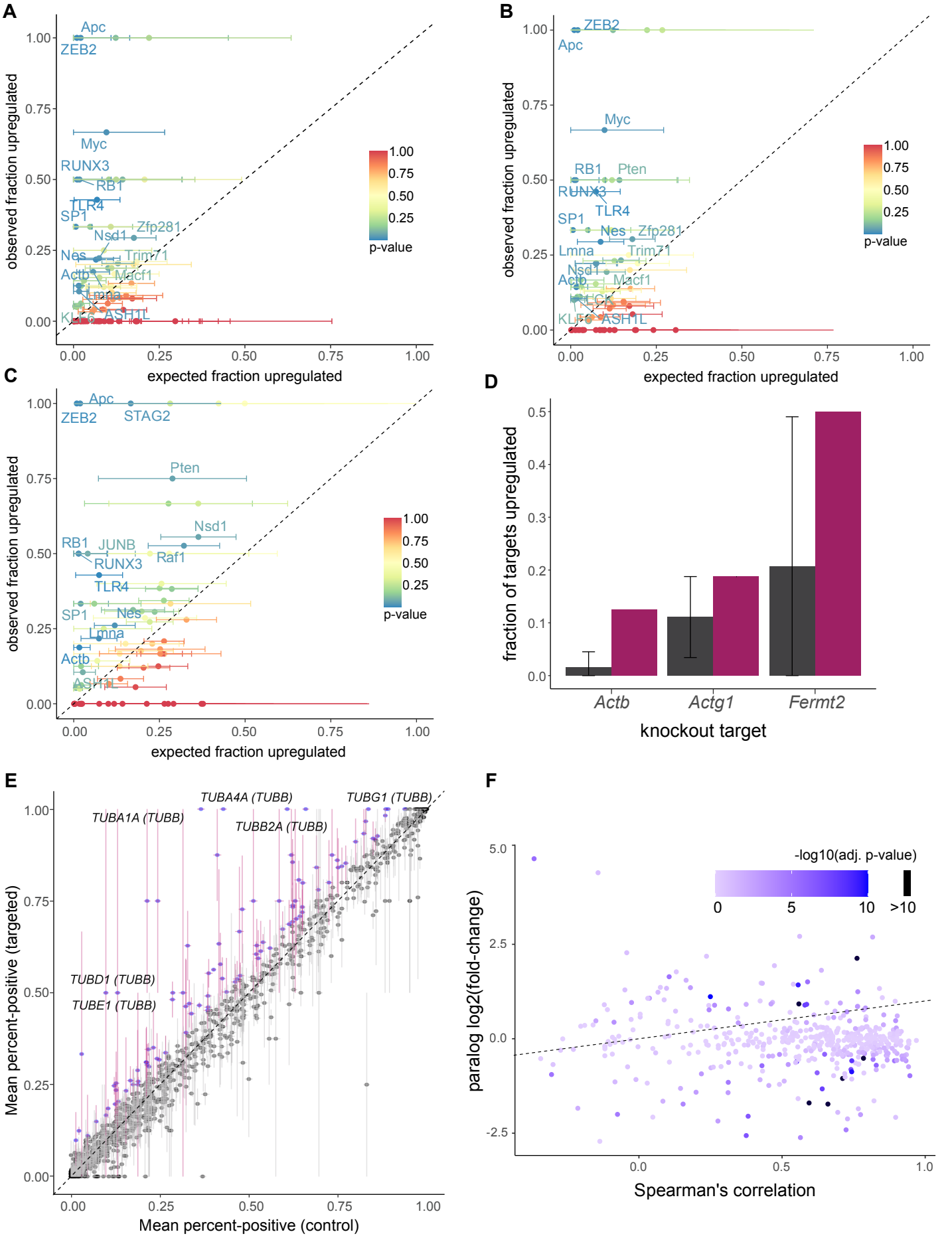


Supplementary Figure 1

S1: Supplementary analysis of paralog upregulation after CRISPR/Cas9-mediated knockouts.

A. Per-knockout-target paralog differential expression results. Per knockout target: expected frequency of paralog upregulation based on bootstrap analysis on the x axis, observed frequency of paralog upregulation on the y axis. Differential expression counted if log2 fold-change > 0.5 and adjusted p-value < 0.05. Error bars represent the standard error of the null expected mean frequency of paralog upregulation across 10000 bootstrap samples (see Methods).

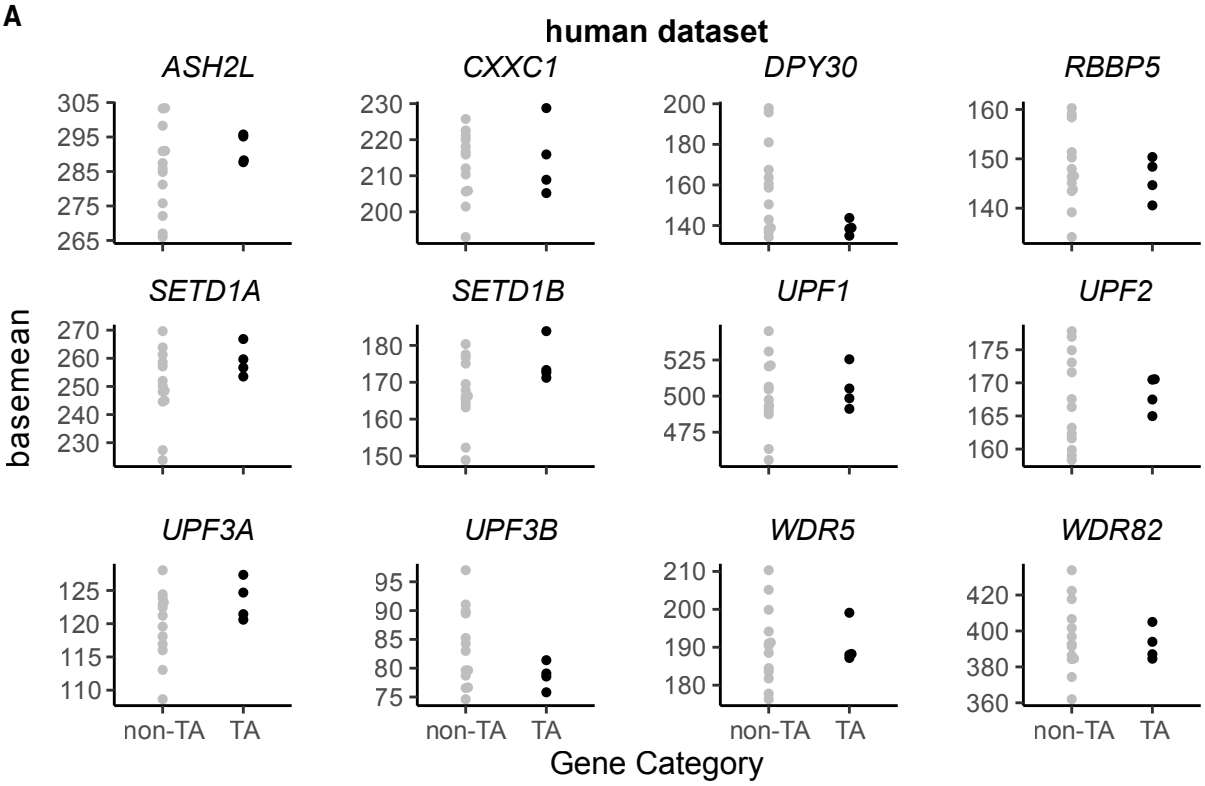
B. Per-knockout-target paralog differential expression results using alternative differential expression inclusion criteria to Fig. 1C. Differential expression counted if log2 fold-change > 0.5, adjusted p-value < 0.05, and if basemean > 10. Per knockout target: expected frequency of paralog upregulation on the x axis, observed frequency of paralog upregulation on the y axis. Error bars represent the standard error of the null expected mean frequency of paralog upregulation across 10000 bootstrap samples (see Methods).

