Science Signaling
AAAS

www.sciencesignaling.org/cgi/content/full/5/206/ra4/DC1

Supplementary Materials for

A Wnt-Bmp Feedback Circuit Controls Intertissue Signaling Dynamics in Tooth Organogenesis

Daniel J. O'Connell, Joshua W. K. Ho, Tadanori Mammoto, Annick Turbe-Doan, Joyce T. O'Connell, Psalm S. Haseley, Samuel Koo, Nobuhiro Kamiya, Donald E. Ingber, Peter J. Park, Richard L. Maas*

*To whom correspondence should be addressed. E-mail: maas@genetics.med.harvard.edu

Published 10 January 2012, *Sci. Signal.* **5**, ra4 (2012)
DOI: 10.1126/scisignal.2002414

The PDF file includes:

Text S1. Quality assessment of microarray data

The microarray data consists of three separate experimental series: (1) Endogenous developmental time-course data (LCM and manually separated tissues); (2) *Pax9* and *Msx1* mutant mouse derived datasets; and (3) datasets from signaling molecule treatment (or "perturbation") experiments (fig. T1). To ensure high quality integrative analysis, we used a rigorous filtering and batch effect removal procedure to process the data.

Time-course

Hand separated

	E10	E11	E13
Epi	2	2	2
Mes	2	2	2

Laser capture microdissection (LCM)

	E11.5	E12	E12.5	E13	E13.5	EK
Epi	3	3	3	3	3	3
Mes	3	3	3	3	3	

Mutants

	Pax9		Msx1	
	WT	Mut	WT	Mut
Epi, E13.5	3	3	3	3
Mes, E13.5	3	3	3	3

Signaling molecule treatment

	Untreated	Endogenous signaling	GSK3i treated	Fgf8 treated	Shh treated	Bmp4 treated
Mes, E10	3	3	3	3	3	3
Epi, E13	3	3	3	3	3	3

Fig. T1. Summary of the microarray datasets used in this study. The numbers in the table represent the number of biological replicates in each biological condition.

All 105 samples were hybridized to Illumina Mouse WG-6 v2 Expression BeadChip microarrays. This microarray contains 45,281 probes for 31,508 genes. Many probes in the microarray did not generate detectable signals. Raw data were loaded and processed by the R package lumi (1). We define a probe to have a 'detectable signal' if its detection *P*-value (as determined by Illumina BeadStudio software) is less than 10^{-8} in more than 6 (out of 105) samples. By this criterion, only 20,379 probes were detectable and were used for further analysis (fig. T2).

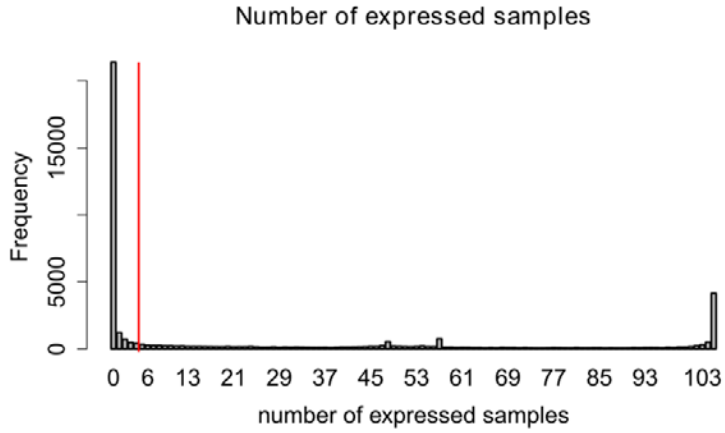

Fig. T2. Distribution of probes with different numbers of detectable samples. More than 50% of the probes in this microarray are detectable (P <10^{-8}) in more than six samples in the Illumina Mouse WG-6 v2 Expression BeadChip microarray. All non-detectable probes were removed from further analysis.

We first log2 transformed all profiles, and then applied robust spline normalization (2) implemented in lumi to each of the four experimental series independently. We further applied quantile normalization to all profiles across the four experimental series. We assessed the quality by analyzing the pairwise correlation between samples (fig. T3A). However, we noticed that quantile normalization alone was insufficient to remove a potential batch effect that was likely caused by the use of different microarray preparation protocols (cRNA or cDNA). This possible systematic difference between cRNA and cDNA in sample preparation has been reported previously (3), and is related to hybridization sensitivity and specificity. Therefore, we performed batch effect removal using an R package called ComBat (4). When using ComBat, we assumed the use of cRNA and cDNA was the only source of variation. After batch effect removal, the average correlation between all samples was much higher (fig. T3B). Notably, the two top-level clusters (as determined by hierarchical clustering) corresponded to samples from epithelium and mesenchyme. This indicated that the batch effect removal step could recover biologically relevant groupings of the profiles. We also plotted the probe intensity distribution of all the samples after batch effect removal to show that the intensity distribution is similar across all samples (fig. T4).

In the Illumina Mouse WG-6 v2 Expression BeadChip microarray, each gene is typically represented by more than one probe. To identify the most representative probe for each gene, we selected the probe with the highest median expression value across all samples as the unique probe for each gene. The final dataset contains 14,032 genes since many genes do not have probes with detectable expression. Next, we investigated whether two time-course datasets generated by different sample preparation protocols (LCM and manually separated tissues) could be integrated in a meaningful manner. We calculated the correlation between all pairs of profiles from the two time-course experiments, and visualized the results using heat maps (fig. T5A). The profile-wide Pearson correlations between samples were generally high (mostly above 0.9), regardless of whether the samples were generated using the cDNA or cRNA protocol. Also, samples were clustered by their expected biological categories (epithelium or mesenchyme, and initiation- or bud-stage, and enamel knot). Similar conclusions were reached by inspecting the gene expression patterns of the 50 most variably expressed genes (as determined by sample variance across all samples; fig. T5B). Thus, we did not observe any bias due to the use of two independently generated datasets.

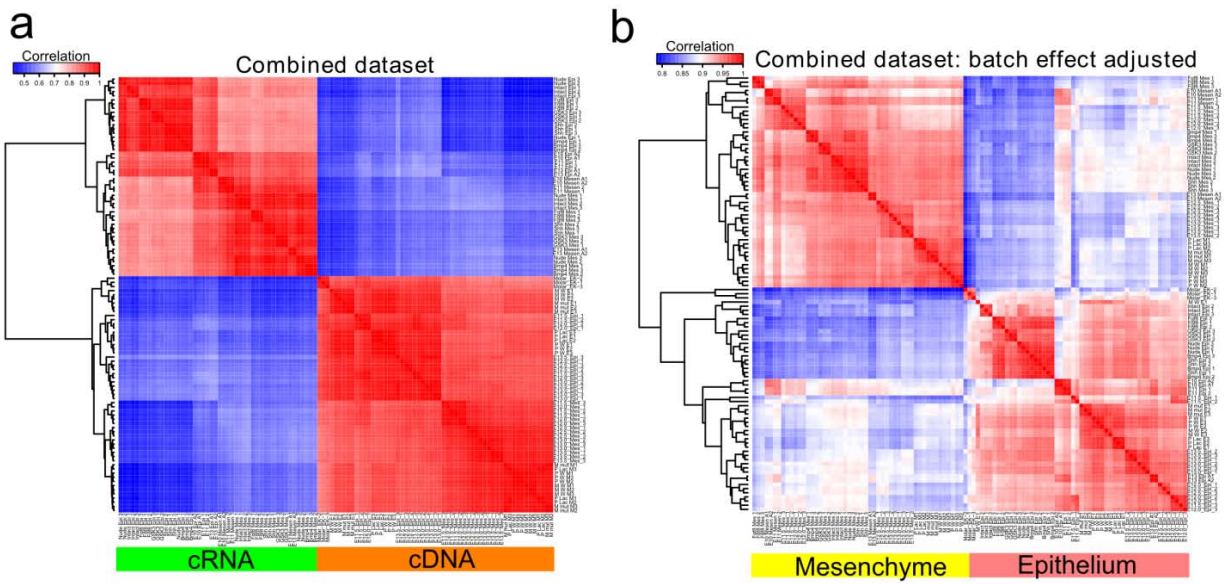

Fig. T3. Heat maps of correlation coefficients between every pair of gene expression profiles before and after batch effect removal. a, Before batch effect removal. There is a systematic difference between samples processed by cDNA and cRNA protocols. **b,** After batch effect removal. The samples cluster by tissue type and the overall correlation between samples is higher.

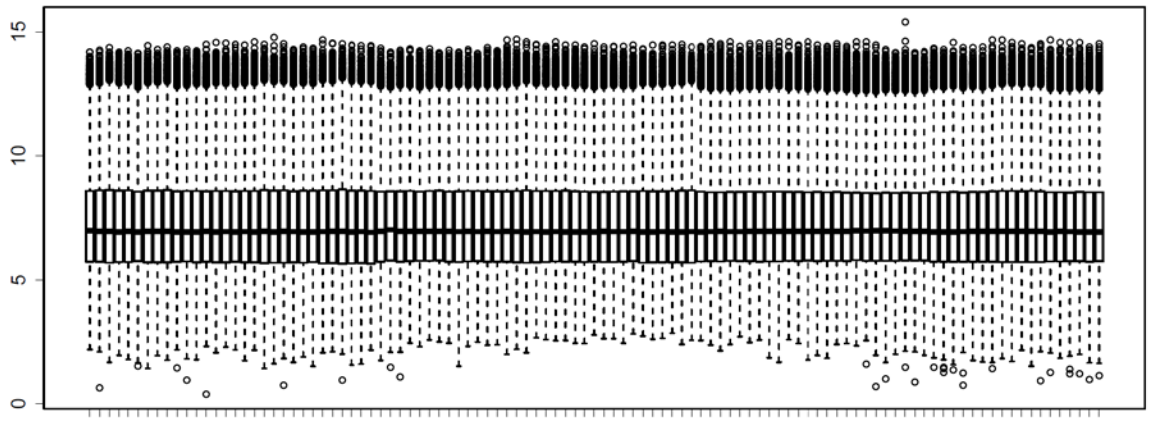

Fig. T4. Distribution of probe intensity of all samples after batch effect removal. The y-axis shows the expression intensity in $\log_2$ scale, and the x-axis denotes the sample.

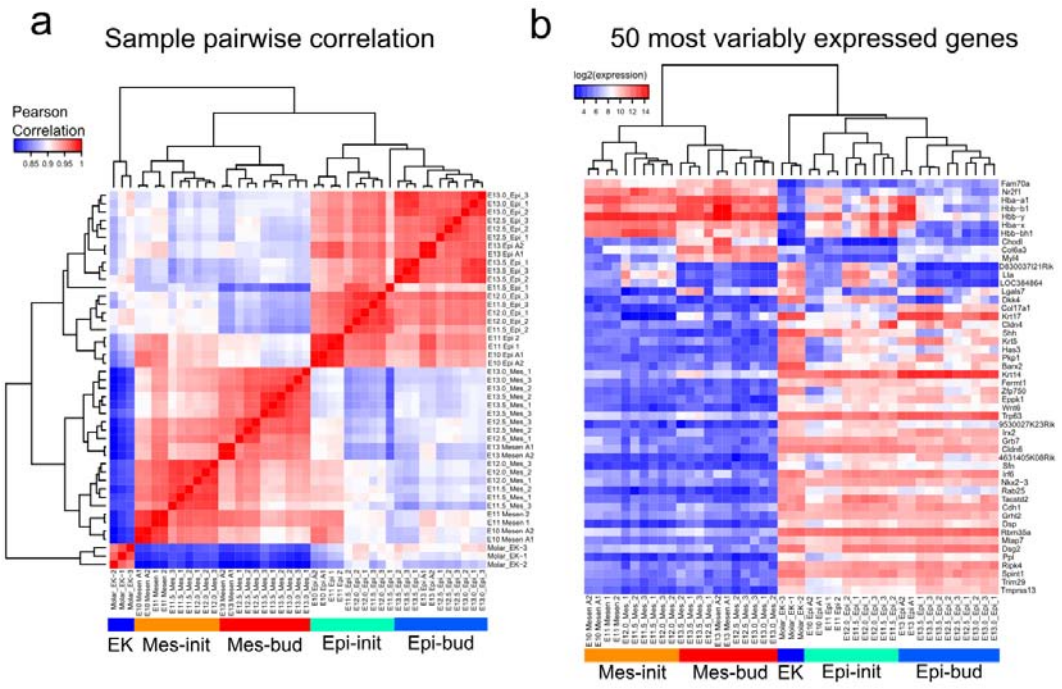

Fig. T5. Clustering of samples from the two time-course datasets. **a,** Correlation coefficients between every pair of samples from the two time-course experiments. Samples are clustered by their expected biological categories (epithelium or mesenchyme, and initiation- or bud-stage, and enamel knot). The average profile-wide Pearson correlation between samples is high (mostly above 0.9). **b,** Heat map showing the expression pattern of the 50 most highly variable genes. Profiles are grouped according to their expected biological categories.

Text S2. Comparison with qualitative gene expression data from the BITE-IT database

The BITE-IT database (http://bite-it.helsinki.fi/) is currently the most comprehensive gene expression database for all stages of mammalian tooth development. It contains qualitative mRNA expression measurements (mostly using in situ hybridization), and protein expression measurement for about one hundred genes that are pertinent to tooth development. Gene expression is recorded as either "+" for expression or "-" for lack of expression. The data in this database were manually assembled from reading and recording results in the primary research literature. We compared our microarray gene expression time series dataset against the BITE-IT records as an independent means to assess the quality of our microarray data.

We manually extracted the qualitative mRNA expression measurement from BITE-IT for dental epithelium and dental mesenchyme at initiation-, bud- and cap-stages (only enamel knot genes were extracted at the cap stage). Of all the records without any missing data, 52 genes can be mapped to a probe in our microarray platform using gene symbols. We conducted a receiver operator characteristic (ROC) analysis to assess the consistency of the qualitative record

in BITE-IT and our microarray gene expression time series data. We calculated the average expression abundance of each gene in each of the five conditions (Epi init, Epi bud, Mes init, Mes bud, EK). Treating the "+" and "-" measurement in BITE-IT as a 'gold standard', we calculated the area under the ROC curve (AUROC) for each gene using the average expression abundance of each gene. An AUROC of 1 indicates a perfect correspondence between the microarray data and BITE-IT, an AUROC of 0.5 indicates that the gene expression measurements correlate no better than expected by random, and an AUROC of 0 indicates a perfect anti-correlation of the gene expression time series and BITE-IT record. Analysis was carried out using the R package ROCR (5). The overall AUROC is about 0.8. For example, if we define the cut-off between non-expressed and expressed genes to be 7.16, our timecourse microarray data can match the BITE-IT gene expression with a sensitivity of 0.83 and a specificity of 0.71 (fig. T6).

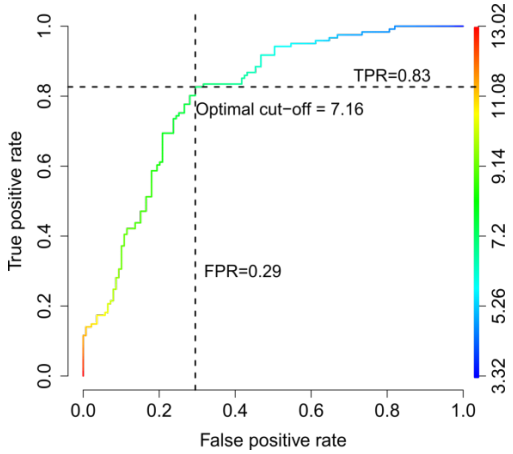

Fig. T6. Consistency between BITE-IT data and our microarray gene expression data. Using a log2 gene expression cut-off of 7.16, we achieved an overall sensitivity of 0.83 at a specificity at 0.71. The overall area under the ROC curve is about 0.8.

We further analyzed the concordance between our microarray data and the BITE-IT record at the gene level to test whether the expression pattern of each gene in our microarray dataset is consistent with BITE-IT across different tissues and time points. In this analysis, there are 42 differentially regulated genes (DRG), and 10 genes are non-DRGs (7 non-expressed and 3 constitutively expressed). The AUROC of each DRG was calculated, and the distribution is shown in fig. T7A. About 70% of the 42 DRGs in the BITE-IT database are perfectly consistent with our microarray dataset (AUROC=1). Of the 10 non-DRG genes, the non-expressed genes have significantly lower expression compared to those constitutively expressed genes (P = 0.017, by one sided Wilcoxon rank sum test; fig. T7B). To further investigate the discrepancy between our microarray gene expression pattern and the BITE-IT records (those genes with AUROC <1), we analyzed the coefficient of variation of each gene across different tissues and time points. We found that genes with low AUROC tend to have significantly lower coefficient of variation compared to genes with

high AUROC ($P < 0.05$, by one sided Wilcoxon rank sum test; fig. T7C). Further, the average expression abundances of all DRGs were similar (fig. T7D). This indicates that genes with expression patterns that differ from the BITE-IT database generally have smaller effect sizes. Therefore, a stringent cut-off when calling for differentially expressed genes should in principle remove most of these potential false positives.

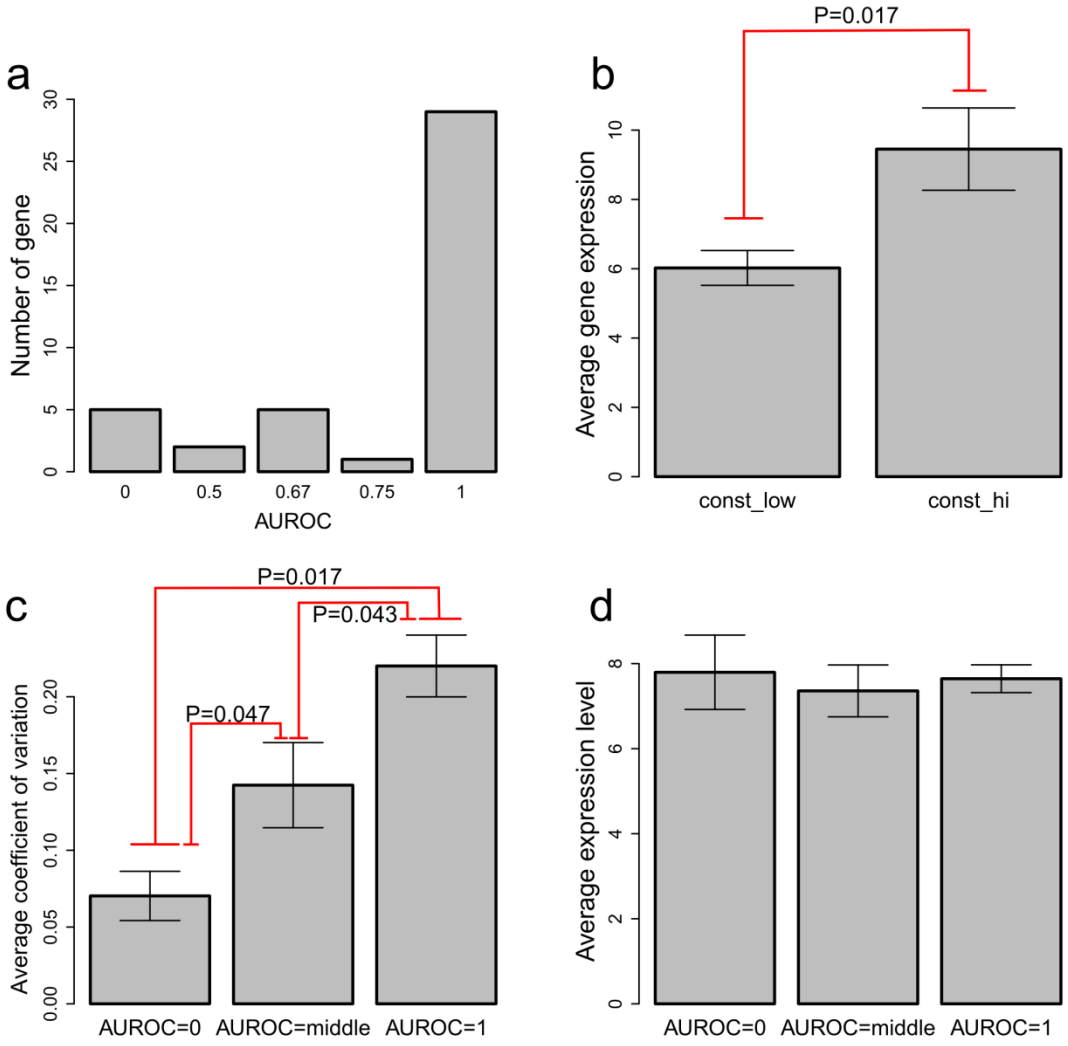

Fig. T7. Comparison of time-course microarray gene expression data (present study) and the BITE-IT database. **a,** Distribution of AUROC values. 70% of the 42 differentially regulated genes (DRGs) in the BITE-IT database are perfectly consistent with our microarray dataset (AUROC = 1). **b,** The mean expression of 10 non-DRGs (also see table S1). The seven genes that are non-expressed (const_low) have a statistically significant lower expression than the three constitutively expressed genes (const_hi) (P = 0.017, by one sided Wilcoxon rank sum test). This indicates that the microarray agrees with BITE-IT records. **c,** The average coefficient of variation of DRGs with AUROC=1 is significantly higher than those DRGs with lower AUROC (one sided Wilcoxon rank sum test). **d,** The average expression of all DRGs are similar regardless of the AUROC values.

