitment	9.389e-14	26
GO:0048598	embryonic morphogenesis	4.317e-13	25
GO:0003002	regionalization	3.157e-12	33
GO:0006928	cell motion	6.954e-12	27
GO:0035295	tube development	7.660e-12	17
GO:0048332	mesoderm morphogenesis	1.030e-11	11
GO:0007369	gastrulation	1.227e-11	15
GO:0035051	cardiac cell differentiation	1.397e-11	10
GO:0001704	formation of primary germ layer	3.511e-11	11
GO:0001709	cell fate determination	4.816e-11	18
GO:0016477	cell migration	1.021e-10	20
GO:0035239	tube morphogenesis	1.095e-10	15
GO:0007444	imaginal disc development	2.178e-10	29
GO:0048565	gut development	2.212e-10	14
GO:0001707	mesoderm formation	2.383e-10	10
GO:0048732	gland development	2.494e-10	19
GO:0048870	cell motility	3.242e-10	20
GO:0051674	localization of cell	7.282e-10	20
GO:0048569	post-embryonic organ development	9.052e-10	24
GO:0007422	peripheral nervous system development	2.474e-09	14
GO:0009791	post-embryonic development	3.419e-09	28
GO:0035107	appendage morphogenesis	4.187e-09	21
GO:0048736	appendage development	5.563e-09	21
GO:0007552	metamorphosis	8.462e-09	25
GO:0035272	exocrine system development	1.189e-08	16
GO:0007431	salivary gland development	1.189e-08	16
GO:0035114	imaginal disc-derived appendage morphogenesis	2.009e-08	20
GO:0009886	post-embryonic morphogenesis	2.507e-08	24
GO:0048737	imaginal disc-derived appendage development	2.633e-08	20
GO:0007442	hindgut morphogenesis	2.696e-08	10
GO:0010002	cardioblast differentiation	3.524e-08	7
GO:0048859	formation of anatomical boundary	4.164e-08	10
GO:0007424	open tracheal system development	4.408e-08	16
GO:0060541	respiratory system development	4.408e-08	16
GO:0035120	post-embryonic appendage morphogenesis	5.846e-08	19
GO:0048567	ectodermal gut morphogenesis	6.286e-08	10
GO:0007439	ectodermal gut development	6.286e-08	10
GO:0048563	post-embryonic organ morphogenesis	6.362e-08	21
GO:0007560	imaginal disc morphogenesis	6.362e-08	21
GO:0007447	imaginal disc pattern formation	7.205e-08	13
GO:0007513	pericardial cell differentiation	7.376e-08	6
GO:0008586	imaginal disc-derived wing vein morphogenesis	7.542e-08	9
GO:0048547	gut morphogenesis	7.663e-08	10

Supplementary Table 18: Top 50 GO FAT BP terms from the top 200 genes from the marginal model for Dl conservation data.