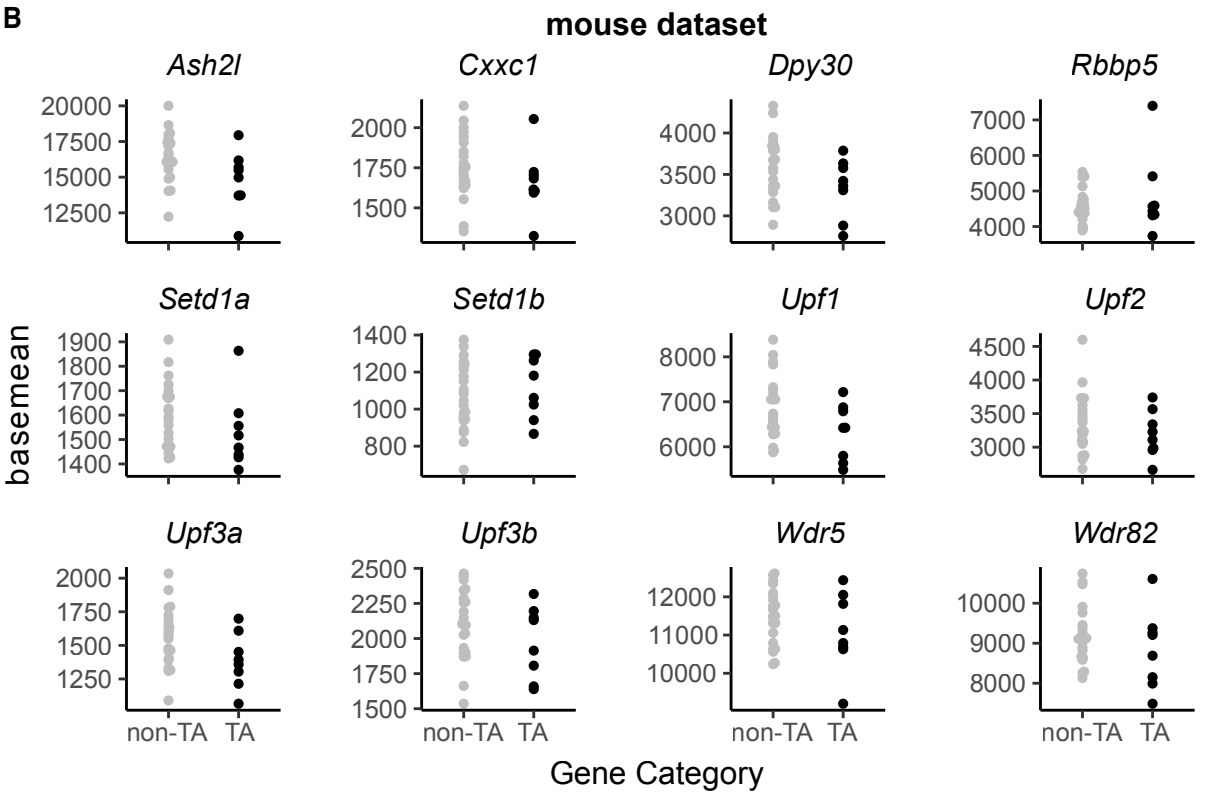
C. Per-knockout-target paralog differential expression results using alternative differential expression inclusion criteria to Fig. 1C. Differential expression counted if log2 fold-change > 0 and if adjusted p-value < 0.05. Per knockout target: expected frequency of paralog upregulation on the x axis, observed frequency of paralog upregulation on the y axis. Error bars represent the standard error of the null expected mean frequency of paralog upregulation across 10000 bootstrap samples (see Methods).

D. Reanalysis of RNA-seq from [14] after three knockouts (see reference manuscript's Fig. 4A) using the paralog upregulation pipeline presented here. Error bars represent standard deviation of bootstrap samples, in gray. Observed fractions of paralogs upregulated, in purple.

E. Perturb-seq-based single-cell paralog expression change after reference gene knockout. Per paralog, percentage of cells positive for that paralog expression in non-template control guide-treated cells on x-axis, percentage of cells positive for that paralog expression in cells treated with guides targeting a reference CRISPR target for which the gene is a paralog. Paralogs ranked in the top-100 of absolute increases per quantification method marked in magenta. All paralogs of all knockouts shown, using cells meeting UMI count inclusion criteria, regardless of whether minimum cell count per guide inclusion criteria were met, listed in Methods.

F. Paralog expression change compared to Spearman's correlation of median expression levels across 54 tissues in GTEx of the paralog with the corresponding knockout target.