Overall, it is encouraging to find that about 75% of 52 genes analyzed here have a gene expression profile that is highly consistent with the qualitative expression pattern recorded in BITE-IT (table S1). The genes with low AUROC values tend to have low coefficients of variation (and therefore small effect size). This indicates that a stringent effect size cut-off can be used to remove most of these false positive genes. It is particularly important that the expression patterns of key genes in our gene regulatory network (*Bmp4*, *Wnt*s, *Fgf*s, *Shh*, *Pax9* and *Msx1*) corroborate the mRNA measurements recorded in BITE-IT. This analysis further supports the quality of our microarray data.

Text S3. Construction of a gene regulatory network from perturbation data

We collected over 1,500 pieces of perturbation evidence from the literature. The majority of these data (1,055 entries) are from single perturbation experiments (single-gene knockout or treatment with a purified signaling molecule) between the developmental stages of E10.0 and E14.5. We restricted the data for constructing the gene regulatory network to these 1,055 entries for simplicity and relevance to the stages of tooth development investigated here. Summary statistics of this dataset are shown in fig. T8. In general, this dataset is comprehensive, with good coverage of the developmental stages, regulation type (positive, negative, and no regulation), tissue type, and year of publication.

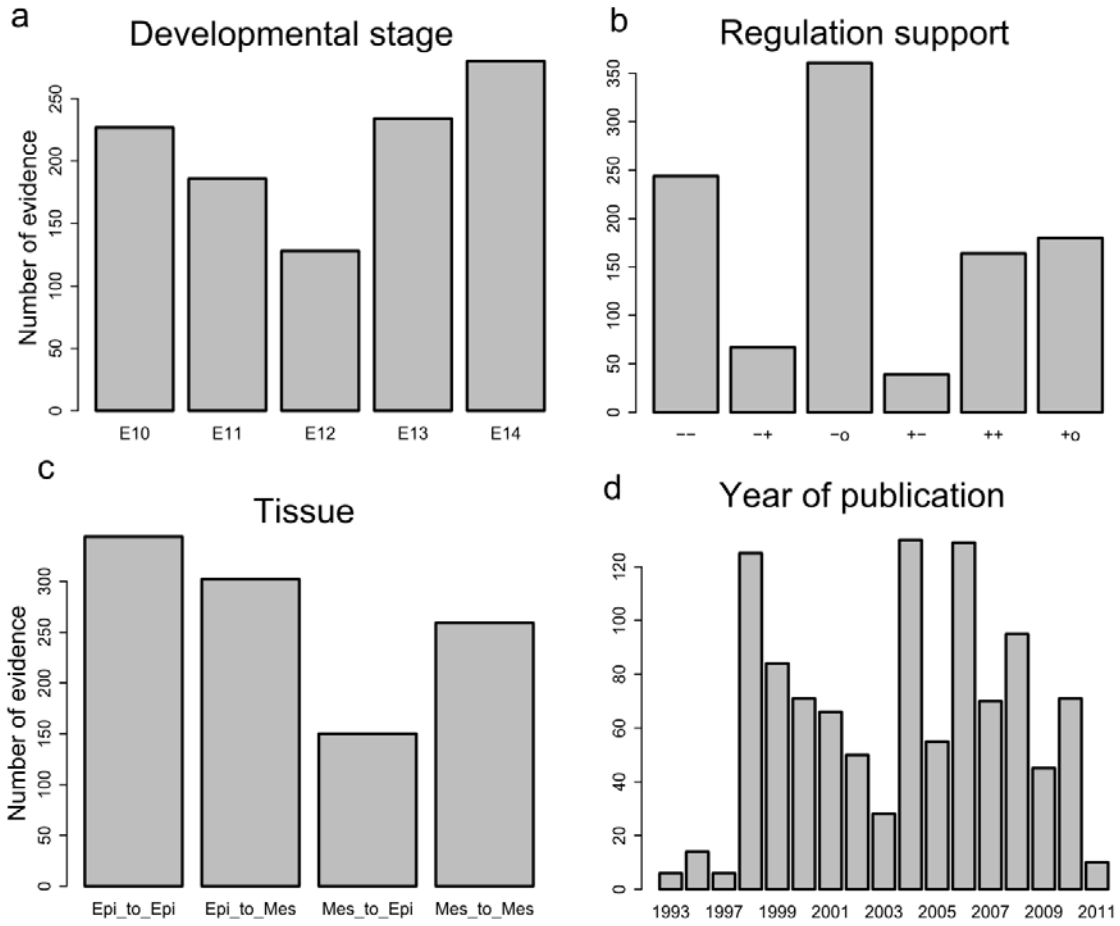

Fig. T8. Distribution of literature-derived perturbation evidence. a, Developmental stage; **b,** Type of regulation support (- equals downregulation; + equals upregulation; o equals no regulation) **c,** Tissue of the regulator and target; **d,** Year of publication. This analysis confirms comprehensive coverage of our perturbation dataset.

In addition to the literature evidence, we also compiled a list of 388 DRGs from the perturbation experiments generated in this study. These data were compiled from the DRGs in our signaling molecule treatment experiments and in our *Pax9* and *Msx1* mutant gene expression experiments [FDR < 0.05; log2(fold change) > 0.5; limma *t*-test]. Further, for inclusion, the genes must be involved in a

signaling pathway or have already been studied in the literature. We also included evidence from our in vivo mouse mutants (Figs. 3 and 4). These data were used to construct gene regulatory networks for E13.5 dental epithelium and dental mesenchyme (fig. T9).

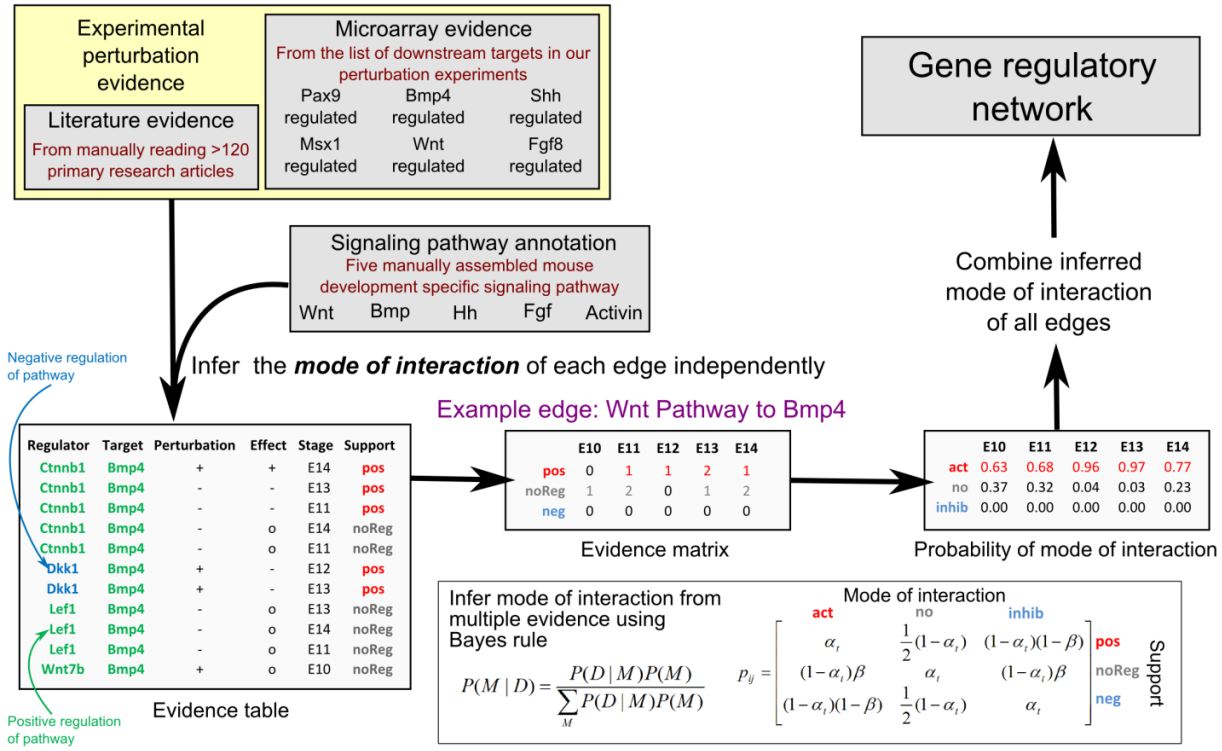

Fig. T9. Construction of a signaling based gene regulatory network from perturbation data. The mode of interaction of each edge is inferred independently of all other edges. This example illustrates how the mode of interaction of one edge (Wnt Pathway to *Bmp4*) is inferred based on perturbation data. Perturbation data that describe the regulation of *Bmp4* by the Wnt Pathway are summarized in an evidence table. The evidence was categorized as providing support for positive, negative, or no regulation. The evidence was then summarized as an evidence matrix, which describes the number of pieces of evidence of each type of support at each time point. We then used Bayes' rule to combine multiple pieces of evidence to estimate the probability of mode of interaction between this regulator and target. The mode of interaction with the highest probability (and greater than 0.6) was selected as the true mode of interaction. This procedure was repeated for all edges. The results of all edges were then combined to create a gene regulatory network.

The details of the likelihood based gene regulatory network construction procedure are summarized in the Materials and Methods section and illustrated in fig. T9. There are three important parameters associated with our likelihood model: α_0, β and, Δ_α. α_0 is the probability of observing the correct experimental evidence at the exact developmental stage in which the model is built, for example, using a piece of evidence from E13.5 to support a network model of E13.5. β is the probability of observing an incorrect experimental evidence due to insensitivity of the detection technology, and Δ_α is the decrease in likelihood

due to the time difference between the developmental stage of the observed evidence (for example, E12.5) and the developmental stage of the network model (for example, E13.5). All parameters are biologically motivated, and have an intuitive interpretation. Due to the sparsity of the perturbation evidence per edge, we do not deem parameter estimation an effective strategy. Therefore in this study, we set these parameters to some biologically intuitive values: α_0=0.9, Δ_α=0.15, and β=0.9. To ensure that our construction procedure is robust to the choice of parameters, we conducted a parameter sensitivity analysis. In this sensitivity analysis, we investigated the effect of changing the three parameters independently. In particular, we tested α_0=0.5,0.55,...,0.9; β =0.5, 0.55,...,0.9; and Δ_α =0.1, 0.15,...,0.5. We then calculated the proportion of consistent mode of interaction. An edge is said to have a consistent mode of interaction upon perturbing a parameter if the probability of mode of interaction is consistently above or below 0.6. Otherwise, an edge is said to have an inconsistent mode of interaction when a parameter is perturbed. The results are shown in figs. T10 to T13. We found that parameter β is very robust against very large changes (between 0.5 to 0.9). The parameter α_0 is quite robust against reasonably large changes (consistency only dropped to 80% when changing α_0 from 0.9 to 0.6). The parameter Δ_α is also quite robust against reasonably large changes (consistency only dropped to 80% when changing α_0 from 0.15 to 0.4). Therefore, we are confident that the structure of the gene regulatory network is robust when based on biologically motivated parameters.

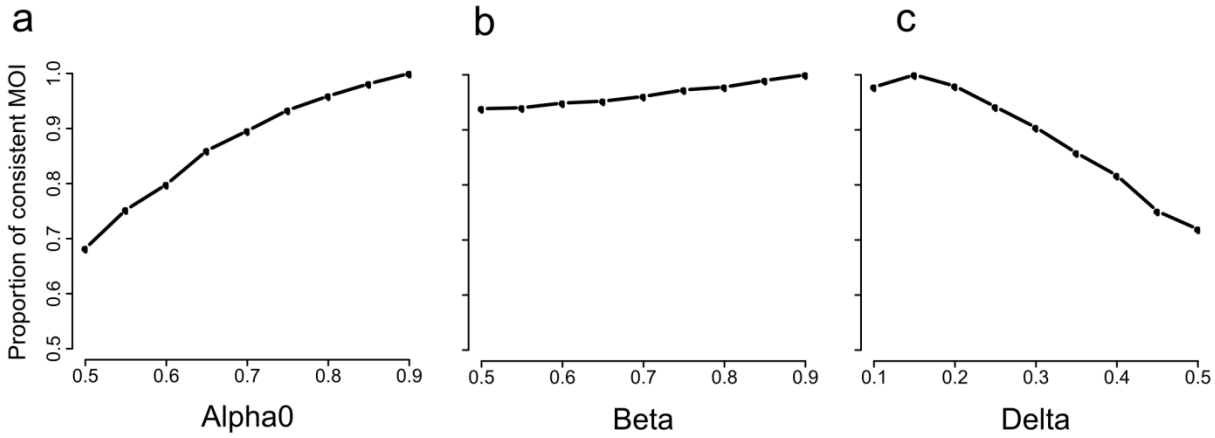

Fig. T10. Parameter sensitivity analysis of the gene regulatory network construction procedure. We tested the consistency of the mode of interaction of each edge upon perturbing each of the three parameters in the algorithm: a, α_0; b, β; c, Δ_α.

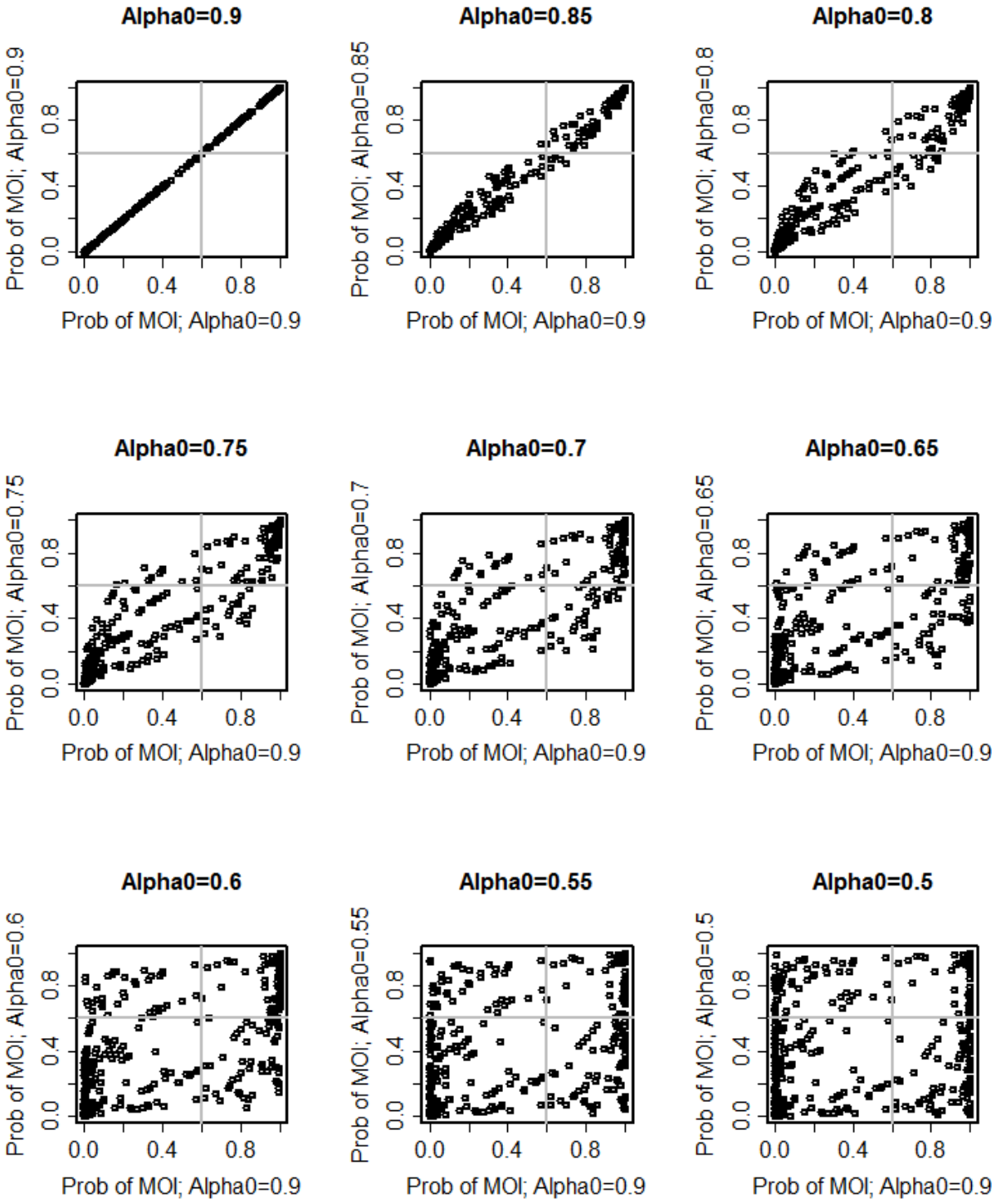

Fig. T11. Parameter sensitivity analysis of α_0.