ID	name	*p*-value	count
GO:0006816	calcium ion transport	3.096e-06	8
GO:0045165	cell fate commitment	1.396e-05	18
GO:0015674	di-, tri-valent inorganic cation transport	1.652e-05	8
GO:0030182	neuron differentiation	3.273e-05	24
GO:0006812	cation transport	4.062e-05	18
GO:0016192	vesicle-mediated transport	4.887e-05	23
GO:0007267	cell-cell signaling	6.479e-05	16
GO:0030001	metal ion transport	7.104e-05	14
GO:0006811	ion transport	7.606e-05	21
GO:0007010	cytoskeleton organization	8.608e-05	25
GO:0000902	cell morphogenesis	1.083e-04	24
GO:0009792	embryonic development ending in birth or egg hatching ...	1.274e-04	16
GO:0060284	regulation of cell development	1.474e-04	12
GO:0032989	cellular component morphogenesis	1.765e-04	26
GO:0006897	endocytosis	1.852e-04	17
GO:0010324	membrane invagination	1.852e-04	17
GO:0046331	lateral inhibition	3.514e-04	4
GO:0008104	protein localization	4.057e-04	21
GO:0030030	cell projection organization	4.444e-04	20
GO:0002009	morphogenesis of an epithelium	4.563e-04	15
GO:0006886	intracellular protein transport	4.672e-04	13
GO:0006350	transcription	4.899e-04	23
GO:0045168	cell-cell signaling involved in cell fate ...	5.530e-04	4
GO:0046907	intracellular transport	5.553e-04	18
GO:0001709	cell fate determination	5.616e-04	11
GO:0034613	cellular protein localization	6.060e-04	13
GO:0051301	cell division	6.098e-04	14
GO:0048666	neuron development	6.578e-04	19
GO:0050767	regulation of neurogenesis	6.632e-04	8
GO:0016044	membrane organization	6.887e-04	18
GO:0060429	epithelium development	7.042e-04	15
GO:0007423	sensory organ development	7.185e-04	21
GO:0007444	imaginal disc development	7.636e-04	21
GO:0048511	rhythmic process	7.721e-04	7
GO:0051960	regulation of nervous system development	8.218e-04	9
GO:0007389	pattern specification process	8.933e-04	23
GO:0001700	embryonic development via the syncytial blastoderm	9.117e-04	14
GO:0000904	cell morphogenesis involved in differentiation	1.093e-03	17
GO:0048813	dendrite morphogenesis	1.119e-03	10
GO:0016358	dendrite development	1.119e-03	10
GO:0048729	tissue morphogenesis	1.144e-03	15
GO:0030036	actin cytoskeleton organization	1.250e-03	11
GO:0030029	actin filament-based process	1.318e-03	11
GO:0007560	imaginal disc morphogenesis	1.559e-03	16
GO:0048563	post-embryonic organ morphogenesis	1.559e-03	16
GO:0048812	neuron projection morphogenesis	1.671e-03	16
GO:0008356	asymmetric cell division	1.699e-03	7
GO:0031175	neuron projection development	1.730e-03	16
GO:0048667	cell morphogenesis involved in neuron ...	1.790e-03	16
GO:0046530	photoreceptor cell differentiation	2.098e-03	10

Supplementary Table 19: Top 50 GO FAT BP terms from genes with fdr ≤ 0.2 from the joint model for all Dl data. 1650 genes are represented.

ID	name	*p*-value	count
GO:0007389	pattern specification process	2.102e-37	172
GO:0045165	cell fate commitment	3.888e-37	111
GO:0003002	regionalization	2.788e-34	161
GO:0001709	cell fate determination	3.456e-33	76
GO:0048598	embryonic morphogenesis	3.611e-32	103
GO:0048729	tissue morphogenesis	5.979e-29	104
GO:0006928	cell motion	1.405e-27	116
GO:0007444	imaginal disc development	1.644e-27	140
GO:0009791	post-embryonic development	1.459e-26	143
GO:0060429	epithelium development	3.338e-26	97
GO:0002009	morphogenesis of an epithelium	3.534e-25	93
GO:0000902	cell morphogenesis	3.853e-25	142
GO:0009880	embryonic pattern specification	3.441e-24	93
GO:0032989	cellular component morphogenesis	2.213e-23	154
GO:0045449	regulation of transcription	3.547e-23	207
GO:0002165	instar larval or pupal development	4.276e-23	133
GO:0007350	blastoderm segmentation	5.393e-23	88
GO:0048569	post-embryonic organ development	1.146e-22	106
GO:0030182	neuron differentiation	5.083e-22	129
GO:0000904	cell morphogenesis involved in differentiation	6.711e-22	106
GO:0035282	segmentation	3.691e-21	95
GO:0048667	cell morphogenesis involved in neuron . . .	5.751e-21	101
GO:0009886	post-embryonic morphogenesis	6.941e-21	115
GO:0051252	regulation of RNA metabolic process	7.203e-21	179
GO:0048812	neuron projection morphogenesis	1.178e-20	100
GO:0031175	neuron projection development	1.570e-20	100
GO:0007552	metamorphosis	2.024e-20	116
GO:0016477	cell migration	4.190e-20	73
GO:0048707	instar larval or pupal morphogenesis	5.393e-20	112
GO:0048666	neuron development	5.393e-20	112
GO:0051674	localization of cell	5.537e-20	78
GO:0007560	imaginal disc morphogenesis	8.649e-20	98
GO:0048563	post-embryonic organ morphogenesis	8.649e-20	98
GO:0048870	cell motility	1.836e-19	75
GO:0030030	cell projection organization	4.162e-19	114
GO:0006355	regulation of transcription, DNA-dependent	1.436e-18	161
GO:0016331	morphogenesis of embryonic epithelium	3.448e-18	56
GO:0048732	gland development	1.932e-17	66
GO:0048858	cell projection morphogenesis	2.554e-17	101
GO:0007369	gastrulation	2.813e-17	41
GO:0048565	gut development	5.405e-17	41
GO:0048610	reproductive cellular process	5.568e-17	136
GO:0035220	wing disc development	6.118e-17	92
GO:0032990	cell part morphogenesis	1.072e-16	102
GO:0048736	appendage development	1.601e-15	82
GO:0007409	axonogenesis	2.497e-15	71
GO:0007423	sensory organ development	3.029e-15	115
GO:0060541	respiratory system development	4.730e-15	60
GO:0007424	open tracheal system development	4.730e-15	60
GO:0007391	dorsal closure	6.065e-15	45