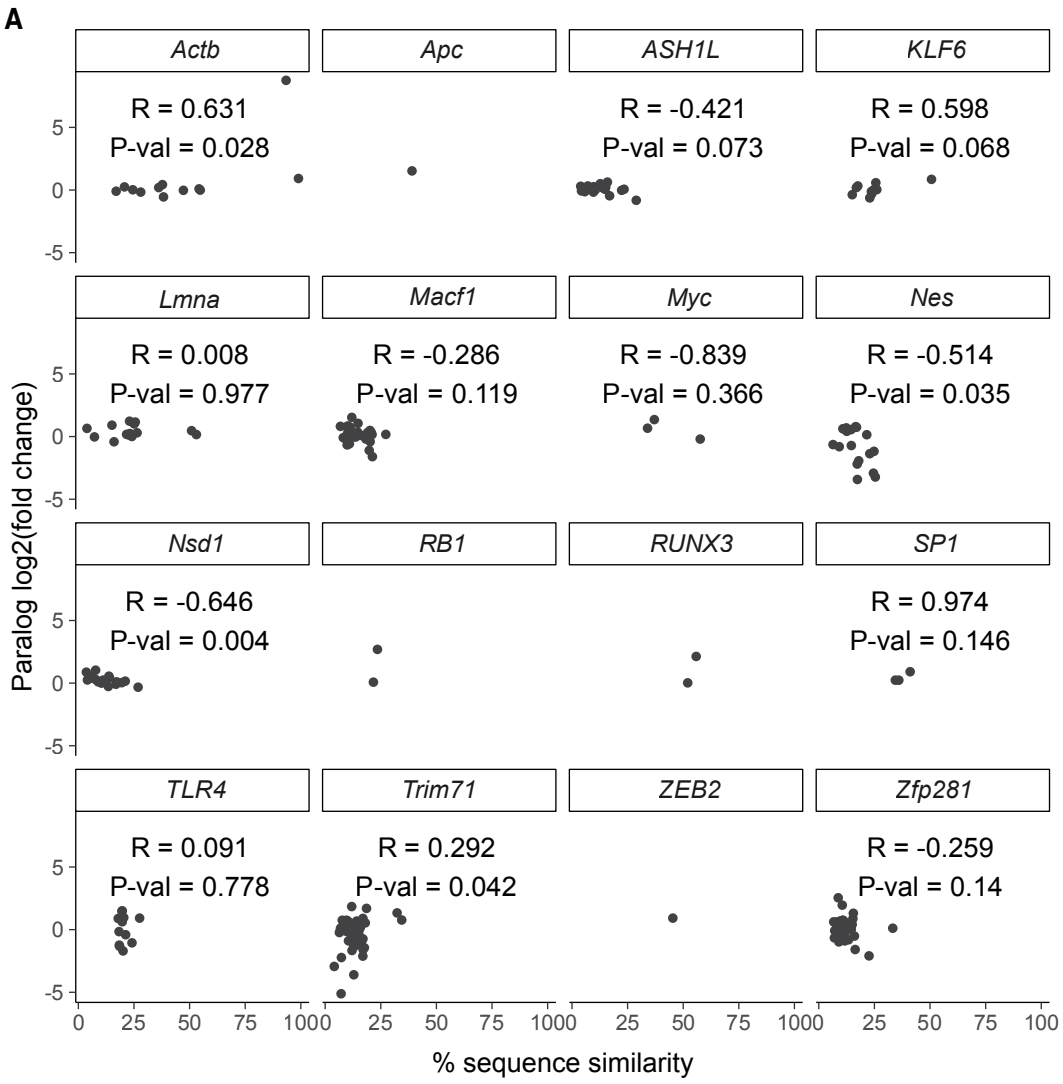


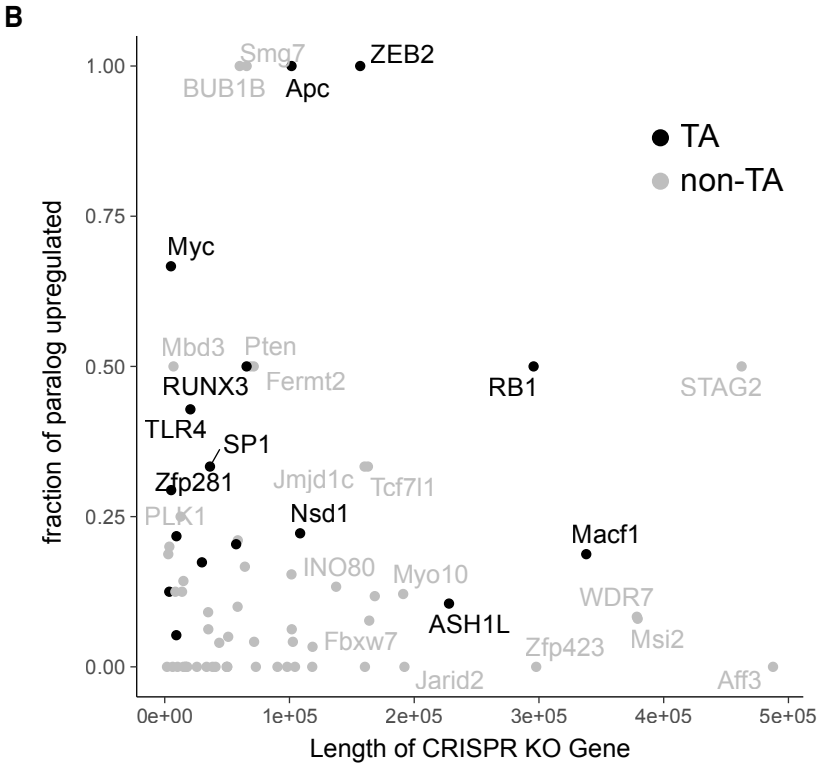


Supplementary Figure 2

S2: Analysis of transcriptional adaptation-associated genes' expression levels with knockout target paralog upregulation

A. Expression levels (DESeq2 basemean) of COMPASS complex and Upf genes important for nonsense-mediated decay in humans, for CRISPR targets in GSE151825 displaying paralog upregulation by transcriptional adaptation ("TA") or not displaying paralog upregulation by transcriptional adaptation ("non-TA").
B. Expression levels (DESeq2 basemean) of COMPASS complex and Upf genes important for nonsense-mediated decay in mice, for CRISPR targets in GSE145653-1 displaying paralog upregulation (with adaptation) or not displaying paralog upregulation (without adaptation).





Supplementary Figure 3

S3: Analysis of whether there are paralog and target gene feature associations with expression change after reference gene knockout

A. Per-knockout-target scatter plots of each paralog's sequence homology with the knockout target versus $\log_2$ fold-change of paralog after reference gene knockout.
B. Per-knockout-target scatter plots of each knockout target gene length with the paralog upregulation frequency after knockout.

A

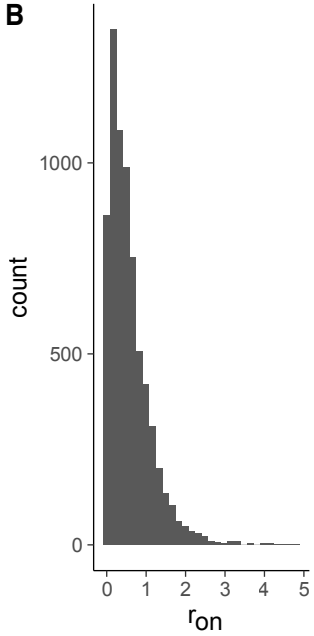
Parameter	Description
r_{on}	Activation rate of an allele
r_{off}	Inactivation rate of an allele
r_{prod}	Production rate of mRNA from an active allele
r_{deg}	Degradation rate of mRNA
r_{add}^{NITC}	Additional activation of an allele (A or A') upon nonsense-induced transcriptional compensation caused by product $A_{nonsense}$
$r_{add}^{A,B}$, $r_{add}^{A',B}$	Activation of B by A or A' to one of two respective B active states specified in the model
d	Factor by which the mRNA production rate of B is lower in the A'-directed B active state than in A-directed B active state
n	Hill coefficient
k	Dissociation constant of the Hill function