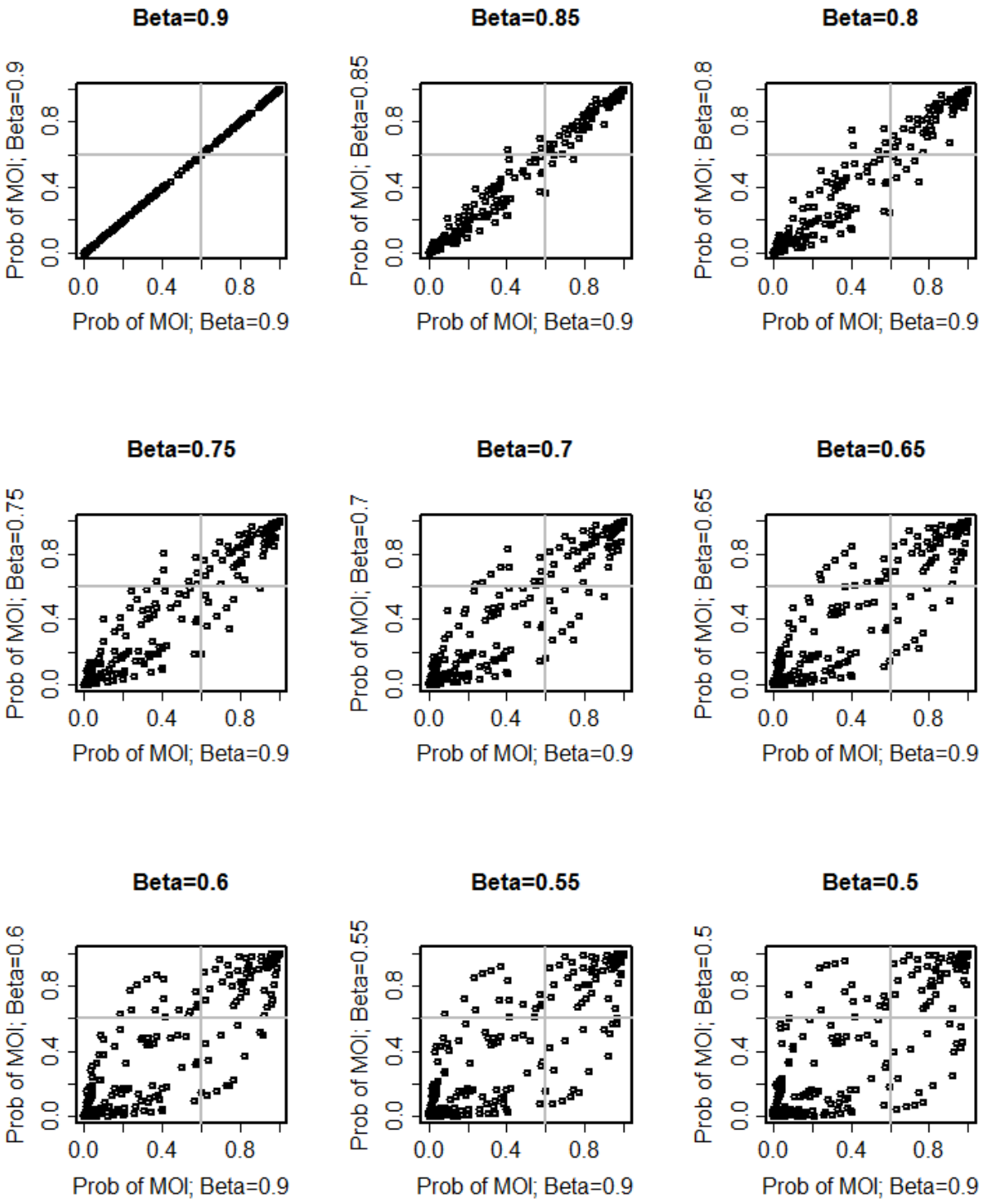

Fig. T12. Parameter sensitivity analysis of β.

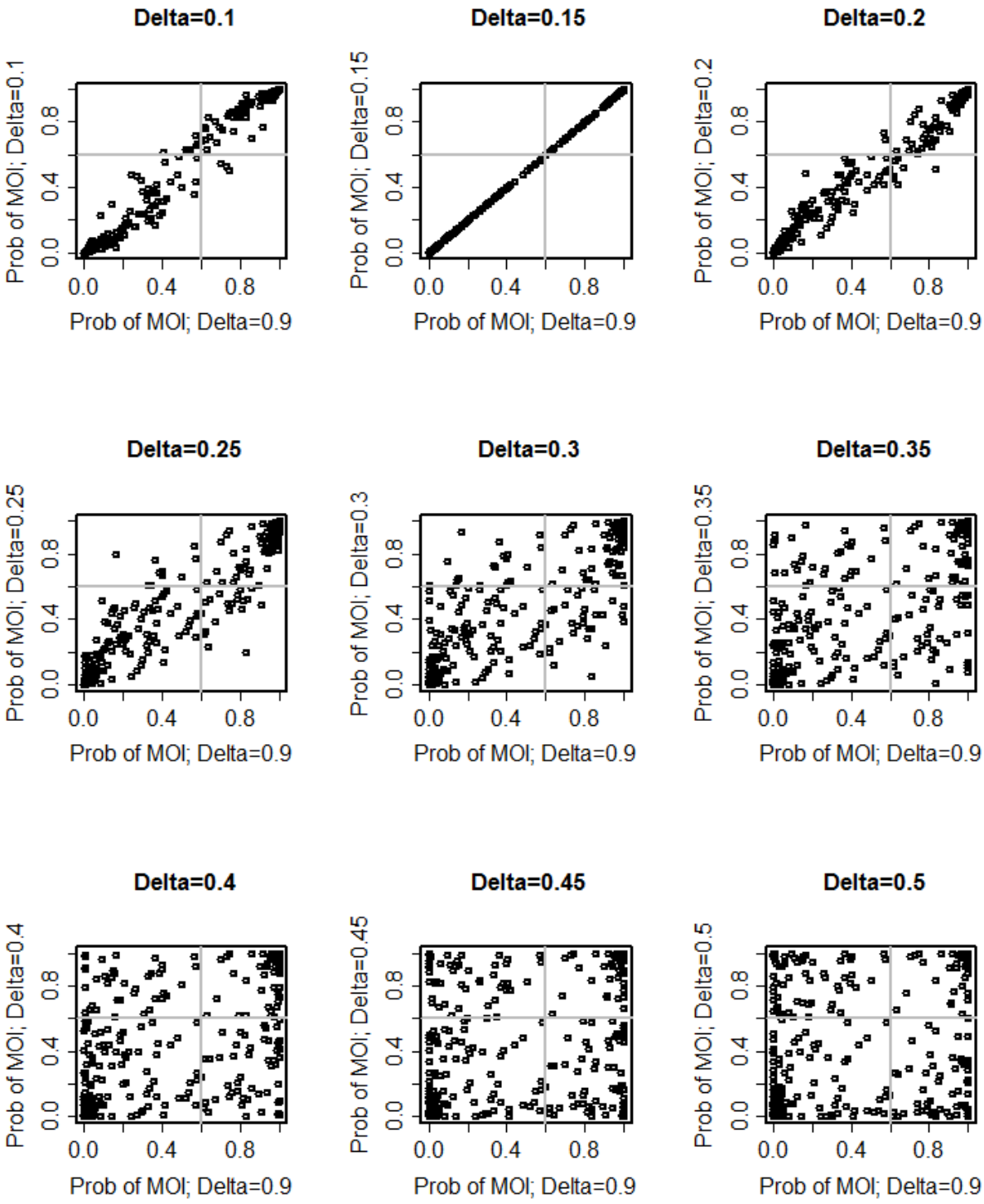

Fig. T13. Parameter sensitivity analysis of Δ_{α}.

Text S4. Simulation of the Wnt-Bmp feedback circuit

We made two observations based on the microarray gene expression analysis:

1. There is a surprising concordance in the changes of gene expression between the epithelium and mesenchyme. This is apparent from the PCA (Fig. 1A) and the contingency table of the differentially expressed genes (Fig. 1B).

2. The expression of extracellular signaling molecules is not concordant between the two tissues. In particular, there is a "shift" in tissue specific signaling molecule expression (Fig. 1C). Most signaling molecules are expressed in the epithelium at E10.0-E11.5, and later they are mostly expressed in the mesenchyme at E13.0-E13.5. At E14.5, the signaling molecules are expressed in both tissues. Canonical *Wnt* genes are expressed largely in dental epithelium, and *Bmp4* is expressed in both the epithelial and mesenchymal compartments in a temporally specific manner. This observation is consistent with the long-standing experimental observation that the instructive potential shifts temporally between the two tissues (6). Our microarray data support this conclusion.

How can we reconcile the globally concordant expression changes observed, with the discordant changes in the gene expression of key signaling molecules? Is it possible to use the structure of the gene regulatory network to explain this phenomenon? Here we seek to build a simple ordinary differential equation (ODE) model to explore whether the structure of the Wnt-Bmp feedback circuit can explain the observed sequential and reciprocal pattern of signaling molecule expression and overall concordance in genome-wide gene expression changes.

In particular, we aim to investigate whether a simple ODE model can generate the essential features that are characteristic of the dynamic expression patterns observed for *Bmp4* and *Wnt*s. These essential features include: [1] *Bmp4* expression is high in epithelium but low in mesenchyme at the initiation-stage (E10-E11.5), while the pattern is reversed at the bud-stage (E13-E13.5); [2] *Wnt* expression is higher in the epithelium at the bud- and cap-stages than the initiation- and placode-stages; and [3] *Bmp4* expression is high in both the epithelial enamel knot and the mesenchyme at E14.5; and [4] *Wnt* expression is generally undetectable in the molar dental mesenchyme at any of these stages. To quantify these features, we converted the average scaled expression of *Bmp4* and *Wnt* ligands based on our timecourse microarray data (fig. T14) into a reduced timecourse of 4 developmental stages: Initiation-stage (E10-E15), placode-stage (E12-12.5), bud-stage (E13-13.5), and cap-stage (E14.5). These reduced timecourse data can be represented as a line graph or bar plots (fig. T14).

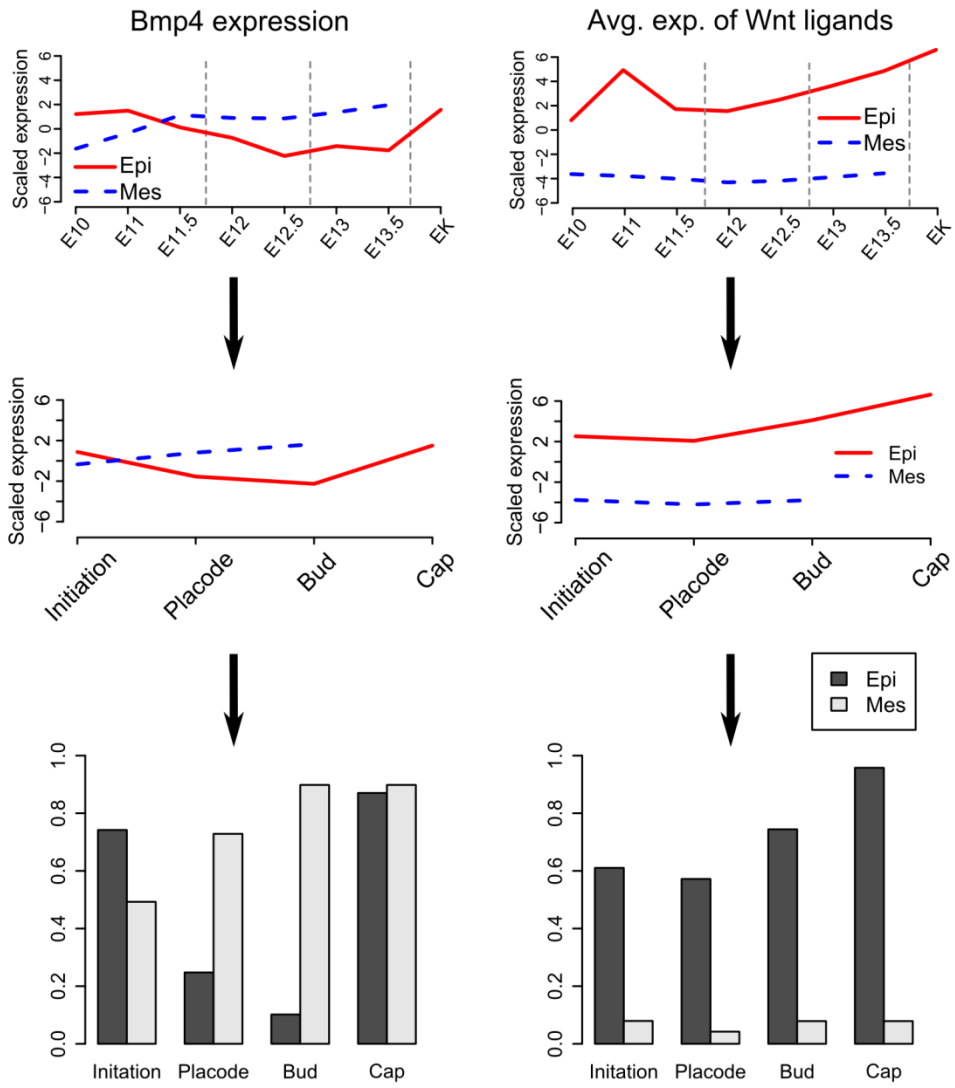

Fig. T14. The expression dynamics of Wnt and Bmp4 ligands from our microarray timecourse data.

4.1 Model description

The ODE model we have built is based on the Wnt-Bmp feedback circuit structure in fig. T15. Signal transduction and transcriptional activation (orange and black edges in the fig. T15, respectively) are modeled using the Hill equation, $h(x) = 1 + \frac{x^n}{x^n + K^n}$, where *x* represents expression of the regulator, *K* represents the activation threshold and *n* is the sigmoidicity constant, to model the non-linear influence a regulator(s) has on its target.

This approach for modeling gene regulatory networks has been used previously as a simple approximation of gene expression dynamics (7). We therefore created eight simple equations to model the activity of Wnt and Bmp pathways in the two tissue compartments (epithelial Bmp pathway activity = B_E, epithelial Wnt pathway activity = W_E, mesenchymal Bmp pathway activity = B_M, mesenchymal Wnt pathway activity = W_M), as well as the expression of the genes encoding the Wnt and Bmp4 ligands by the two compartments (epithelial *Bmp4* expression = b_E, epithelial *Wnt*s expression = w_E, mesenchymal *Bmp4* expression = b_M, mesenchymal *Wnt*s expression = w_M).

$$B_E : \frac{dx_1}{dt} = h(x_5 + x_7) - Dx_1$$

$$W_E : \frac{dx_2}{dt} = h(x_6 + x_8) - Dx_2$$

$$B_M : \frac{dx_3}{dt} = h(x_5 + x_7) - Dx_3$$

$$W_M : \frac{dx_4}{dt} = h(x_6 + x_8) - Dx_4$$

$$b_E : \frac{dx_5}{dt} = h(x_2) - Dx_5$$

$$w_E : \frac{dx_6}{dt} = h(x_1) - Dx_6$$

$$b_M : \frac{dx_7}{dt} = \frac{1}{2}\left[h(x_3) + h(x_4)\right] - Dx_7$$

$$w_M : \frac{dx_8}{dt} = 0$$

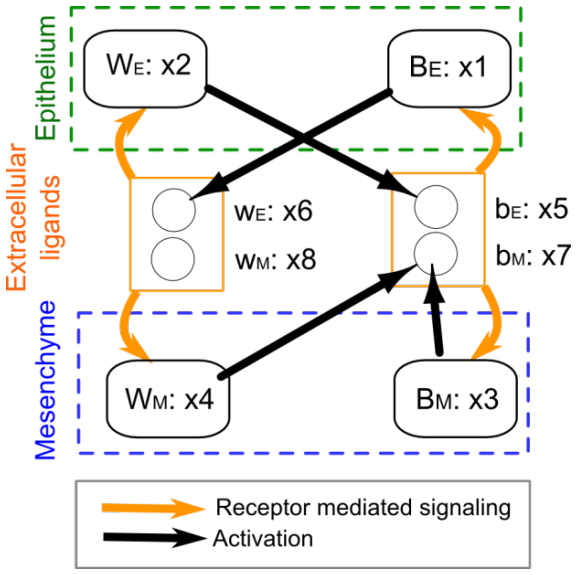

Fig. T15. The Wnt-Bmp feedback circuit, and the names of the variables used in the ODE model.

The model contains two tunable parameters: *K* and *D*. The two parameters have the following intuitive meanings:

K: The threshold in which the target expression is activated by its regulator

D: The rate of expression degradation

The parameter *n* in the Hill equation controls the sigmoidicity of the function *h(x)*. In the simulation, we fixed this parameter as *n* = 20. We note that different values of *n* would also be appropriate as long as they allow the generation of a sharp activation curve (fig. T16). A sigmoidal function was chosen mainly to capture the non-linear dynamics pervasive in biochemical signaling systems. The rate of degradation, *D*, and the activation threshold, *K*, were identified by an exhaustive search. Both *D* and *K* are identical across all 8 ODEs in the simulation of wildtype and mutant conditions.

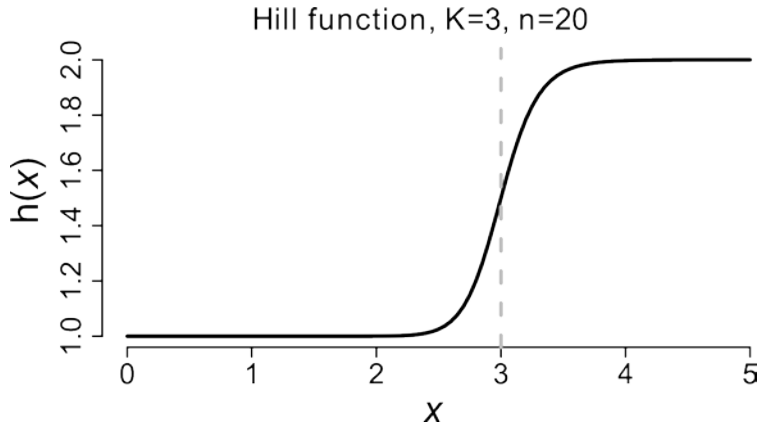

Fig. T16. The Hill equation. The parameter *n* controls the sigmoidicity of the curve: the larger the *n*, the sharper the bend at the inflection point (*n*=20 in this example). The parameter *K* determines the inflection points. In this example, the inflection is set to *K*=3.

Modeling assumptions:

1. The activity of one pathway in both tissue compartments (for example, the Wnt pathway in epithelium and mesenchyme) is identical for all time points in wild-type. This assumption enforces the observation that the genome-wide expression patterns are concordant between tissues. The expression patterns of Wnt and Bmp ligands are allowed to differ.

2. The activation threshold (*K*), and coefficient of signal degradation (*D*) is the same across all eight equations. This implies that we do not make an assumption about pathway specific dynamics. This assumption reduces the number of parameters in the overall model, and therefore guards against over-fitting the model.

The model was implemented and simulated using GNU Octave (http://www.gnu.org/software/octave/). The ODE model was solved using the *lsode* function.

4.2 Implementation and parameter estimation

The initial values of signaling molecule expression and pathway activity were assigned based on the microarray data at the initiation-stage (Fig. 2). Since the microarray data only provide relative expression measurements, the actual numerical values of these initial values cannot be specified. Nonetheless, the relative magnitude of expression should be sufficient to capture this set of initial values. The results shown in this section are not sensitive to small changes in the initial values.

$$B_E : x_1 = 0.5 \qquad b_E : x_5 = 0.9$$
$$W_E : x_2 = 0.5 \qquad w_E : x_6 = 0.6$$
$$B_M : x_3 = 0.5 \qquad b_M : x_7 = 0.1$$
$$W_M : x_x = 0.5 \qquad w_M : x_8 = 0.1$$

We conducted a systematic parameter search to identify a set of parameters that could qualitatively capture the observed expression dynamics of the *Wnt* and *Bmp4* signaling molecules. We tested all combinations of *K* and *D* within the following range (below).