Supplementary Table 20: KEGG pathways from selected genes from Dl gene lists.

(a) From top 200 genes from the joint model for all Dl data.

ID	name	*p*-value	count
dme04340	Hedgehog signaling pathway	5.569e-03	4
dme00010	Glycolysis / Gluconeogenesis	7.777e-03	5
dme04144	Endocytosis	3.348e-02	5

(b) From top 200 genes from the marginal model for Dl binding data.

ID	name	*p*-value	count
dme00020	Citrate cycle (TCA cycle)	4.005e-02	4
dme00620	Pyruvate metabolism	4.731e-02	4
dme00010	Glycolysis / Gluconeogenesis	6.077e-02	4

(c) From top 200 genes from the marginal model for Dl expression data.

ID	name	*p*-value	count
dme04340	Hedgehog signaling pathway	6.079e-03	4
dme04310	Wnt signaling pathway	2.880e-02	5

(d) From top 200 genes from the marginal model for Dl conservation data.

ID	name	*p*-value	count
dme04310	Wnt signaling pathway	1.504e-02	7
dme04330	Notch signaling pathway	2.146e-02	4
dme04144	Endocytosis	6.803e-02	6
dme03010	Ribosome	9.390e-02	6

(e) From genes with fdr ≤ 0.2 from the joint model for all Dl data.

ID	name	*p*-value	count
dme04340	Hedgehog signaling pathway	1.646e-03	11
dme04320	Dorso-ventral axis formation	3.346e-03	11
dme00010	Glycolysis / Gluconeogenesis	3.397e-03	17
dme00620	Pyruvate metabolism	2.043e-02	14
dme03040	Spliceosome	4.397e-02	25
dme00071	Fatty acid metabolism	4.619e-02	11

(f) From genes identified as Dorsal targets by Biemar et al. (2006). 42 genes are represented.

ID	name	*p*-value	count
dme04340	Hedgehog signaling pathway	6.817e-02	2

(g) From genes identified as Dorsal targets by Zeitlinger et al. (2007). 38 genes are represented.

ID	name	*p*-value	count
dme04340	Hedgehog signaling pathway	3.466e-02	2
dme04350	TGF-beta signaling pathway	4.887e-02	2

References

L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.

F. Biemar, D.A. Nix, J. Piel, B. Peterson, M. Ronshaugen, V. Sementchenko, I. Bell, J.R. Manak, and M.S. Levine. Comprehensive identification of Drosophila dorsal-ventral patterning genes using a whole-genome tiling array. *Proceedings of the National Academy of Sciences*, 103(34):12763–12768, 2006. doi: 10.1073/pnas.0604484103.

D.W. Huang, B.T. Sherman, and R.A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1):44–57, 2009a.

D.W. Huang, B.T. Sherman, and R.A. Lempicki. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1):1–13, 2009b.

M. Kanehisa and S. Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28:27–30, 2000.

M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M Tanabe. KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Research*, 40:D109–D114, 2012.

G.J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley, second edition, 2008.

K.M. Mullen and I.H.M. van Stokkum. *nnls: The Lawson-Hanson algorithm for non-negative least squares (NNLS)*, 2012. R package version 1.4.

J. Zeitlinger, R.P. Zinzen, A. Stark, M. Kellis, H. Zhang, R.A. Young, and M. Levine. Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes & Development*, 21(4):385–390, 2007. doi: 10.1101/gad.1509607.