$$\text{NITC ratio} = \frac{r_{on}^{A}}{r_{add}^{NITC}}$$

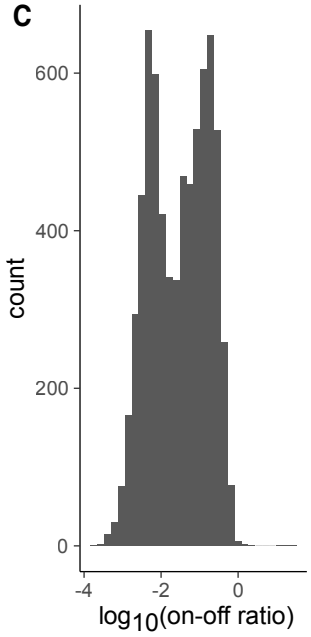
$$\text{activation ratio} = \frac{r_{add}^{A,B}}{r_{add}^{A',B}}$$

$$\text{production ratio} = d$$

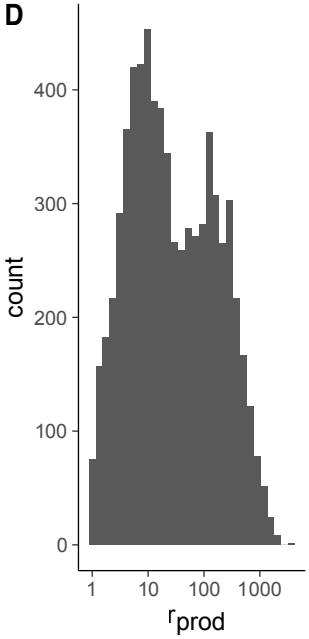
B



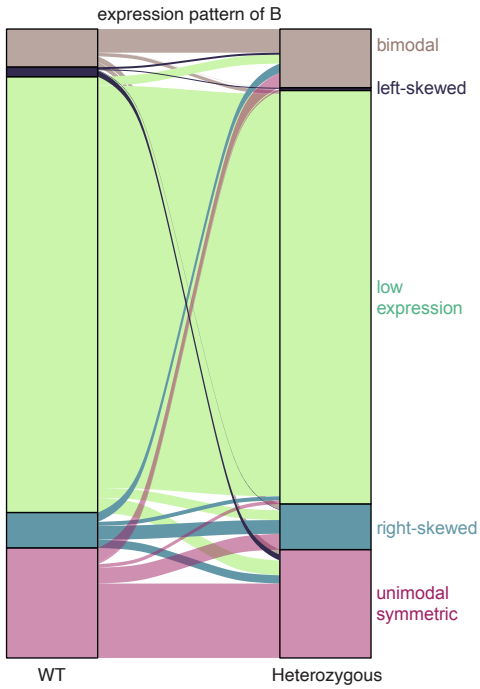
C



D



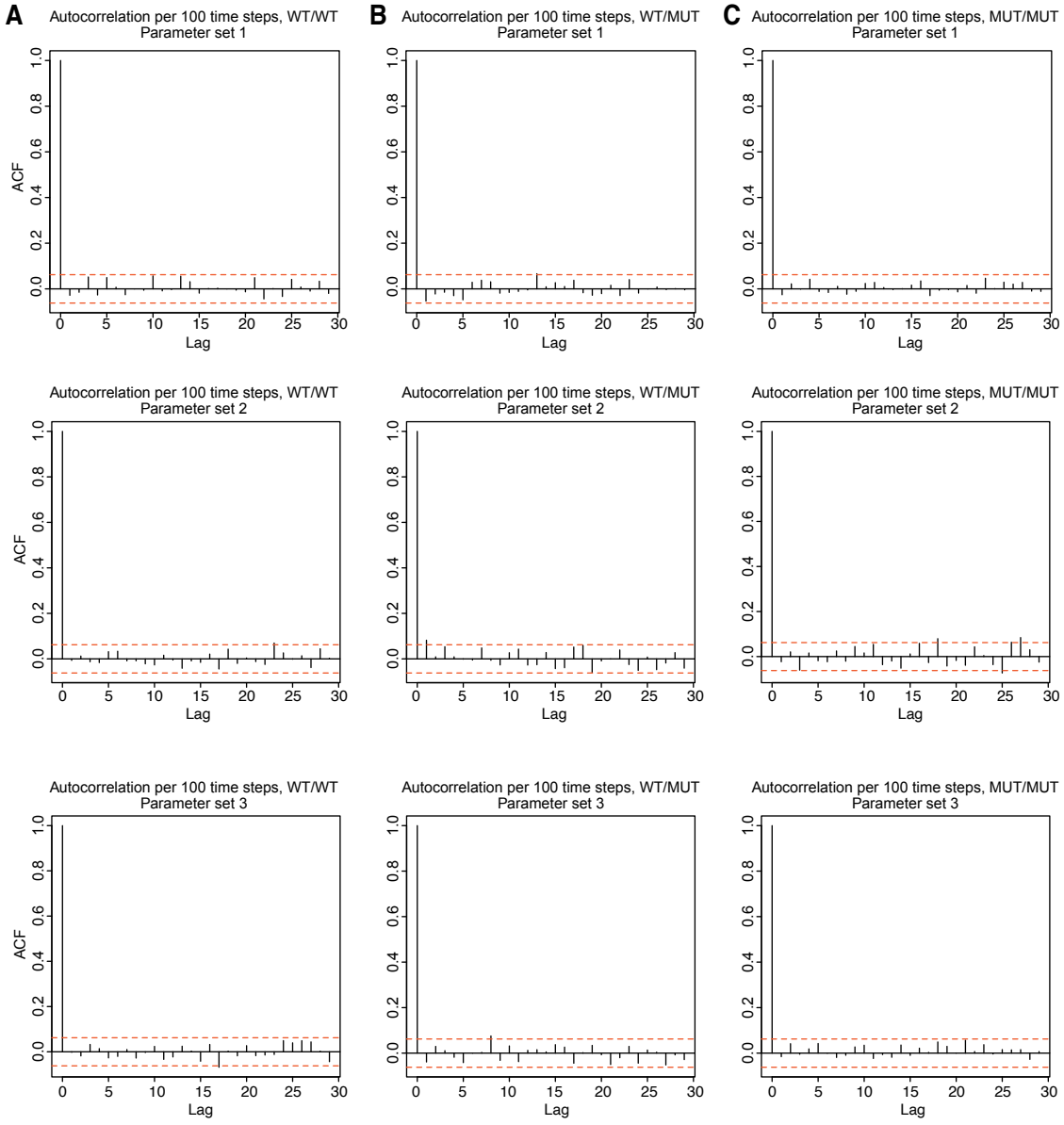
E



Supplementary Figure 4

S4: Simulation transcriptional burst kinetics summary and resulting parameter set classifications, before and after mutation