Parameter	Number of candidate searched	Range
K	50	[0.6,1.0]
D	50	[1.8,2.2]

For each set of parameters, we simulated a time-course profile, and checked whether there was a single cut-off value such that the expression could be characterized as high (H) or low (L), and that would match the observed expression profile (below).

Signaling	Initiation	Placode	Bud	Cap
Epi Bmp4	H	L	L	H
Epi Wnt	L	L	H	H
Mes Bmp4	L	H	H	H
Mes Wnt	L	L	L	L

If the qualitative expression profile matched all 20 qualities (4 signaling molecules and 5 time points), this set of parameters was termed “valid”. All “valid” parameters are shown in fig. T17. This experiment shows that *K* and *D* are correlated, and must be balanced to obtain the observed expression patterns. For the following discussion, we chose one valid parameter set, *K*=0.9 and *D*=2, for subsequent analysis.

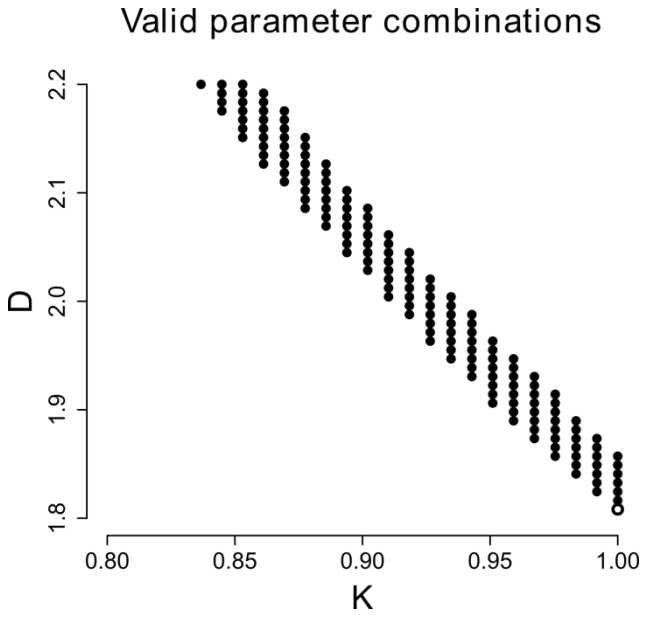

Fig. T17. Estimation of valid parameters for ODE modeling. Each valid combination of parameters *D* and *K* is indicated as a dot in this parameter space.

4.3 Model simulation for mouse mutants

To check the validity of the model, we applied this ODE model to simulate the gene expression dynamics of canonical *Wnt*s and *Bmp4* in wild-type tooth development from E10.0 to E14.5, as well as in various genetic mutants: [1] epithelial *Apc* loss-of-function (LOF) (*Apc* LOF); [2] *Pax9* null mutant (*Pax9* null); [3] compound epithelial *Apc* LOF; *Pax9* null; (*Apc LOF ; Pax9* null); [4] epithelial *Bmpr1a* LOF mutant (*Bmpr1a* LOF); and [5] compound epithelial *Apc* LOF; *Bmpr1a* LOF (*Apc LOF ; Bmpr1a* LOF) (figs. T18 to T23).

The data from our microarray timecourse experiment were used to identify valid values for parameters K and D to capture the essential features of the observed signaling dynamics. Not only can the essential features of Bmp4 and Wnt expression dynamics be captured by our simulation, we can observe that the activity of the Bmp and Wnt pathways (the third and forth panels of fig. T18) remain identical across the two tissues compartments, suggesting our simulation can also capture simultaneous reception of signals to the two compartments. This result therefore provides evidence that the Wnt-Bmp feedback circuit structure by itself is sufficient to account for reciprocal expression of signaling molecule genes (*Bmp4* and *Wnt*s), and suggests that simultaneous signal transduction may account for the genome-wide concordant gene expression changes that are observed in dental epithelium and mesenchyme.

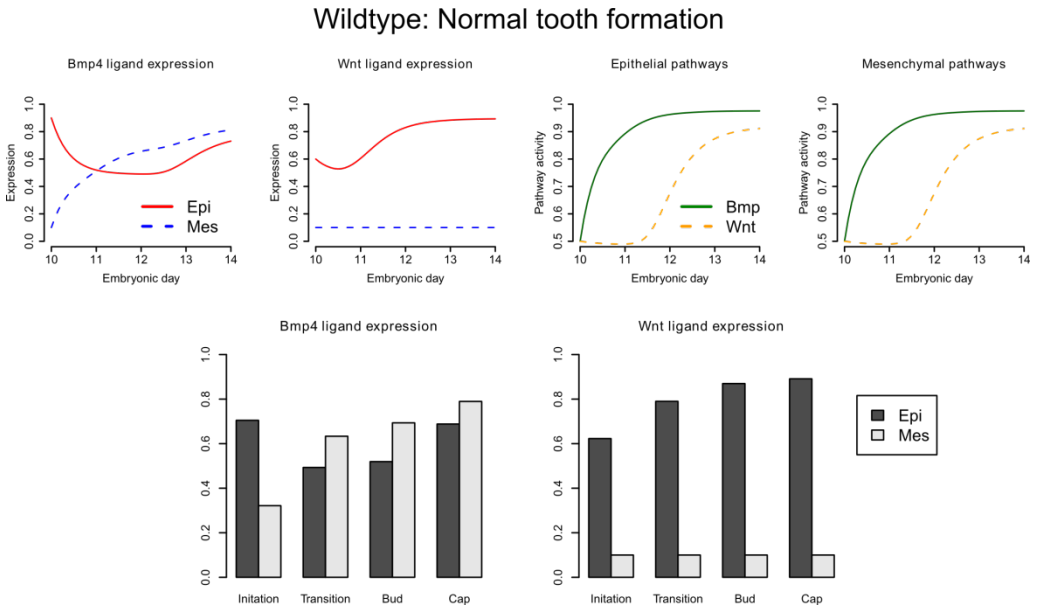

Fig. T18. Simulation of wildtype gene expression dynamics based on the ODE model.

To simulate epithelial *Apc* LOF, we assumed that the activation threshold, K, of the epithelial Wnt pathway is halved (for example, $K_{WE}=K/2$), which mimics the situation where the Wnt pathway is activated more easily in the absence of the inhibition mediated by *Apc*. Our simulation (fig. T19) shows that epithelial Wnts and Bmp4 expression remains high throughput bud- and cap-stages, which is consistent with our experimental observation (Figs. 3 and 4).

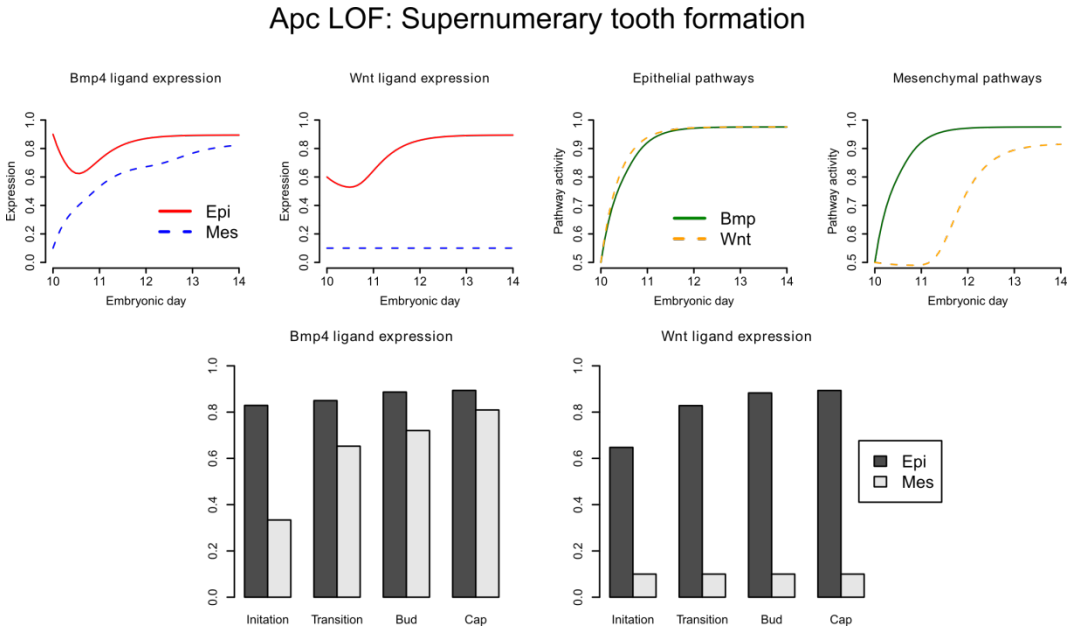

Fig. T19. Simulation of epithelial *Apc* LOF gene expression dynamics based on the ODE model.

To simulate the *Pax9* null mutant, we assumed the rate of change of mesenchymal *Bmp4* production to be zero:

$$b_M : \frac{dx_7}{dt} = 0$$

This assumption decouples the influence of mesenchymal Bmp4 and Wnt pathways on *Bmp4* expression, which is the main functional defect in *Pax9* null mice. Our simulation (fig. T20) shows that the expression of *Bmp4* and *Wnt*s are largely lost in both compartments, which is consistent with our observations from in vivo validation (Fig. 3).

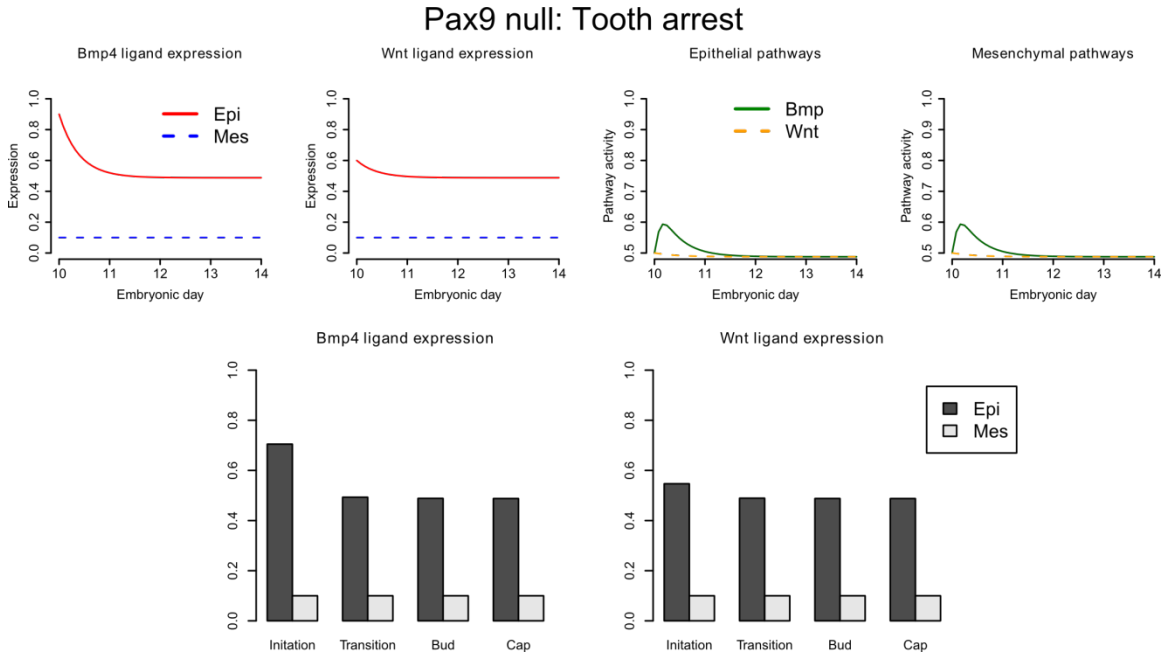

Fig. T20. Simulation of *Pax9* null gene expression dynamics based on the ODE model.

To stimulate the expression patterns of epithelial *Apc* LOF and *Pax9* null compounds, we combined the alterations of the abovementioned mutant ODEs. Our simulation (fig. T21) shows that the epithelial expression of *Bmp4* and *Wnt*s are regained in compund *Apc* LOF and *Pax9* null mutants. This is consistent with our hypothesis that ectopic constitutive activation of epithelial Wnt signaling can lead to up-regulation of *Bmp4* expression in the epithelium, and this epithelial derived Bmp4 can substitute for the ablated mesenchymal *Bmp4* expression and satisfy the epithelial Bmp4 signaling requirement.

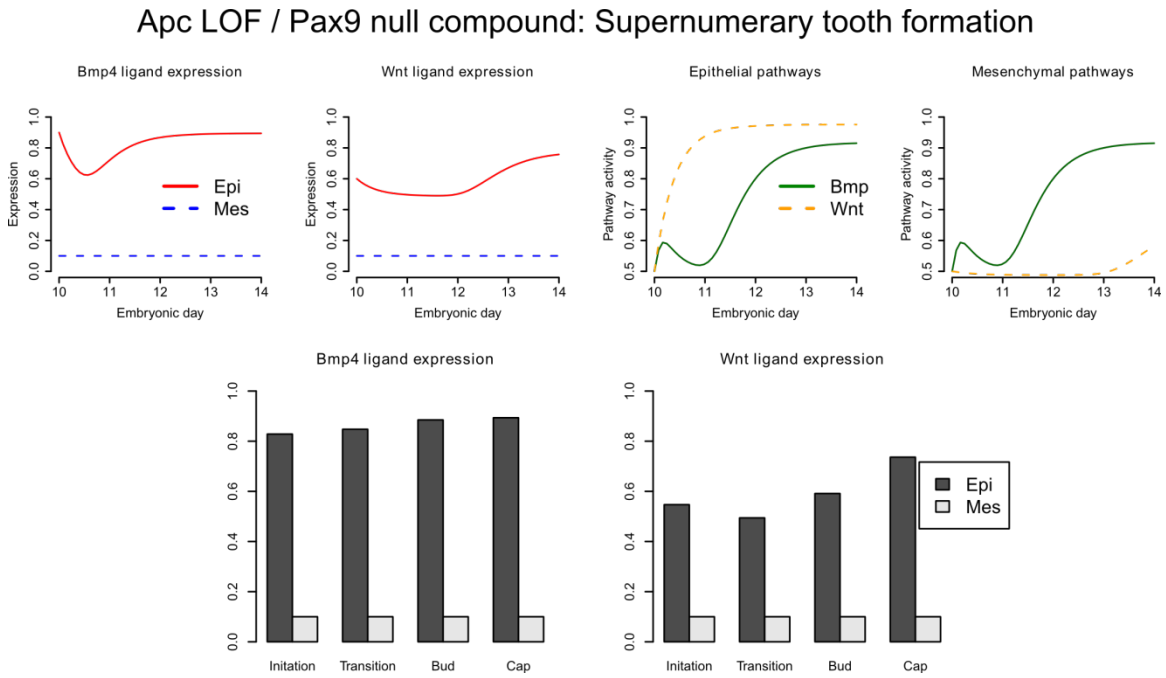

Fig. T21. Simulation of epithelial *Apc* LOF and *Pax9* null compound mutant gene expression dynamics based on the ODE model.

To stimulate the expression patterns of epithelial *Bmpr1*a LOF, we assume the rate of change of epithelial Bmp pathway activity to be zero:

$$B_E : \frac{dx_1}{dt} = 0$$

This assumption decouples the influence of the epithelial Bmp pathway from the extracellular *Bmp4* concentration. Our simulation (fig. T22) shows that the expression of *Bmp4* and *Wnt*s is largely lost in the epithelium while mesenchymal *Bmp4* expression remains high, which is consistent with our observations from *in vivo* experiments (Fig. 4).

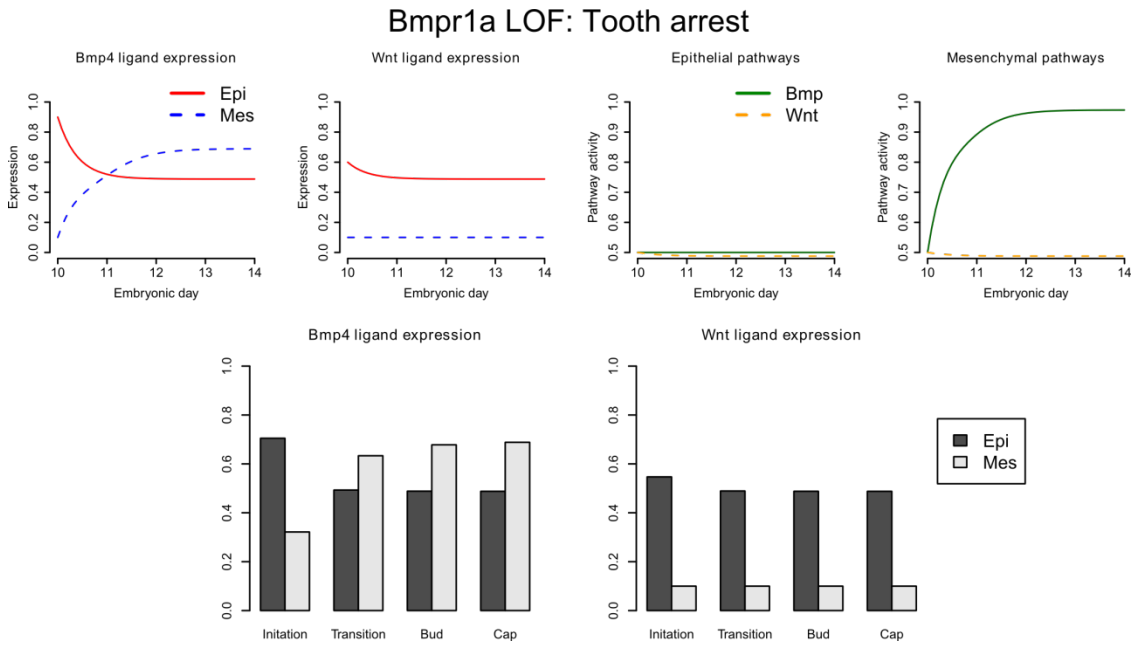

Fig. T22. Simulation of epithelial *Bmpr1a* LOF gene expression dynamics based on the ODE model.

To stimulate the expression patterns of epithelial *Apc* and *Bmpr1a* LOF compound mutants, we again combined the alterations of the abovementioned mutant ODEs. Our simulation (fig. T20) predicts increased abundances of epithelial *Bmp4* expression and decreased abundances of epithelial *Wnts* expression in *Apc* and *Bmpr1a* LOF mutants, which is confirmed in vivo (Fig. 4).

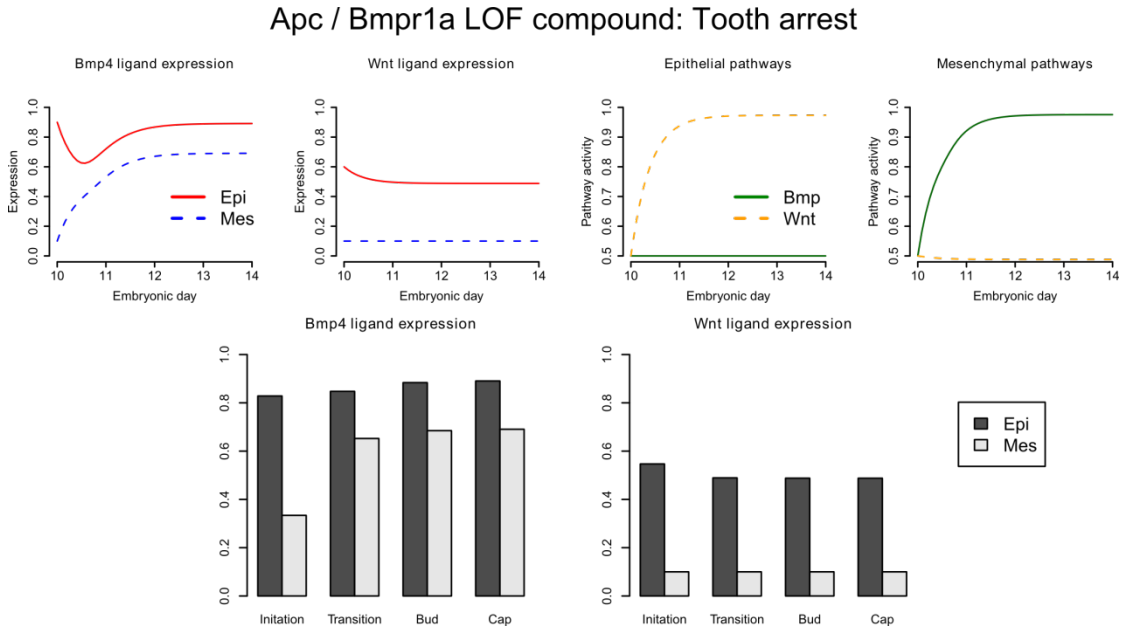

Fig. T23. Simulation of epithelial *Apc* and *Bmpr1a* LOF compound mutant gene expression dynamics based on the ODE model.

4.4 Discussion

The aim of this analysis was to explore whether the essential features of the sequential and reciprocal expression dynamics of *Bmp4* and *Wnt*s can be recapitulated using a network structure inferred from systematic analyses. Currently, our simple ODE model has a small number of parameters. If we include additional tissue- or pathway-specific terms to the ODE model, the simulation should fit the observed expression abundances even better. Nonetheless, this simple model is already sufficient to yield qualitative expression dynamics that are consistent with the expression time-course data observed in endogenous tooth development, and correctly predicts the expression of *Wnt*s and *Bmp4* in different mutant mice.

The network structure itself contains sufficient information to explain the observed temporal and tissue specific expression patterns of canonical *Wnt* and *Bmp4* signaling molecule genes. In the epithelium, *Wnt*s and *Bmp4* expression are each cross-regulated by the complementary signaling pathway, while in the mesenchyme, *Bmp4* is jointly regulated by both pathways. This circuit structure

leads to ODE solutions that predict a temporal shift in signaling between epithelium and mesenchyme. Thus, the reciprocal expression of signaling molecules, typical of most epithelial-mesenchyme interactions, is in this case an inherent property of the Wnt-Bmp feedback circuit itself.

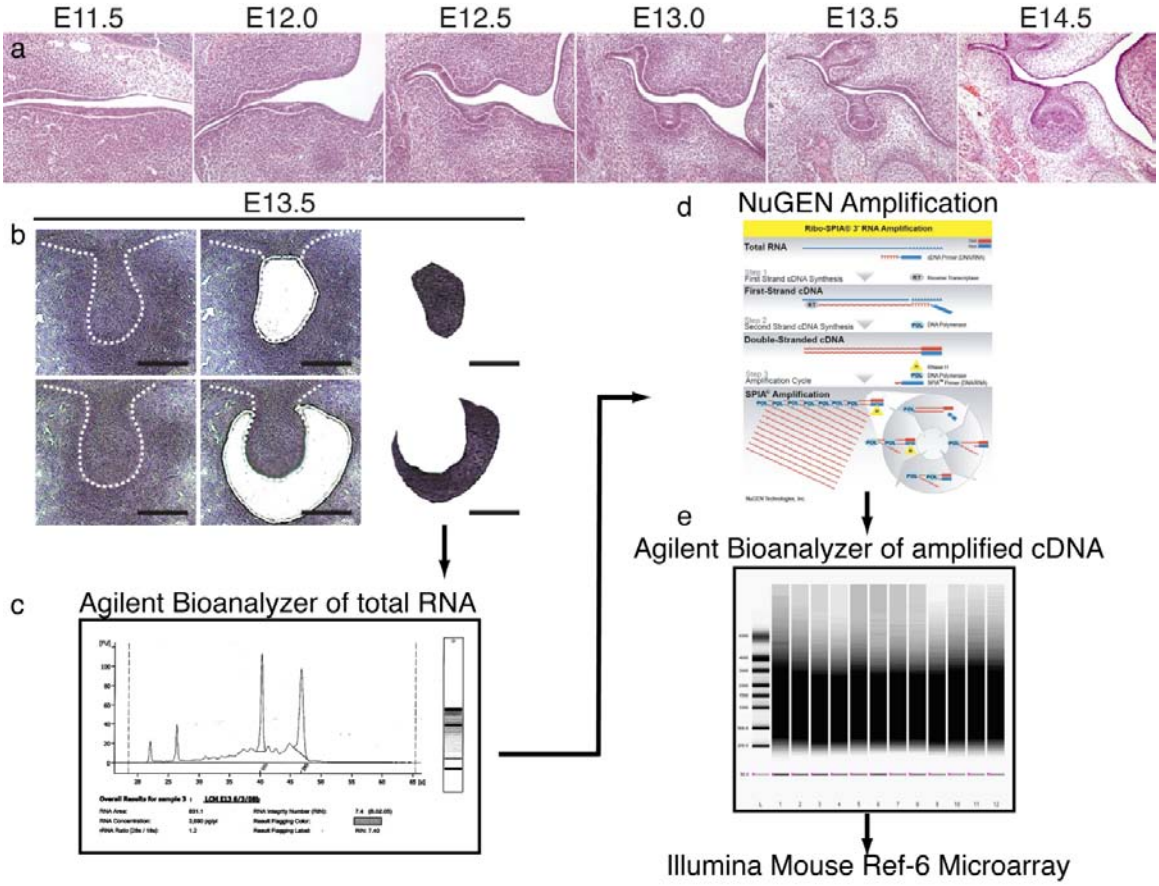

Fig. S1. LCM workflow. **a,** Coronal sections of embryonic molar tooth morphology representing the stages for which RNA was analyzed by gene expression microarray after laser capture microdissection (LCM). **b,** Representative LCM captures of epithelial and mesenchymal tissue compartments from freshly frozen E13.5 coronal cryosections. Upper: epithelium; lower: mesenchyme. Scale bars: 100 μm. **c,** Representative Bioanalyzer tracings of total RNA purified from LCM sections. **d,** NuGEN PolydT based total RNA amplification protocol results in cDNA suitable for microarray analysis. **e,** Bioanalyzer gel trace displaying representative size distribution of amplified and end-labeled cDNA prior to hybridization on Illumina Mouse WG-6 v2 Expression BeadChip whole genome array.

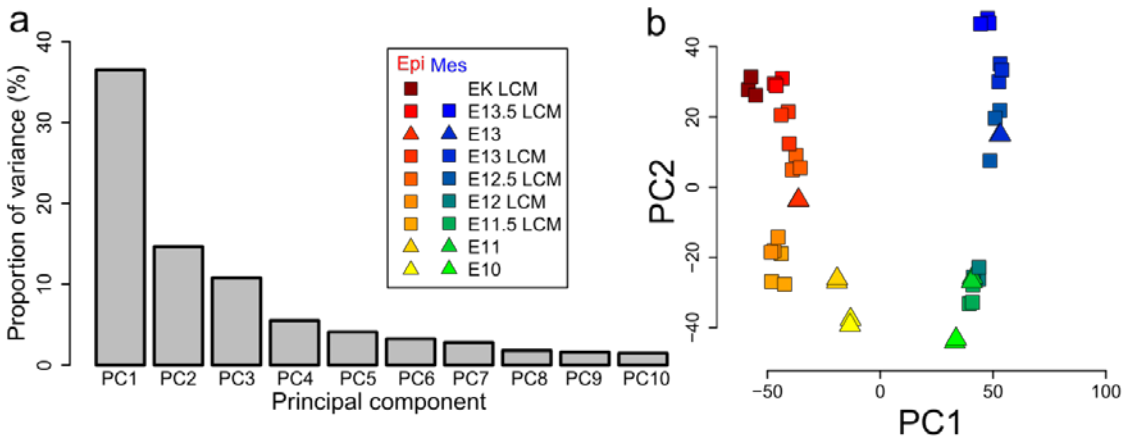

Fig. S2. Principal components analysis of time series gene expression profiles. a, Bar chart showing proportion of variance explained by each principal component (PC). PCs 1 and 2 collectively account for over 50% of the variance in the dataset. **b,** Two dimensional scatter plot representation of the first two principal components (PC1 and PC2). PC1 separates samples based on their tissue type (dental epithelium or dental mesenchyme). PC2 separates samples based on developmental stages [E10.0, E11.0, E11.5, E12, E12.5, E13, E13.5 and enamel knot (EK)]. Samples from each tissue compartment form an approximate line along the direction of PC2, and lines formed by the epithelium and mesenchyme roughly parallel each other. This pattern suggests that both tissue compartments experience similar changes in global gene expression through developmental time. Alternatively stated, the temporal changes in gene expression are concordant between the epithelial and mesenchymal tissue compartments. Both LCM and manually dissected time-course data were included in this analysis and give consistent results. N = 3 biological microarray replicates for LCM generated samples and N = 2 biological microarray replicates for manually separated tissue.

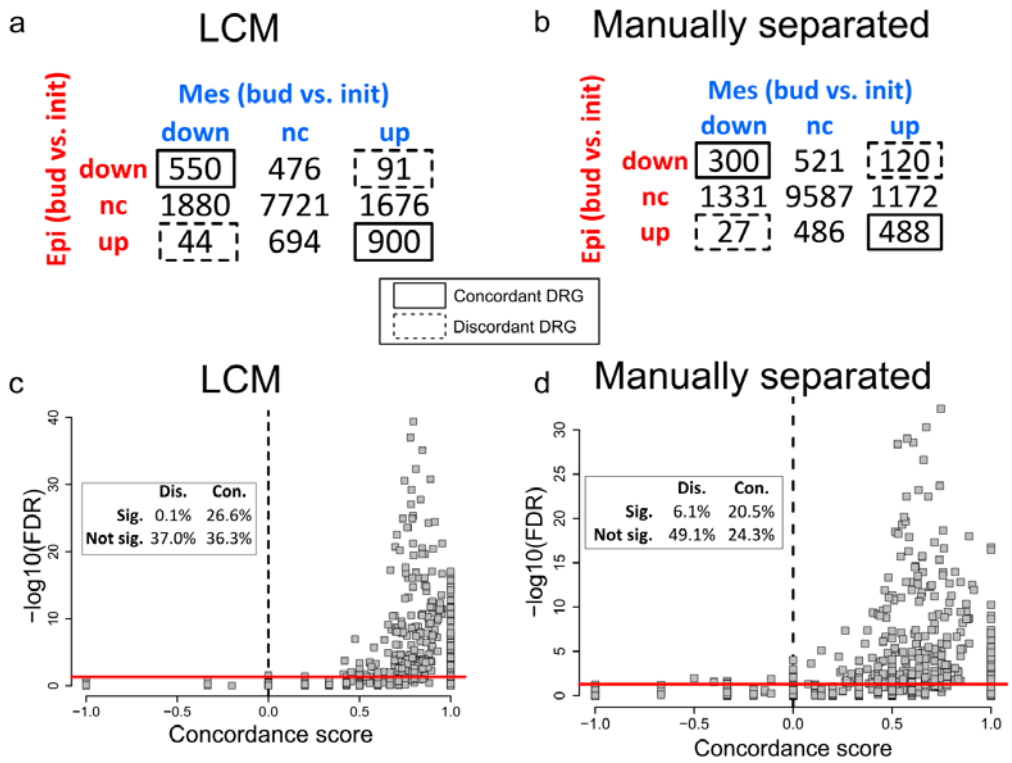

Fig. S3. Analysis of gene expression concordance between epithelium and mesenchyme with the LCM and manually dissected time course data sets. **a** and **b,** Contingency table showing number of genes with significantly increased or decreased expression in epithelium and mesenchyme. The distribution of genes in each category is significantly non-random ($P<10^{-15}$, χ^2 test). **c** and **d,** Analysis of gene expression concordance in each gene set. We first divided all genes according to different functional gene categories (Gene ontology, KEGG pathways, and so on), and then computed the statistical significance (χ^2 test) and concordance score for each gene set. Each square represents a gene set, and its location is determined by the false discovery rate (FDR) using the χ^2 test and the concordance score. The majority of statistically significant gene sets (FDR < 0.05) have concordant gene expression changes (concordance score > 0). Both LCM and manually separated time-course datasets, independently collected, confirm the statistically significant concordance of differentially regulated genes (DRGs) between the epithelial and mesenchymal tissue compartments, indicating that this concordance is a genome-wide property across many functional gene sets.

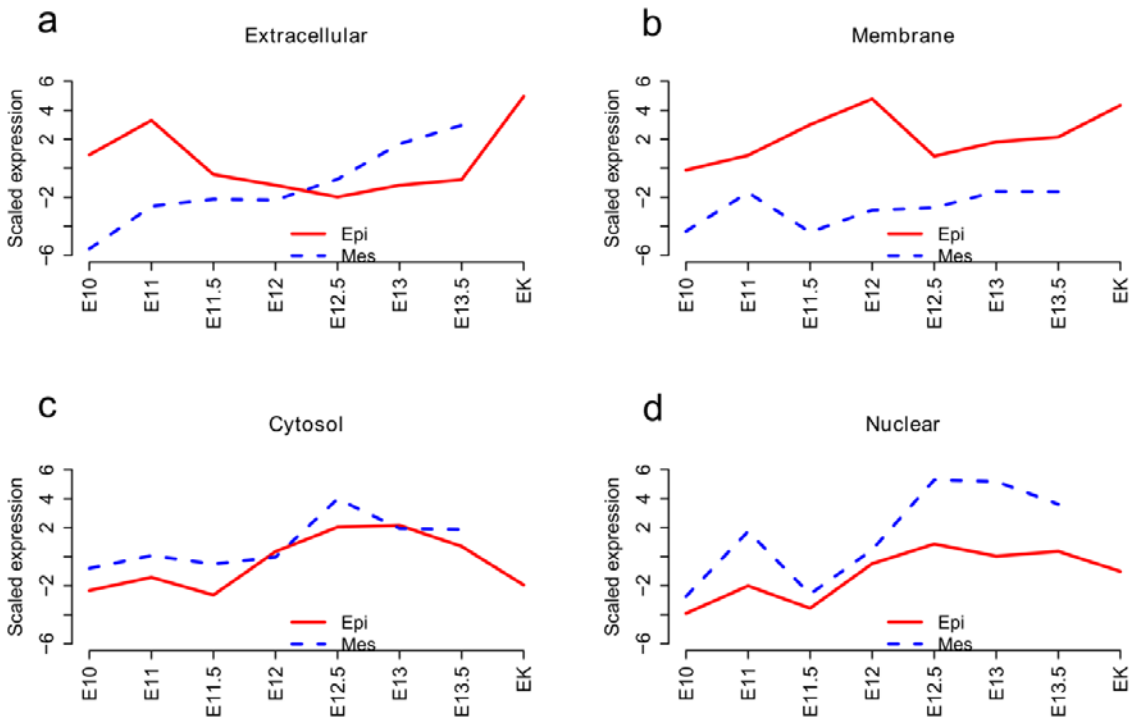

Fig. S4. Average scaled gene expression patterns of different signaling pathway components. **a,** Average expression profiles of extracellular signaling molecules. **b,** Average expression profiles of membrane-bound components (mostly receptors). **c,** Average expression profiles of cytosolic components (mostly intracellular signal transduction components). **d,** Average expression profiles of nuclear components (mostly transcription factors or transcriptional co-activators or co-repressors). The mesenchymal-derived extracellular signaling molecule average gene expression profile (a) shows a progressive increase between E10 and 13.5, whereas the comparable epithelial expression profile shows a decrease between E11 and 12.5, followed by an increase. Thus, from comparison of the gene expression patterns of different signaling components, the spatiotemporal expression patterns of extracellular signaling molecules (a) appears most closely matched to the shift in tooth instructive potential from dental epithelium at E10-E11.5 to dental mesenchyme at E12.5-E13.5 (6).

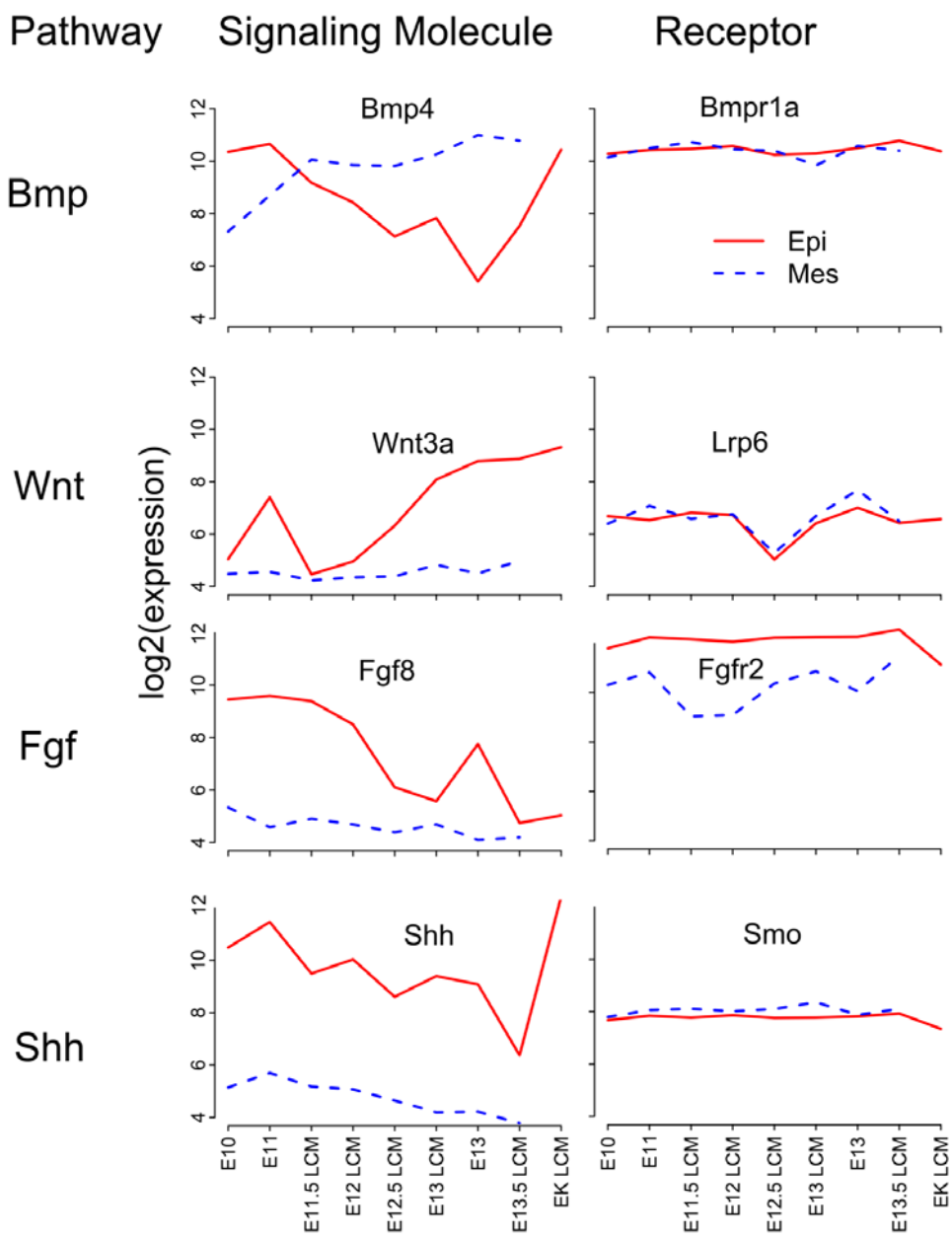

Fig. S5. Expression dynamics of representative extracellular signaling molecules and receptors (Bmp, Wnt, Fgf, and Shh pathways). Gene expression is reported as average log2(expression) of the combined microarray time-course dataset (LCM and manually separated tissues). Signaling molecules show highly dynamic expression changes, whereas the expression of their receptors is relatively unchanged over time. These data suggest that both epithelial-mesenchyme compartments are competent to transduce receptor-mediated signals over time, and that signaling dynamics are most likely controlled by ligand expression.