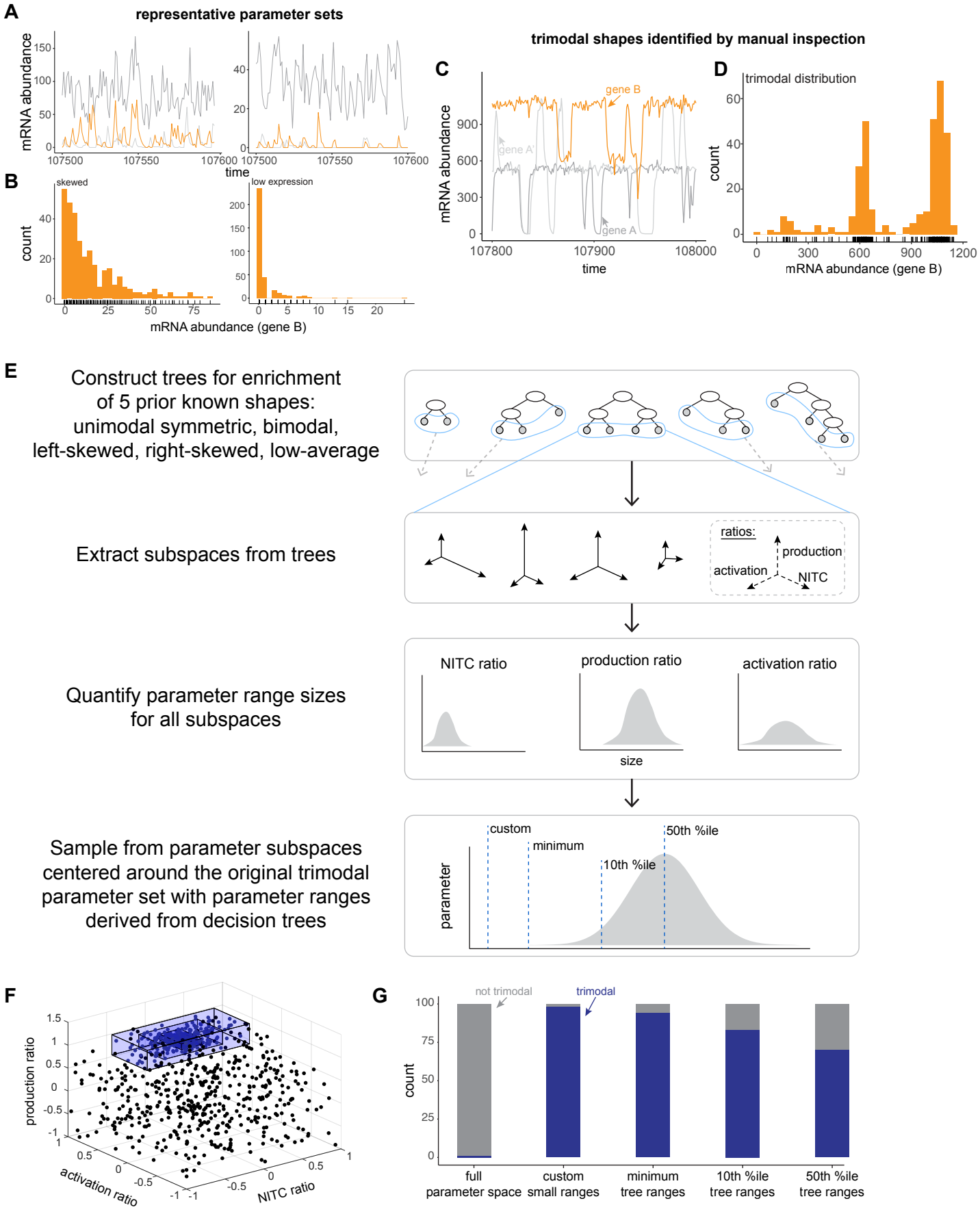
A. (left) Parameter descriptions and key parameter ratios for the model described in Fig. 3A. (right) equations for key parameter ratios.
B. From [70] Supplementary Table 1, for mouse fibroblasts transcriptome-wide, histograms of (left), burst on-rates,
C. ratios, per allele, of burst on to burst off rates, and
D. mRNA production rates
E. Prevalence of distribution classes for gene product B across the sampled parameter space for 10000 sampled parameter sets. Sankey plot demonstrates classes of gene B in the wildtype A genotype and, after mutation, in the heterozygous A genotype. Height of each section proportional to number of classified parameter sets.



Supplementary Figure 5

S5: Autocorrelation analysis of simulated gene expression traces

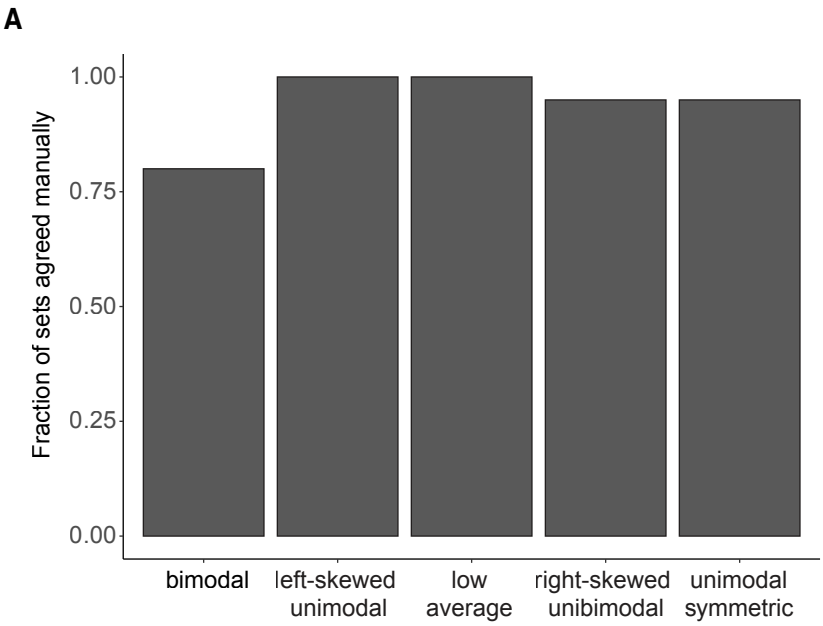
A. Three random, representative autocorrelation plots for per-100-timestep samples from gene expression traces in the wildtype genotype. Dashed red line represents the 95% confidence bound around zero autocorrelation.
B. Equivalent autocorrelation plots in the heterozygous mutant genotype.
C. Equivalent autocorrelation plots in the homozygous mutant genotype.