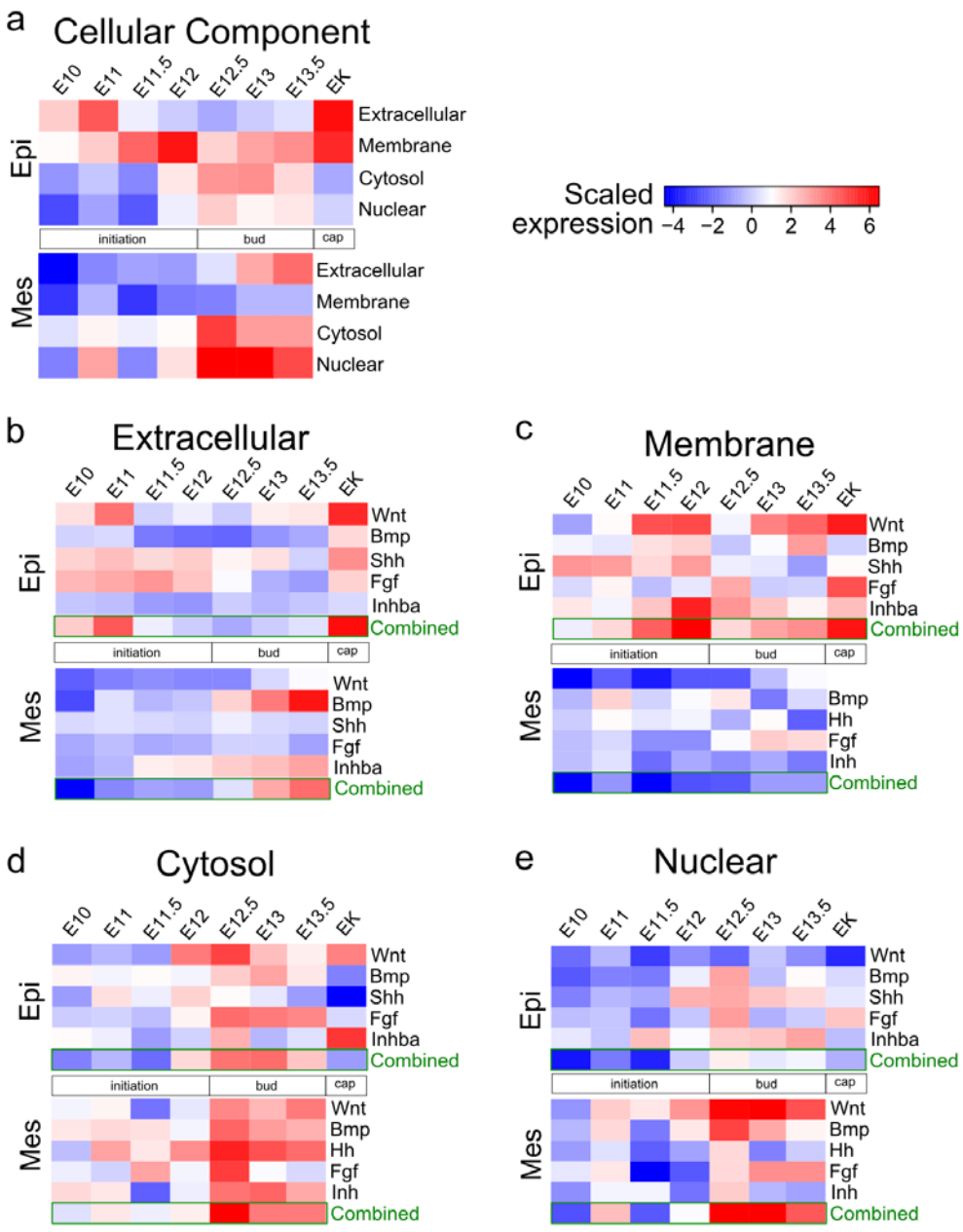

Fig. S6. Heat maps of average expression of cellular components in different signaling pathways. **a,** Average expression profiles of different cellular components (averaging over selected pathways). **b,** Average expression profiles of extracellular signaling molecules. **c,** Average expression profiles of membrane-bound components (mostly receptors). **d,** Average expression profiles of cytosolic components (mostly intracellular signal transduction components). **e,** Average expression profiles of nuclear components (mostly transcription factors or coactivators or corepressors).

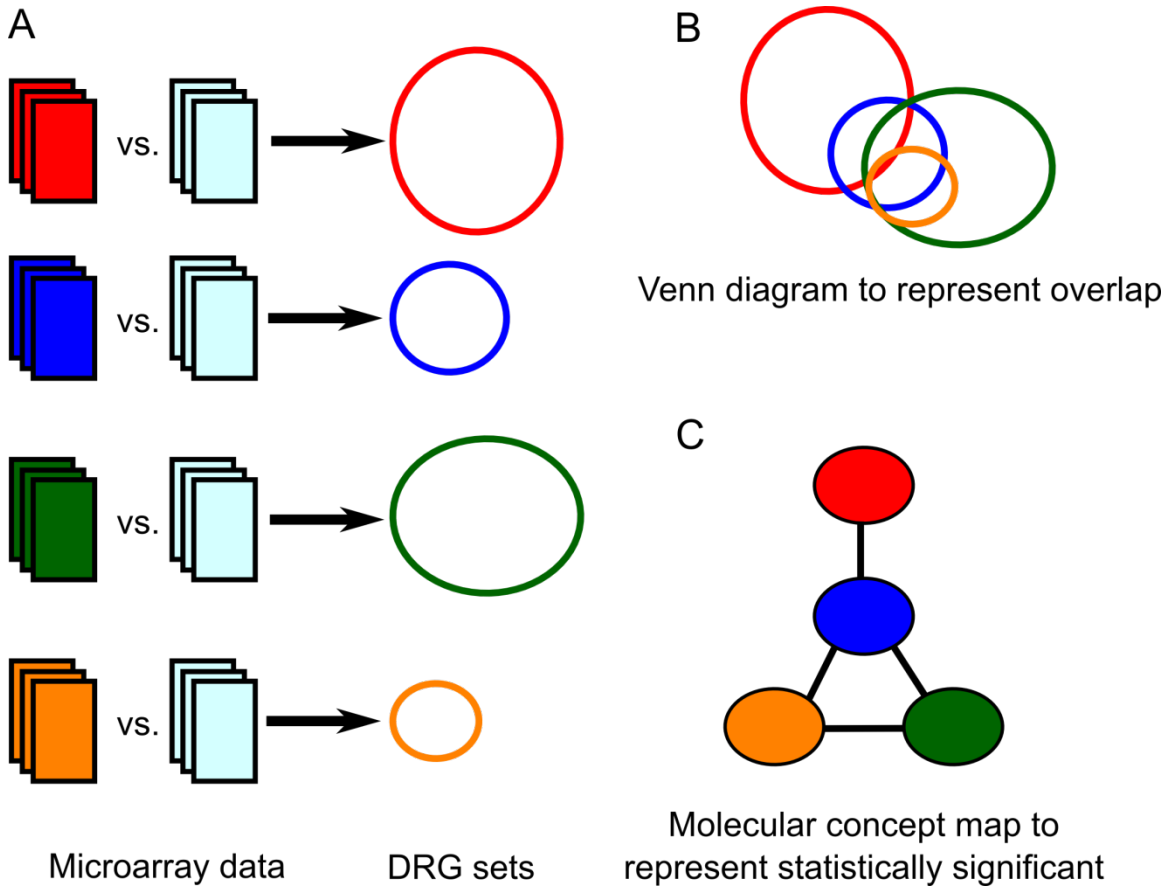

Fig. S7. A schematic illustration of the construction of a molecular concept map. a, Two-group statistical tests are performed to identify the differentially regulated genes (DRGs) for each biologically interesting comparison. **b**, A Venn diagram is a simple means to visualize the extent of overlap among multiple gene sets. However, it easily becomes cluttered when large numbers of gene sets are compared. **c**, A MCM is a simple way to visualize statistically significant overlap among many gene sets. Each node in the MCM represents a gene set, and an edge between two nodes represents a statistically significant overlap between the two gene sets. In this study, we used Fisher's Exact test to determine the statistical significance of gene set overlap.

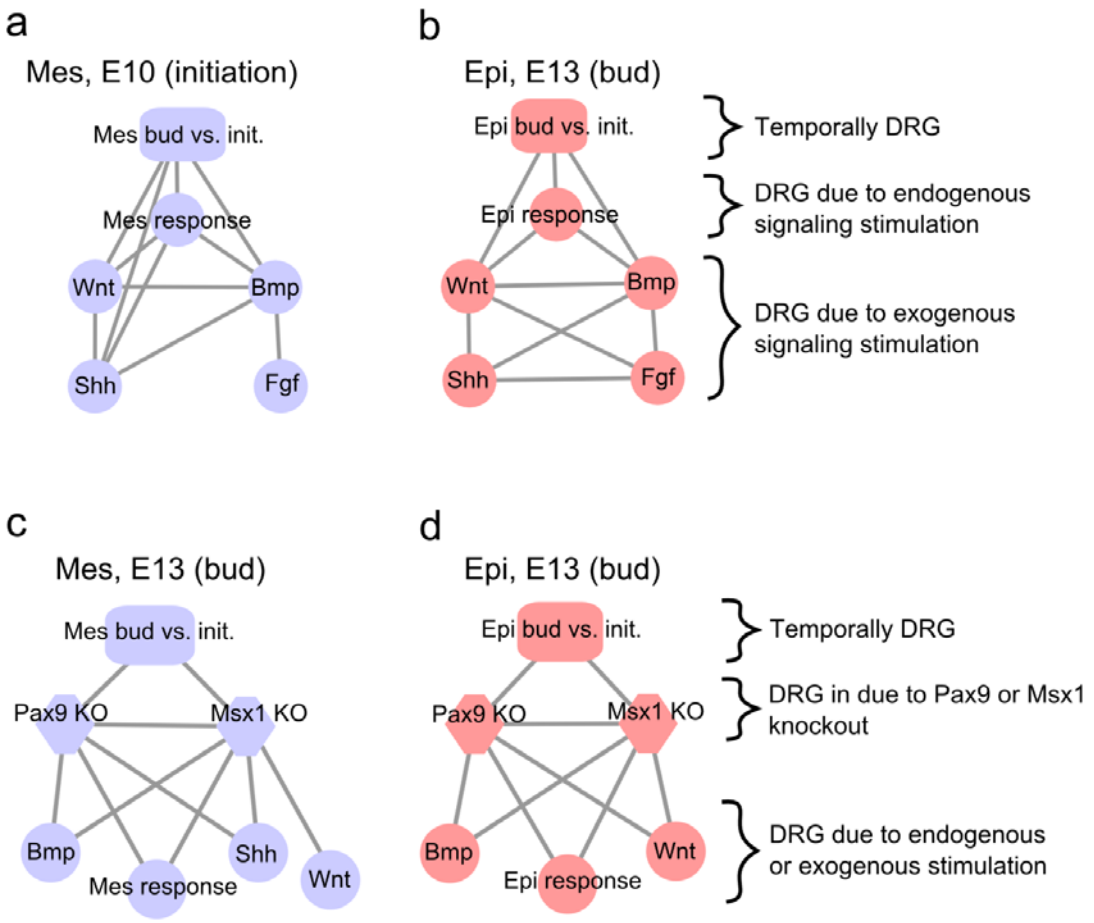

Fig. S8. Molecular concept maps showing significant overlap between various sets of differentially regulated genes. **a** and **b,** MCMs of DRG sets related to mesenchymal (initiation-stage) and epithelial (bud-stage) induction: endogenous tissue inductive response (Mes response; Epi response), temporally differentially expressed genes (Mes bud-stage compared to initiation-stage; Epi bud-stage compared to initiation-stage), and tissue response to exogenous stimulation of Wnt, Bmp, Shh, and Fgf signaling (Wnt, Bmp, Fgf, Shh). **c** and **d**, MCMs showing significant overlap between Pax9 and Msx1 target genes in epithelium and mesenchyme, and their association with other gene sets related to inductive signaling. Edges displayed in (a) and (b) are hidden in (c) and (d) for clarity. Each edge demonstrates a statistically significant overlap of gene sets (FDR < 0.05, χ^2; odds ratio > 2). This analysis, based on three independent datasets, suggests that Wnt and Bmp signaling most closely recapitulates the effect of endogenous signaling during early odontogenesis. N = 3 biological microarray replicates per sample.

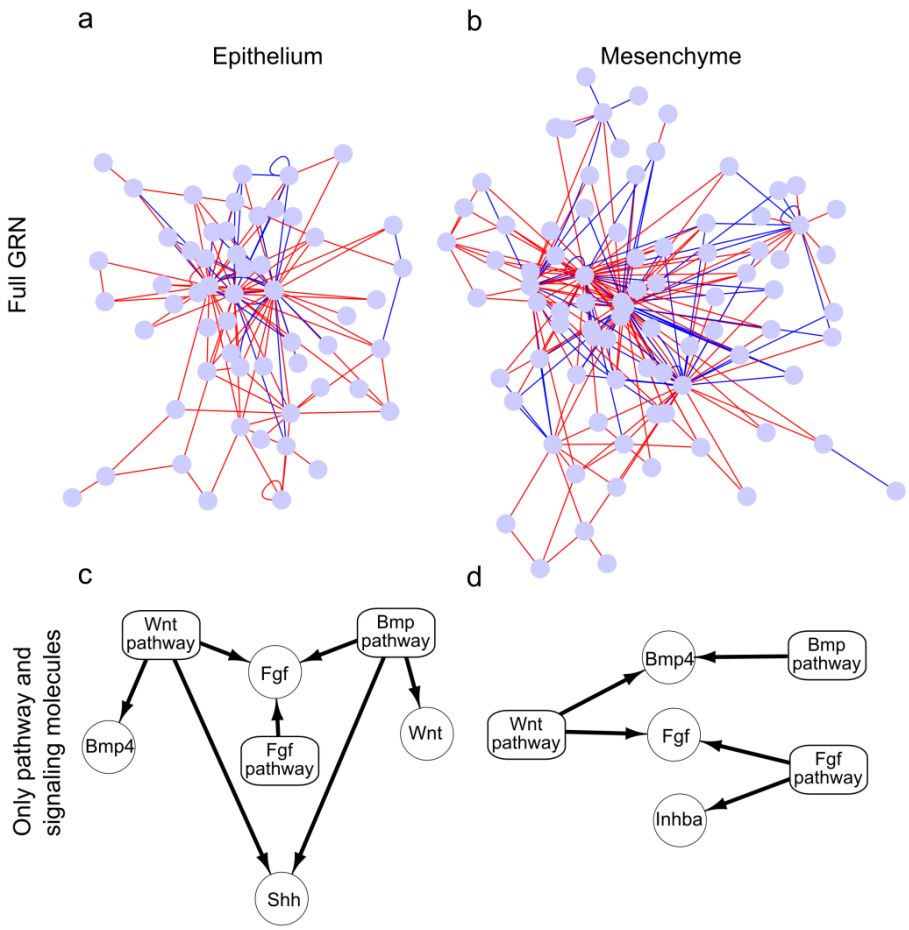

Fig. S9. Epithelial and mesenchymal gene regulatory networks reconstructed using our probabilistic method. **a** and **b,** "Full GRNs" (gene regulatory networks) for E13.5 dental epithelium and E13.5 dental mesenchyme are shown. Each node represents a pathway, a set of extracellular signaling molecules, or a gene. Red and blue edges represent activation and inhibition, respectively. **c** and **d**, A small subnetwork of the 'Full GRN' for each tissue compartment is extracted to show the regulatory relationships between key signaling pathways and extracellular signaling molecules. We focus here on the control of major extracellular signaling molecules by the Wnt, Bmp, Fgf, Shh and Activin pathways. Shh and Activin pathways are not shown in these two figures because there is no evidence yet that they directly control the expression of other signaling molecules. In the epithelium, both Bmp4 and Wnt ligands are under the control of the Wnt and Bmp pathways, respectively. Together, the Wnt and Bmp pathways jointly control mesenchymal *Bmp4*, whereas Wnt is not under the control of either the Wnt or Bmp pathway and is not strongly expressed in dental mesenchyme at these stages.

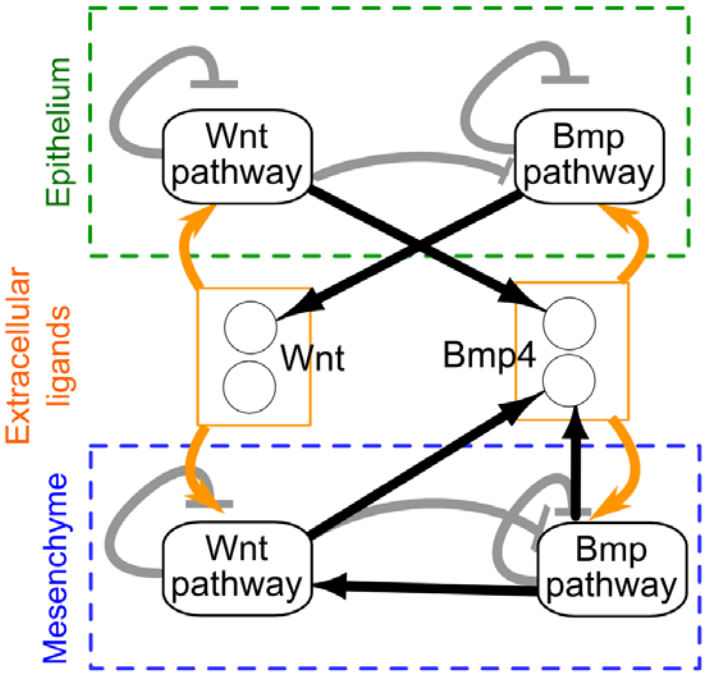

Fig. S10. An expanded view of the Wnt-Bmp feedback circuit. Each edge in this GRN is inferred from a combination of experimental data from the literature and this study. We have developed an interactive web-based network viewer to allow users to interrogate the evidence associated with each edge. The network viewer is part of ToothCODE: http://compbio.med.harvard.edu/ToothCODE/network.php. For clarity, we have removed all pathway-to-pathway edges in Fig. 2 of the main text. The inhibitory edges associated with pathway-to-pathway connections may contribute to the fine tuning of signaling activity, whereas Wnts and Bmp4 act as the major mediators of epithelial-mesenchymal interaction.

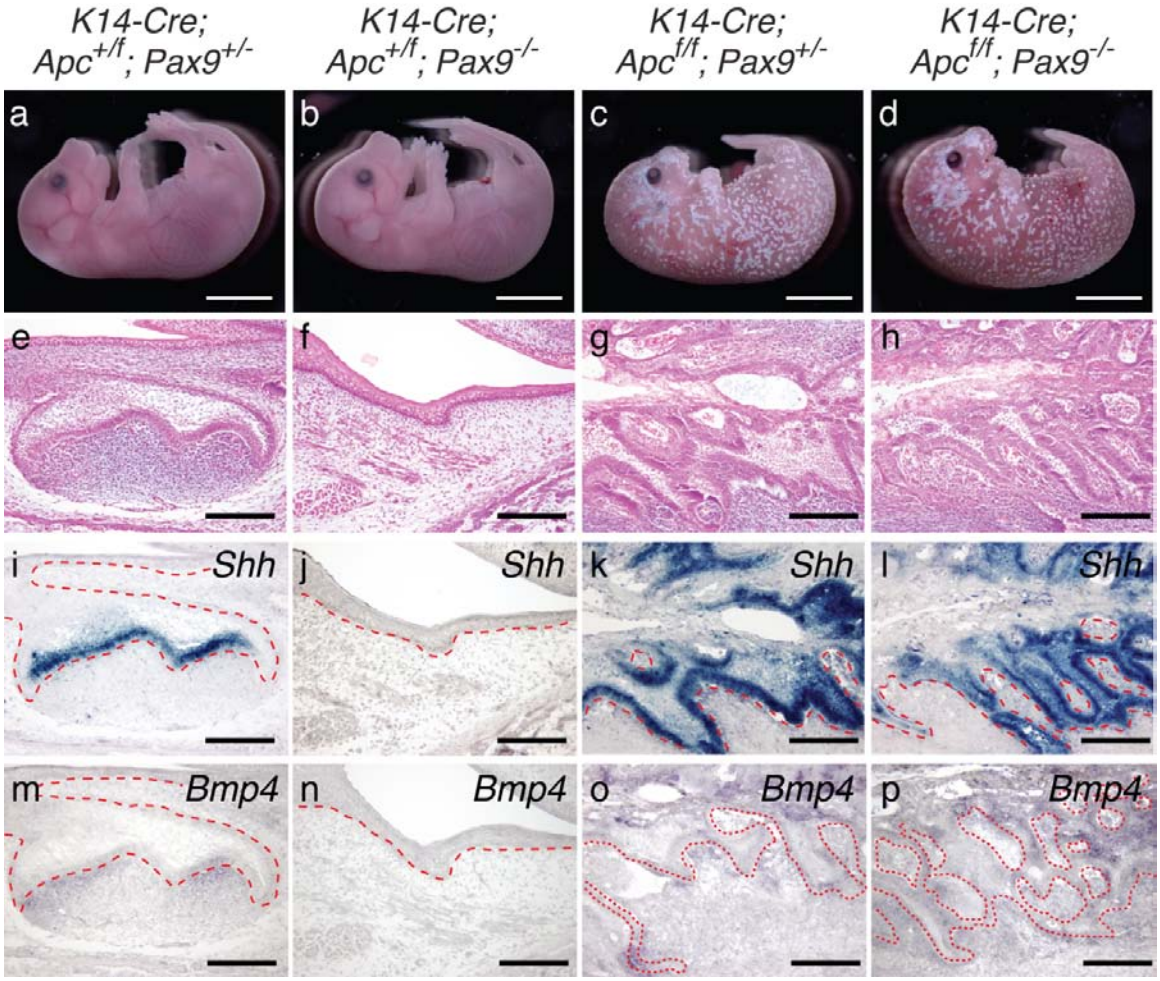

Fig. S11. Constitutive canonical Wnt signaling induces supernumerary tooth development in the absence of *Pax9*. **a-d,** Whole mount phenotypes of E17.5 control (*K14-Cre*; *Apc*$^{+/f}$; *Pax9*$^{+/-}$), *Pax9*-null (*K14-Cre*; *Apc*$^{+/f}$; *Pax9*$^{-/-}$), epithelial *Apc* loss-of-function (*K14-Cre*; *Apc*$^{f/f}$; *Pax9*$^{+/-}$), and compound epithelial *Apc* loss-of-function; *Pax9*-null mutants (*K14-Cre*; *Apc*$^{f/f}$; *Pax9*$^{-/-}$). **e-h,** Hematoxylin, eosin, and alcian blue stained molar region sagittal sections from respective E17.5 embryos. **i-l,** *Shh* expression marks differentiating ameloblasts. **m-p,** *Bmp4* expression in epithelial and mesenchymal compartments of *K14-Cre*; *Apc*$^{f/f}$; *Pax9*$^{+/-}$ and *K14-Cre*; *Apc*$^{f/f}$; *Pax9*$^{-/-}$. Scale bars: 5 mm (**a-d**) and 200 μm (**e-p**). N = 3 nonadjacent sections.

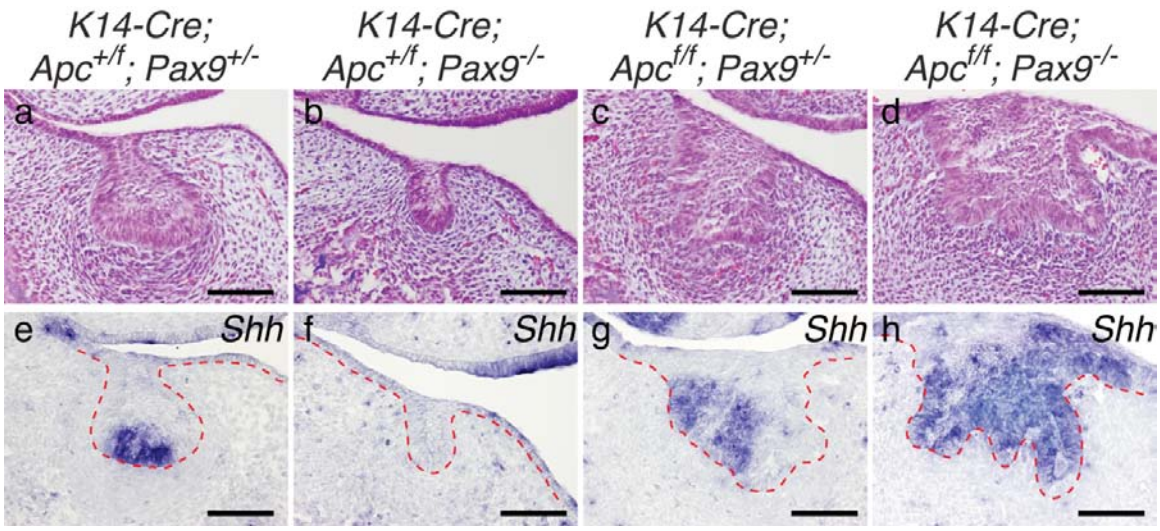

Fig. S12. Coronal sections of molar regions from control, *Pax9*-null, epithelial *Apc* loss-of-function, and compound epithelial *Apc* loss-of-function; *Pax9*-null mutants at E14.0. a-d, Hematoxylin, eosin, and alcian blue stained sections. **e-h,** Increased *Shh* expression in epithelial *Apc* loss-of-function (*K14-Cre*; *Apc*$^{f/f}$; *Pax9*$^{+/-}$) and compound epithelial *Apc* loss-of-function; *Pax9*-null mutants (*K14-Cre*; *Apc*$^{f/f}$; *Pax9*$^{-/-}$) mutants compared to control mice (*K14-Cre*; *Apc*$^{+/f}$; *Pax9*$^{+/-}$). Scale bars: 100 μm. N = 3 nonadjacent sections.

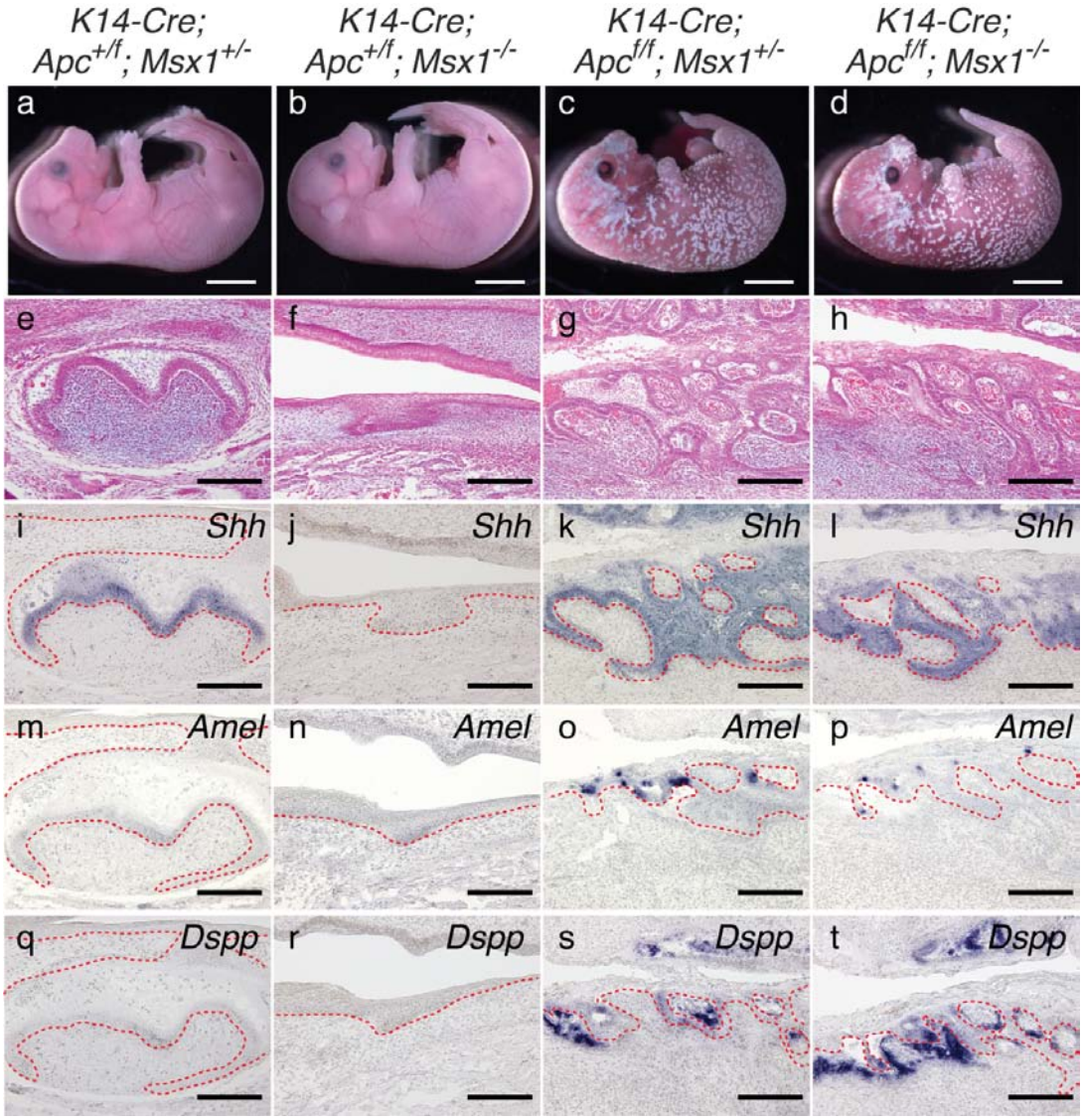

Fig. S13. Constitutive canonical Wnt signaling induces supernumerary tooth formation in the absence of *Msx1*. **a-d,** Whole mount phenotypes of E17.5 control (*K14-Cre*; *Apc*$^{+/f}$; *Msx1*$^{+/-}$), *Msx1*-null (*K14-Cre*; *Apc*$^{+/f}$; *Msx1*$^{-/-}$), epithelial *Apc* loss-of-function (*K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{+/-}$), and compound epithelial *Apc* loss-of-function; *Msx1*-null mutants (*K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{-/-}$). **e-h,** Hematoxylin, eosin, and alcian blue stained sagittal sections of molar regions from respective E17.5 embryos. **i-l,** *Shh* expression marks differentiating ameloblasts. **m-p,** *Amel* expression in *K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{+/-}$ and *K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{-/-}$ marks differentiating ameloblasts. **q-t,** *Dspp* expression in *K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{+/-}$ and *K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{-/-}$ marks differentiating odontoblasts. Scale bars: 5 mm (**a-d**) and 200 μm (**e-t**). N = 3 nonadjacent sections.

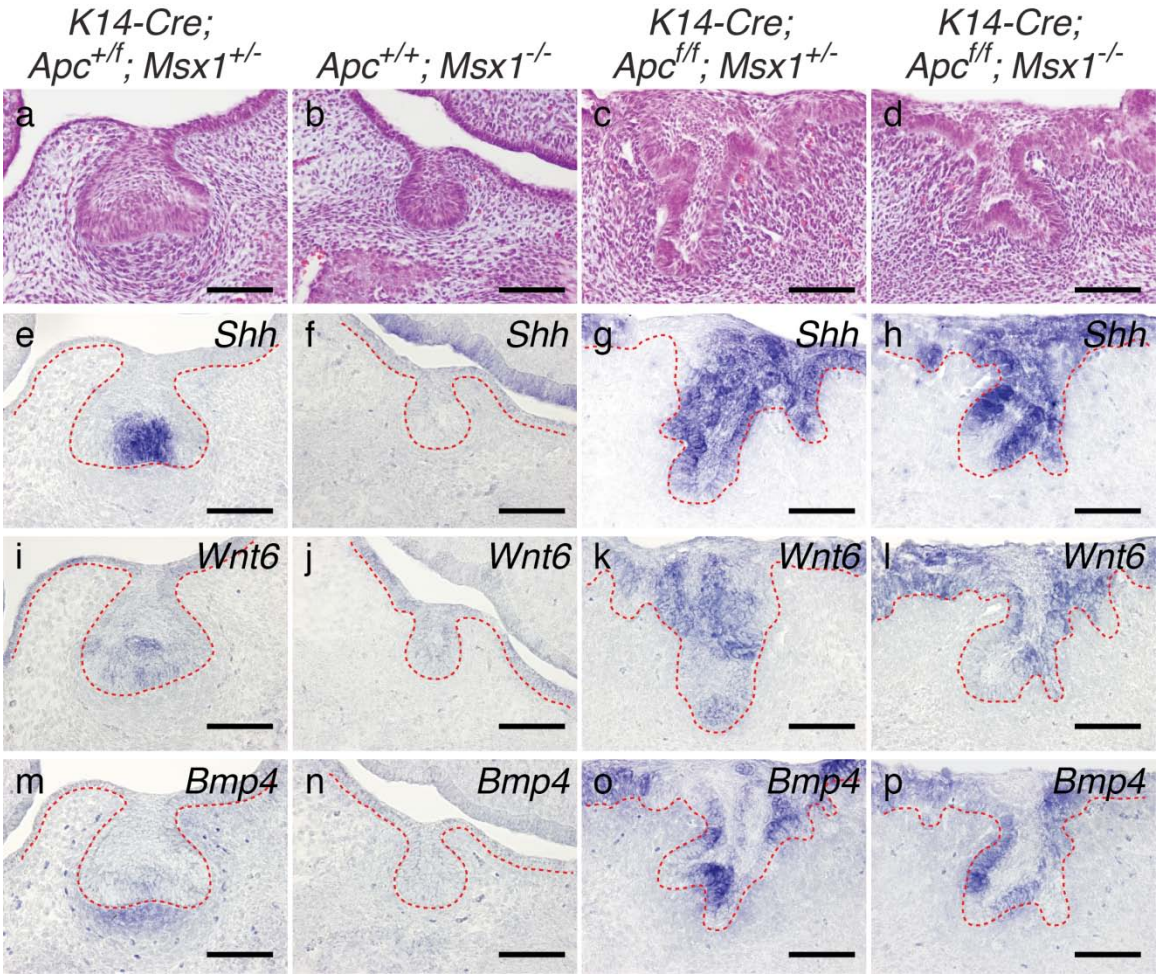

Fig. S14. Coronal sections of molar regions from control, *Msx1*-null, epithelial *Apc* loss-of-function, and compound epithelial *Apc* loss-of-function; *Msx1*-null mutants at E14.0. a-d, Hematoxylin, eosin, and alcian blue stained sections. **e-h,** Increased *Shh* expression in epithelial *Apc* loss-of-function (*K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{+/-}$) and compound epithelial *Apc* loss-of-function; *Msx1*-null (*K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{-/-}$) mutants compared to control mice (*K14-Cre*; *Apc*$^{+/f}$; *Msx1*$^{+/-}$). **i-l,** Increased *Wnt6* expression in epithelial *Apc* loss-of-function (*K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{+/-}$), and compound epithelial *Apc* loss-of-function; *Msx1*-null (*K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{-/-}$) mutants compared to control mice. **m-p,** Increased epithelial *Bmp4* expression in epithelial *Apc* loss-of-function (*K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{+/-}$), and compound epithelial *Apc* loss-of-function; *Msx1*-null (*K14-Cre*; *Apc*$^{f/f}$; *Msx1*$^{-/-}$) mutants compared to control mice.. Decreased mesenchymal *Bmp4* expression in *Msx1*-null mice compared to control mice. Scale bars: 100 μm. N = 3 nonadjacent sections.

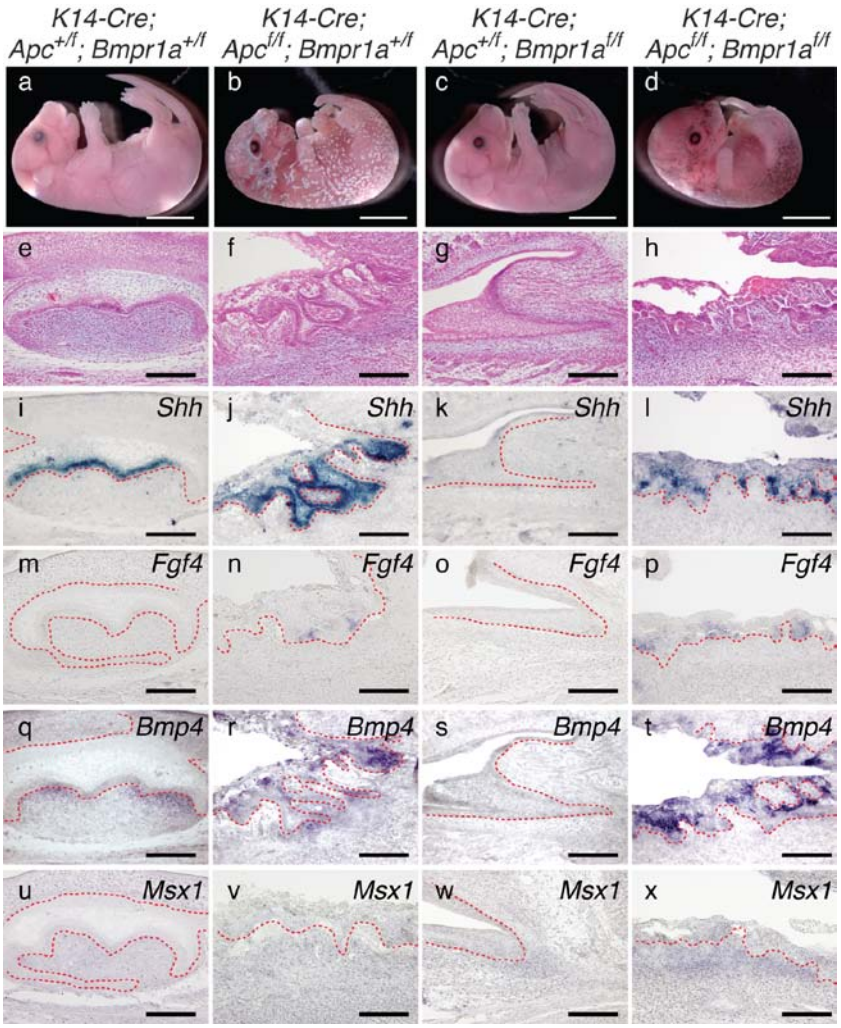

Fig. S15. Constitutive canonical Wnt signaling fails to induce tooth formation in the absence of epithelial *Bmpr1a*. a-d, Whole mount phenotypes of E17.5 control (*K14-Cre*; *Apc*$^{+/f}$; *Bmpr1a*$^{+/f}$), epithelial *Apc* loss-of-function (*K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{+/f}$), epithelial *Bmpr1a* loss-of-function (*K14-Cre*; *Apc*$^{+/f}$; *Bmpr1a*$^{f/f}$), and compound epithelial *Apc* loss-of-function; *Bmpr1a* loss-of-function (*K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{f/f}$). **e-h,** Hematoxylin, eosin, and alcian blue stained sagittal molar sections of E17.5 embryos. **i-l,** Increased *Shh* expression in *K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{+/f}$ and *K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{f/f}$. **m-p,** Increased *Fgf4* expression in *K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{+/f}$ and *K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{f/f}$. **q-t,** Increased epithelial *Bmp4* expression in *K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{+/f}$ and *K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{f/f}$, and decreased mesenchymal *Bmp4* expression in *K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{f/f}$. **u-x,** *Msx1* expression in *K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{f/f}$. Scale bars: 5 mm (**a-d**) and 200 μm (**e-x**). N = 3 nonadjacent sections.