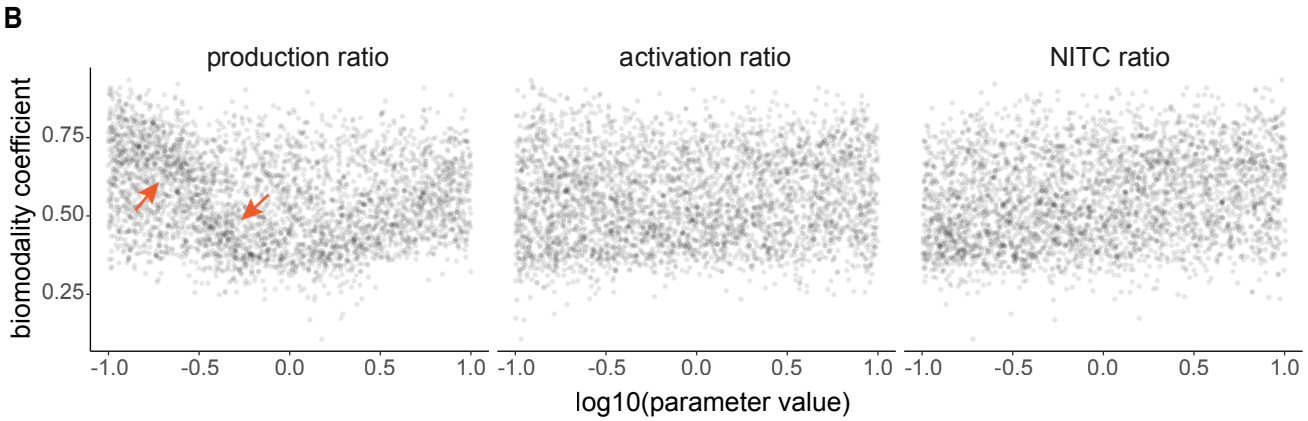


Supplementary Figure 6

S6: Distribution shapes and resampling to identify parameter subspaces enriched for trimodal distributions

A. Example simulation output and (skewed and low-average) from pseudo-single-cells taken every 300 time-steps.
B. Example inference of single-cell expression distributions (skewed and low-average). See Methods. Note that left-skewed and right-skewed distributions are classified separately (see Supplementary Note).
C. Example simulation output for a manually-identified trimodal distribution.
D. Example expression distribution of a trimodal distribution. See Methods.
E. Analysis schematic: How parameter subspace is constructed from decision trees.
F. Parameter subspaces around the initially identified trimodal distribution's parameter set, projected onto only the dimensions of NITC ratio, activation ratio, and production ratio.
G. Enrichment of trimodal gene expression distribution shapes after sampling parameter sets from the subspaces vs. from the full parameter space.



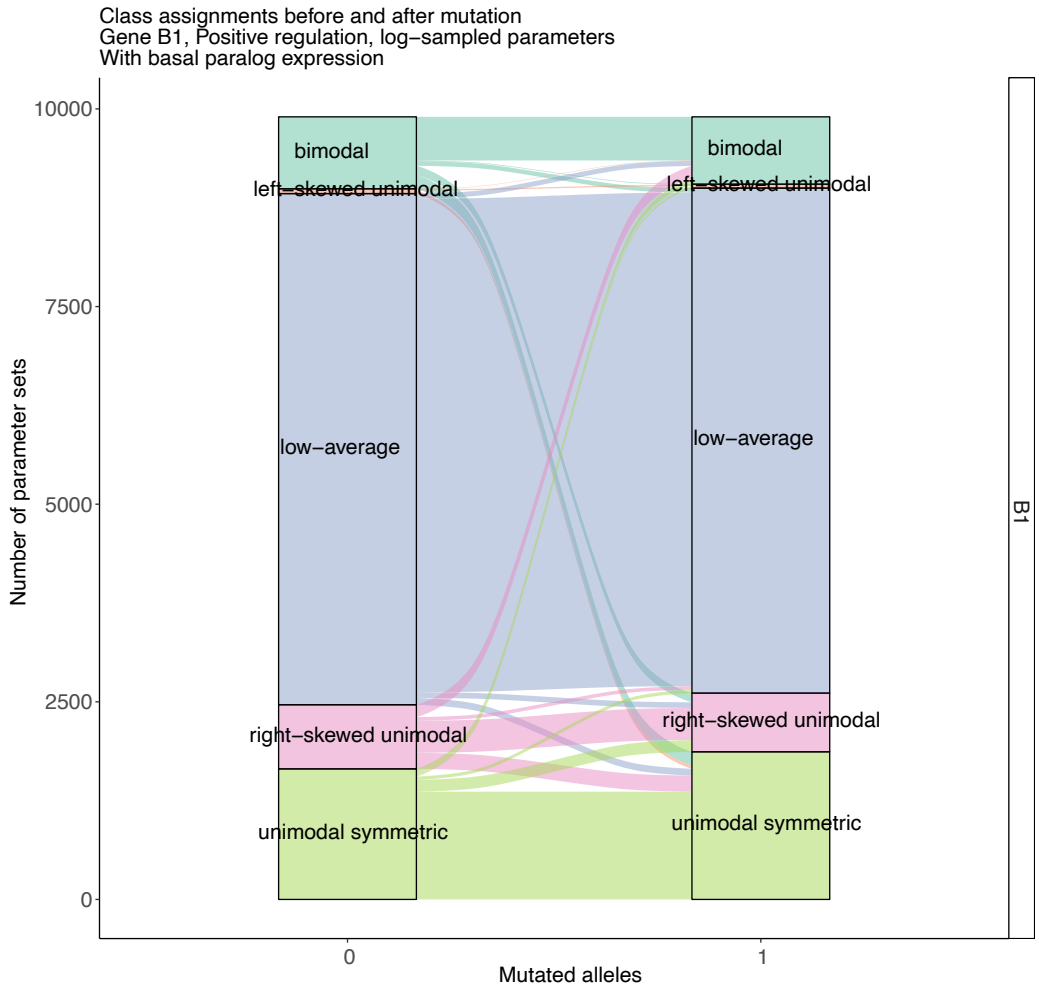


Supplementary Figure 7

S7: Distribution shape classifier and model parameter associations with classifier component statistics

A. Random samples of distributions assigned to each class were sampled (20-40 per class) and manually assessed for whether they matched class assignment. Fraction matching each class assignment on y-axis.
B. Scatter plots and correlation between selected summary statistics of B distributions in the heterozygous genotype and model parameters. Red arrows highlighting density of observations, see correlation reported in Results.