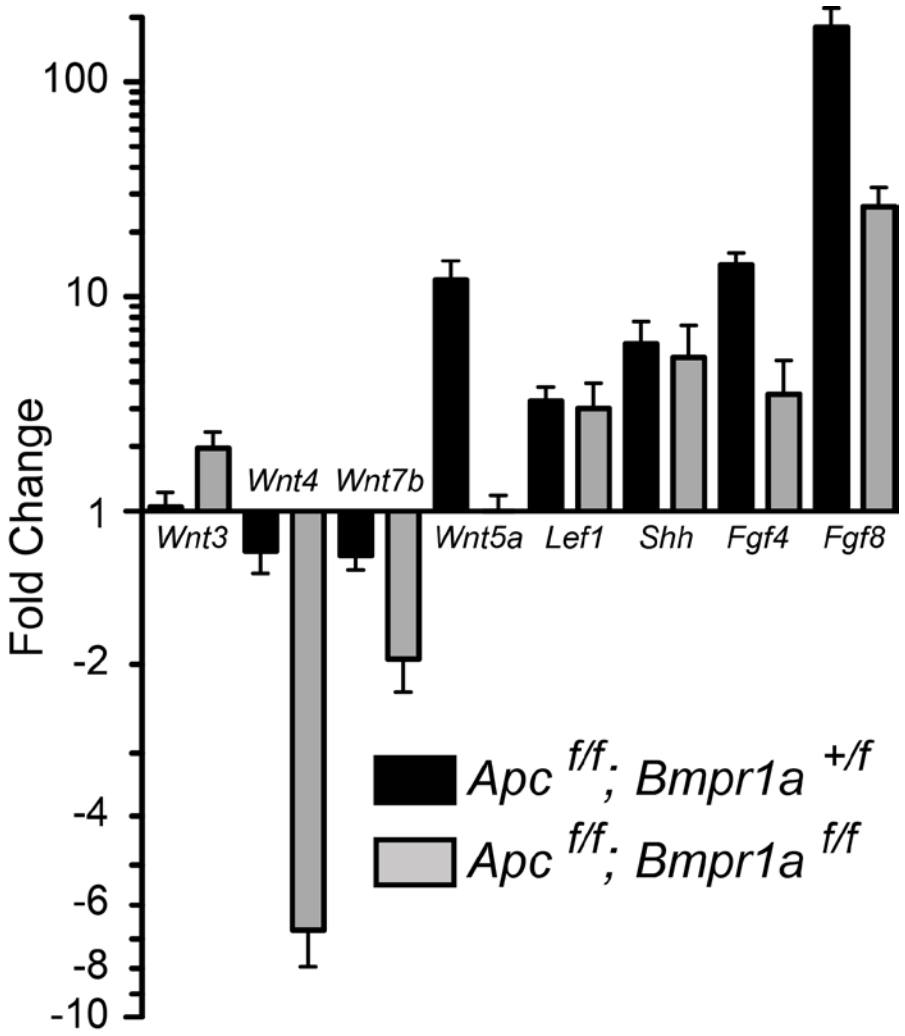

Fig. S16. qRT-PCR expression analysis at E14.5 in the isolated epithelium of epithelial *Apc* loss-of-function and of compound epithelial *Apc* loss-of-function; *Bmpr1a* loss-of-function mutants. Fold changes were determined by comparison with the dental epithelium of control samples (*Apc*$^{+/f}$*; Bmpr1a*$^{+/f}$). The expression of *Wnt4*, *Wnt7b*, *Wnt5a*, *Fgf4*, and *Fgf8* was reduced in compound epithelial *Apc* loss-of-function; *Bmpr1a* loss-of-function mutants (*Apc*$^{f/f}$*; Bmpr1a*$^{f/f}$) compared to epithelial *Apc* loss-of-function mutants (*Apc*$^{f/f}$*; Bmpr1a*$^{+/f}$). *Wnt3* was increased 2-fold in compound epithelial *Apc* loss-of-function; *Bmpr1a* loss-of-function mutants (*Apc*$^{f/f}$*; Bmpr1a*$^{f/f}$) compared to epithelial *Apc* loss-of-function mutants (*Apc*$^{f/f}$*; Bmpr1a*$^{+/f}$). *Lef1* and *Shh* expression was similar in compound epithelial *Apc* loss-of-function; *Bmpr1a* loss-of-function mutants (*Apc*$^{f/f}$*; Bmpr1a*$^{f/f}$) and epithelial *Apc* loss-of-function mutants (*Apc*$^{f/f}$*; Bmpr1a*$^{+/f}$). Data are mean ± s.d. (*n*=6; 2 biological replicates run in technical triplicate), normalized to *Hprt*.

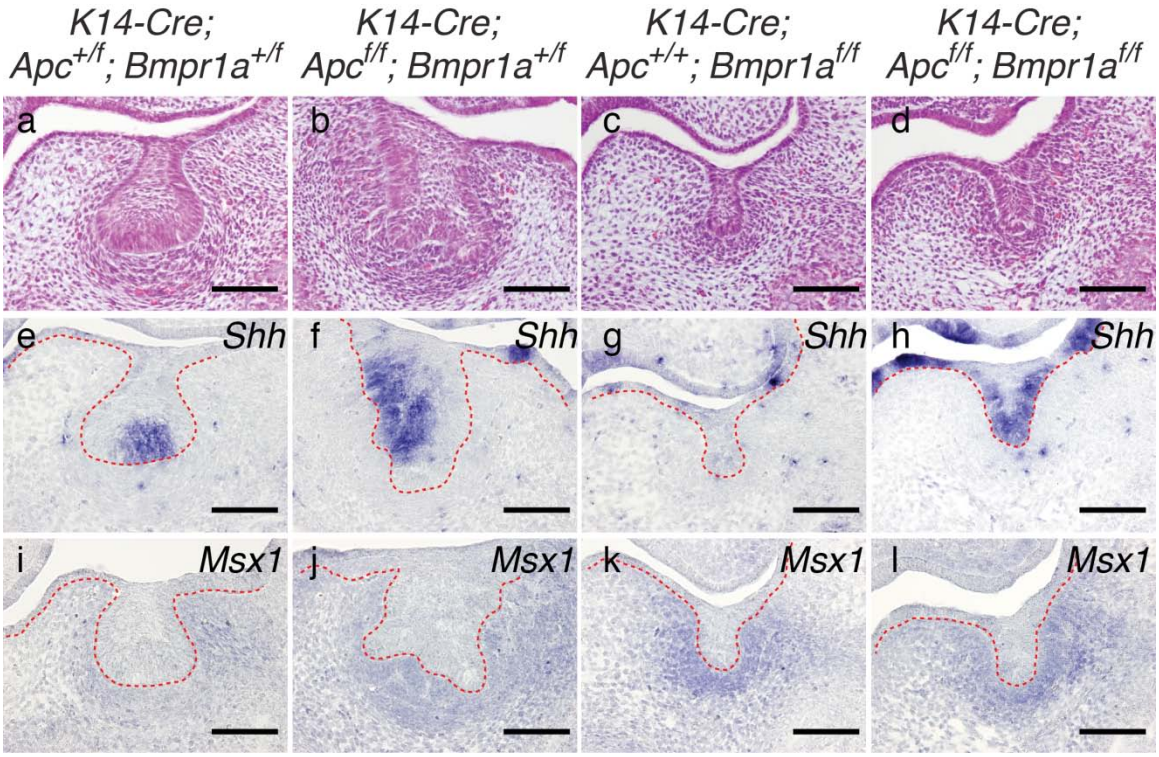

Fig. S17. Coronal sections of molar regions from control, epithelial *Apc* loss-of-function, epithelial *Bmpr1a* loss-of-function, and compound epithelial *Apc* loss-of-function; *Bmpr1a* loss-of-function mutants at E14.0. a-d, Hematoxylin, eosin, and alcian blue stained sections. **e-h,** Increased *Shh* expression in epithelial *Apc* loss-of function (*K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{+/f}$) and compound epithelial *Apc* loss-of-function; *Bmpr1a* loss-of-function (*K14-Cre*; *Apc*$^{f/f}$; *Bmpr1a*$^{f/f}$) mutants compared to control mice (*K14-Cre*; *Apc*$^{+/f}$; *Bmpr1a*$^{+/f}$). **i-l,** *Msx1* expression in dental mesenchyme is nominally similar in the different genotypes. Scale bars: 100 μm. N = 3 nonadjacent sections.

Table S1. Comparison of BITE-IT and microarray data from this study

Molecule	BITE-IT records					Microarray data					AUROC	avg exp	CV
	Epi.init	Mes.init	Epi.bud	Mes.bud	Epi.cap.EK	Epi.init	Mes.init	Epi.bud	Mes.bud	Epi.cap.EK			
	Differentially Expressed					Differentially Expressed							
Inhbb	-	+	-	+	-	4.7	4.6	5.8	4.7	5.6	0.0	5.1	0.1
Dlx3	-	-	-	+	-	10.8	9.0	10.7	8.8	10.9	0.0	10.0	0.1
Notch3	+	-	+	-	-	8.7	9.0	8.1	9.5	9.1	0.0	8.9	0.1
Ptch1	+	+	+	+	-	8.4	8.5	8.0	8.1	8.8	0.0	8.4	0.0
Rab23	+	+	+	+	-	6.6	6.7	6.4	6.5	7.0	0.0	6.6	0.0
Sema3f	+	+	+	-	+	10.3	9.8	10.7	10.2	9.6	0.5	10.1	0.0
Tnc	-	+	-	-	-	5.7	7.4	7.9	11.0	6.5	0.5	7.7	0.3
Gli1	-	+	+	+	-	5.5	5.7	5.2	6.0	5.7	0.7	5.6	0.0
Msx2	-	+	+	-	+	6.0	6.3	5.6	5.6	7.8	0.7	6.3	0.1
Notch1	+	-	+	-	-	8.2	6.4	8.7	7.7	9.8	0.7	8.2	0.2
Tgfb1	-	-	+	-	+	4.5	4.9	4.8	4.9	5.9	0.7	5.0	0.1
Wnt5a	-	+	-	+	-	8.3	8.8	7.0	10.0	10.6	0.7	9.0	0.2
Fgf10	-	-	-	+	-	6.8	8.8	5.0	8.4	6.1	0.8	7.0	0.2
Axin2	-	+	+	+	+	8.9	9.1	9.0	9.8	9.9	1.0	9.3	0.0
Barx1	-	+	-	+	-	8.4	11.5	7.1	10.7	6.9	1.0	8.9	0.2
Bmp4	+	+	-	+	+	9.5	9.2	7.1	10.6	10.5	1.0	9.4	0.2
Bmp6	-	-	-	+	-	4.0	4.4	4.5	6.1	3.5	1.0	4.5	0.2
Bmp7	+	-	+	-	+	8.5	5.7	7.2	5.3	9.5	1.0	7.2	0.3
Sostdc1	+	+	+	+	-	9.1	7.7	10.2	10.8	7.7	1.0	9.1	0.2
Edar	+	-	+	-	+	7.7	4.8	9.1	5.6	9.6	1.0	7.4	0.3
Egr1	-	-	-	+	-	9.1	8.8	10.5	12.5	11.3	1.0	10.4	0.1
Fgf8	+	-	-	-	-	9.2	4.9	5.8	4.4	5.0	1.0	5.9	0.3
Fgf9	+	-	+	-	+	7.2	5.2	6.0	5.2	7.9	1.0	6.3	0.2
Gli2	+	+	+	+	-	9.1	9.1	9.4	9.7	7.7	1.0	9.0	0.1
Gli3	+	+	-	+	-	11.2	11.2	10.6	11.7	9.2	1.0	10.8	0.1
Lhx6	-	+	-	+	-	4.8	6.7	4.5	7.4	4.6	1.0	5.6	0.2
Lhx8	-	+	-	+	-	6.0	8.8	5.8	9.1	5.6	1.0	7.1	0.2
Msx1	-	+	-	+	-	8.7	11.0	7.9	11.9	7.1	1.0	9.3	0.2
Pax9	-	+	-	+	-	4.7	6.2	4.7	7.1	3.6	1.0	5.3	0.3
Pitx2	+	-	+	-	+	12.6	8.6	13.0	10.4	12.5	1.0	11.4	0.2
Prrx2	-	+	-	+	-	6.5	8.7	5.9	9.1	6.6	1.0	7.4	0.2
Sema3a	+	+	+	+	-	9.5	9.1	8.7	9.8	6.8	1.0	8.8	0.1
Sema3c	+	+	+	-	+	8.1	7.2	7.6	6.8	8.9	1.0	7.7	0.1
Smo	+	+	+	+	-	7.8	8.0	7.8	8.1	7.3	1.0	7.8	0.0
Shh	+	-	+	-	+	10.2	5.2	8.2	4.0	12.4	1.0	8.0	0.4
Timp1	-	-	-	+	-	6.7	7.6	6.8	8.2	6.9	1.0	7.2	0.1
Wnt10a	+	-	+	-	+	6.2	4.6	8.3	4.8	11.5	1.0	7.1	0.4
Wnt10b	+	-	+	-	+	5.1	3.3	6.2	3.5	8.6	1.0	5.3	0.4
Wnt3	-	-	-	-	+	5.7	4.1	5.7	4.1	8.6	1.0	5.6	0.3
Wnt4	+	-	+	-	-	7.6	4.8	8.1	4.8	5.7	1.0	6.2	0.3
Wnt6	+	-	+	-	+	9.2	4.5	8.8	4.7	9.1	1.0	7.3	0.3
Wnt7b	-	-	+	-	-	8.2	4.1	8.8	4.3	6.2	1.0	6.3	0.3
	non-expressed genes					non-expressed genes							
Bmp3	-	-	-	-	-	4.4	3.9	5.4	7.0	5.1	NA	5.2	0.2
Dlx5	-	-	-	-	-	7.0	8.4	7.7	7.9	7.9	NA	7.8	0.1
Dlx6	-	-	-	-	-	4.8	6.9	4.2	6.3	5.0	NA	5.4	0.2
Fgf7	-	-	-	-	-	4.6	4.9	4.5	6.6	4.7	NA	5.1	0.2
Fbln2	-	-	-	-	-	7.1	8.7	7.4	9.9	7.5	NA	8.1	0.1
Lfng	-	-	-	-	-	5.8	4.7	5.5	4.4	6.1	NA	5.3	0.1
Wt1	-	-	-	-	-	7.0	4.4	4.6	4.3	6.1	NA	5.3	0.2
	Constituively expressed genes					Constituively expressed genes							
Fbln1	+	+	+	+	+	11.3	11.5	11.9	12.0	12.2	NA	11.8	0.0
Lef1	+	+	+	+	+	7.6	7.4	7.8	8.5	8.6	NA	8.0	0.1
Sdc1	+	+	+	+	+	8.6	8.7	8.4	9.1	8.0	NA	8.6	0.0

Table S2. Gene sets used in this study

Collection	set size	Download	Reference
Gene ontology	10,705	13-Dec-10	http://www.geneontology.org/GO.downloads.annotations.shtml
KEGG pathways	214	16-Sep-09	KEGG, from bioconductor package org.Mm.egPATH
MouseCyc pathways	322	14-Dec-10	ftp://ftp.informatics.jax.org/pub/curatorwork/MouseCycDB/pathways_summary.txt
MGI's Mouse phenotype genes	33	14-Dec-10	ftp://ftp.informatics.jax.org/pub/reports/index.html#pheno
FANTOM4 tissue specific TF	21	30-Dec-10	http://fantom.gsc.riken.jp/4/download/Tables/mouse/qRT-PCR/
Custom gene set	90		This study
Total	11,385		

Table S3. Custom gene sets used in this study

Custom gene sets	set size	Description	Evidence
Enamel knot set	1	Genes highly expressed in enamel knot	Manually curated from on literature and our microarray data
Tooth phenotype set	1	Genetic disruption causes tooth phenotype	Manually curated from literature
Signaling pathways	70	Wnt, Bmp, Fgf, Hh and Activin pathway	Manually curated from literature
Histone genes	11	Major and variant histone genes	Manually curated from literature
PRC1 and 2	2	Components of PRC1 and PRC2 complexes	Manually curated from on literature
Core ES cell module	1	Target genes of the core ES pluripotency factors	Manually curated from literature
Polycomb group ES module	1	Target genes of polycomb group factors in ESCs	Manually curated from literature
Myc ES module	1	Target genes of Myc in ESCs	Manually curated from literature
Secreted factors and receptors	2	Genes of secreted factors and receptors	Manually curated from literature

References

1. Du, P., W. A. Kibbe, and S. M. Lin. 2008. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 24: 1547-1548.
2. Lin, S. M., P. Du, W. Huber, and W. A. Kibbe. 2008. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res* 36: e11.
3. Eklund, A. C., L. R. Turner, P. Chen, R. V. Jensen, G. deFeo, A. R. Kopf-Sill, and Z. Szallasi. 2006. Replacing cRNA targets with cDNA reduces microarray cross-hybridization. *Nat. Biotechnol* 24: 1071-1073.
4. Johnson, W. E., C. Li, and A. Rabinovic. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8: 118-127.
5. Sing, T., O. Sander, N. Beerenwinkel, and T. Lengauer. 2005. ROCR: visualizing classifier performance in R. *Bioinformatics* 21: 3940-3941.
6. Mina, M., and E. J. Kollar. 1987. The induction of odontogenesis in non-dental mesenchyme combined with early murine mandibular arch epithelium. *Arch Oral Biol* 32: 123–7.
7. Mendes, P. 2003. Artificial gene networks for objective comparison of analysis algorithms. *Bioinformatics* 19: 122ii-129.