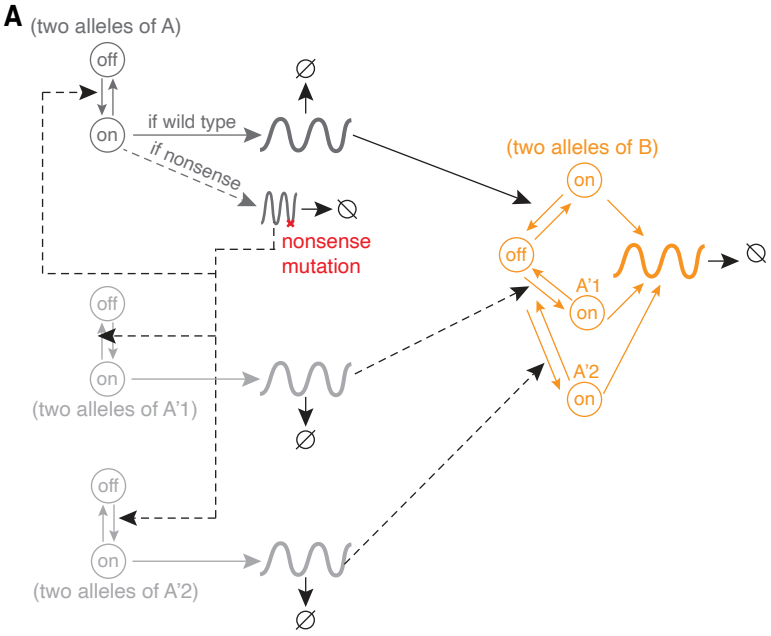
A

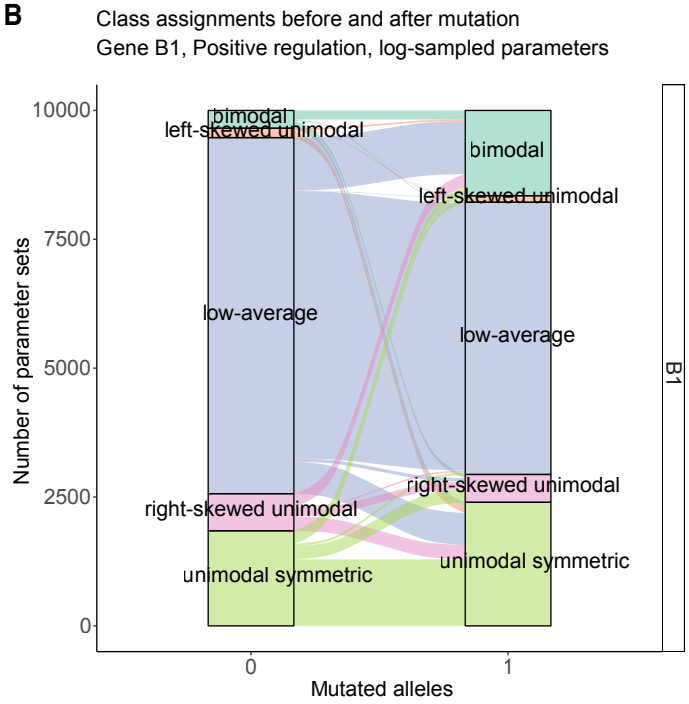


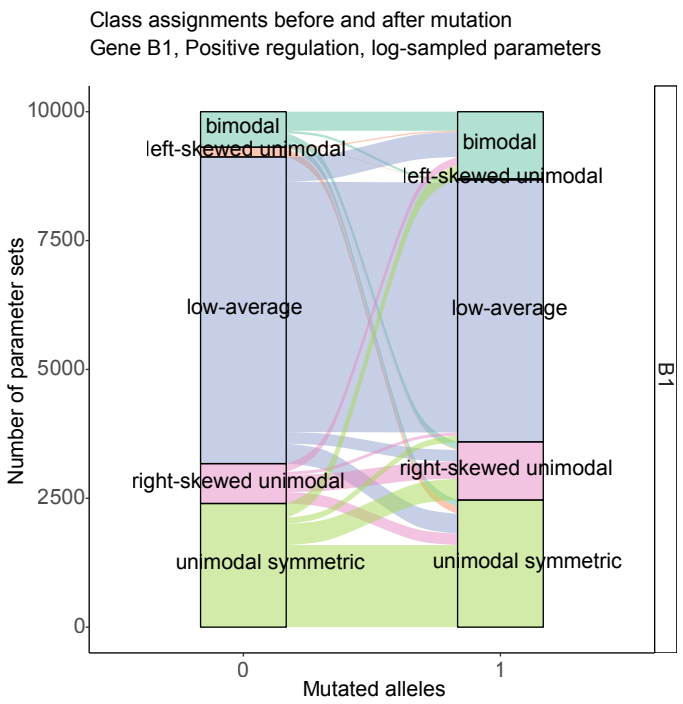
Supplementary Figure 8

S8: Distribution class results with basal paralog expression

A. Prevalence of distribution classes for gene product B across the sampled parameter space for 10000 sampled parameter sets, including some (variable) basal paralog expression. Sankey plot demonstrates classes of gene B in the wildtype A genotype and, after mutation, in the heterozygous A genotype.



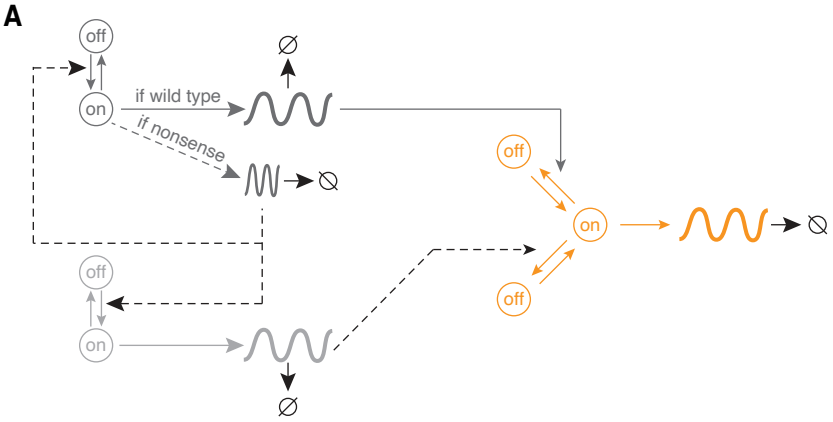


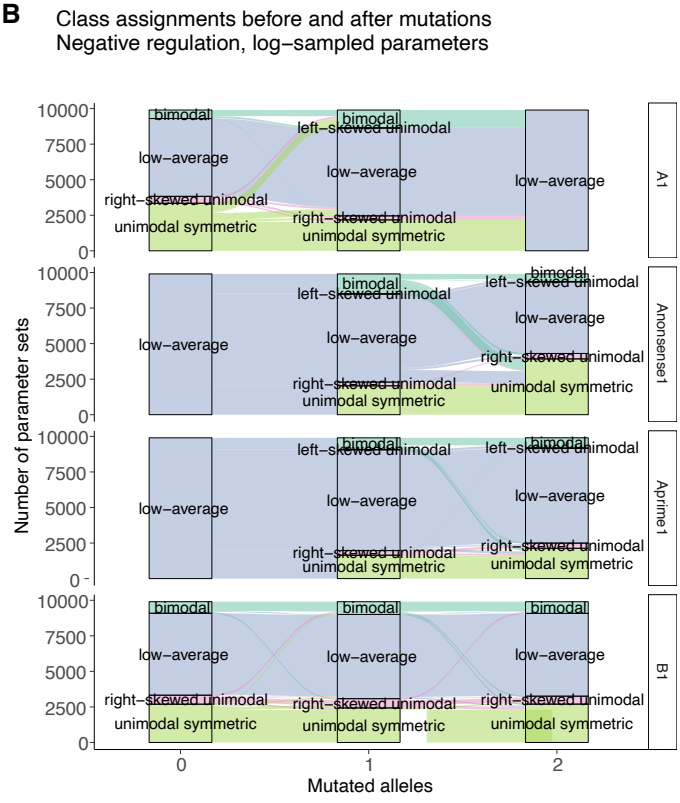


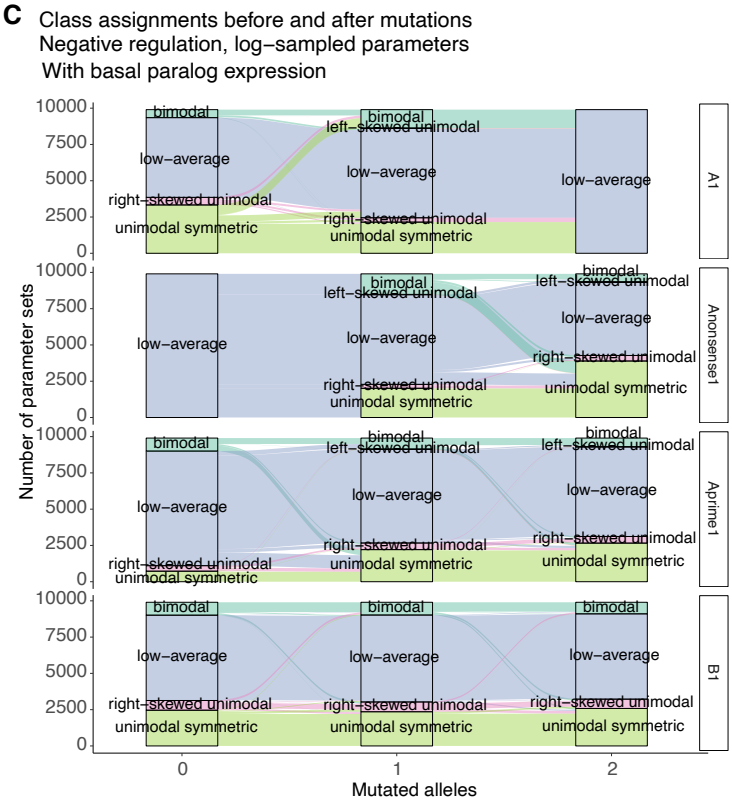
Supplementary Figure 9

S9: A gene regulatory network with transcriptional adaptation by multiple paralogs

A. Schematic of a gene regulatory network with transcriptional adaptation to mutation, with two paralogs. Two alleles of each gene, with bursty transcription of gene products at each allele. A mutated reference gene, A, regulates downstream effector gene B. When mutated, nonsense copies of A product stochastically upregulate paralogs of A, called A'1 and A'2. A'1 and A'2 can also regulate B, albeit with different strengths than A. In one set of simulations, A'1 and A'2 have the same parameter values. In a second set of simulations, they have different effects on the production rate of B. Hill functions are used in propensities for regulatory relationships between gene products and target alleles. See Methods for full model specification.
B. Prevalence of distribution classes for gene product B across the sampled parameter space for 10000 sampled parameter sets. Both paralogs have the same parameter values. Sankey plot demonstrates classes of gene B in the wildtype A genotype and, after mutation, in the heterozygous A genotype.
C. Prevalence of distribution classes for gene product B across the sampled parameter space for 10000 sampled parameter sets. A'1 and A'2 have different effects on the production rate of B. Sankey plot demonstrates classes of gene B in the wildtype A genotype and, after mutation, in the heterozygous A genotype.



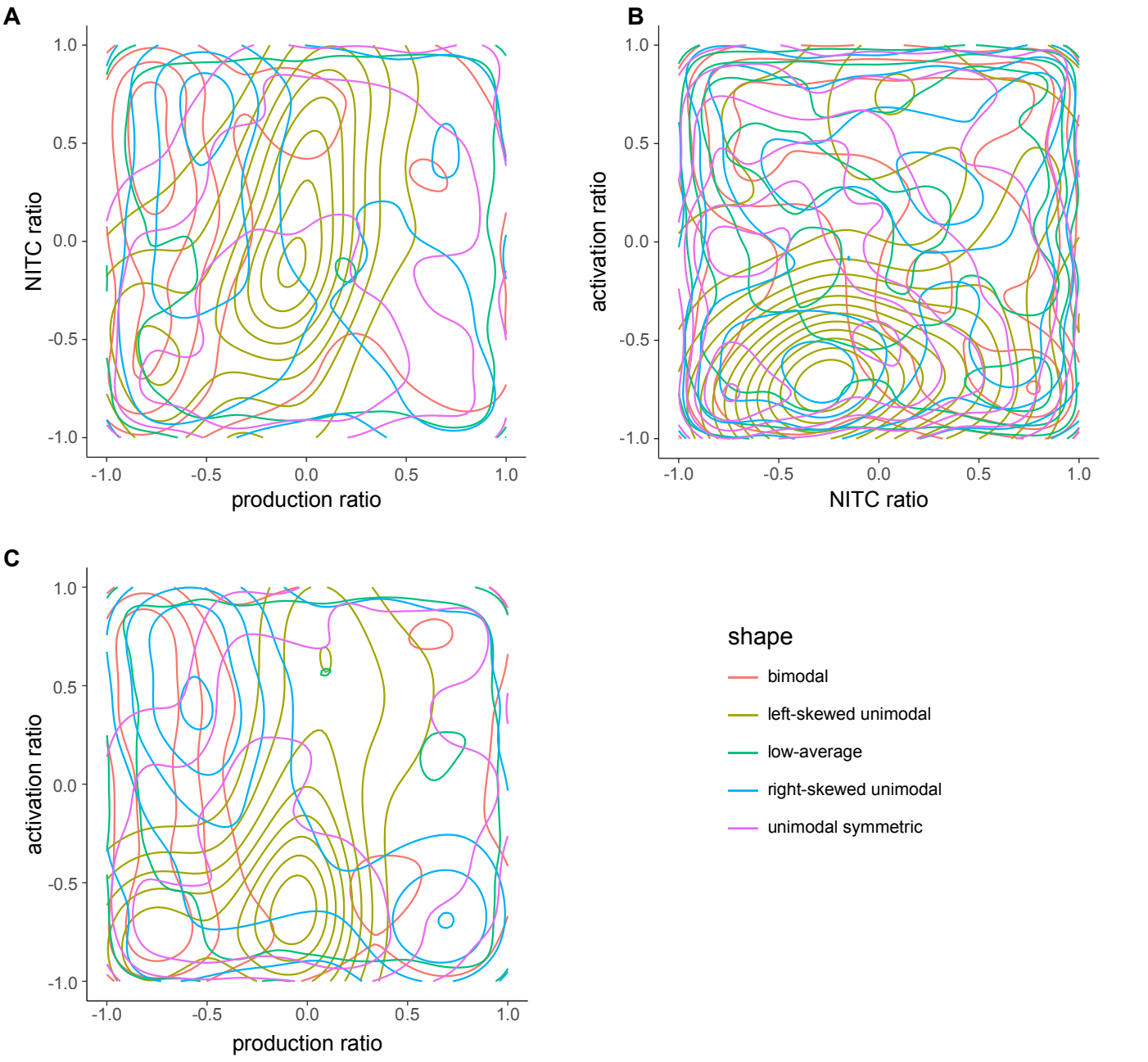




Supplementary Figure 10

S10: A gene regulatory network with transcriptional adaptation for a repressor

A. Schematic of a gene regulatory network with transcriptional adaptation to mutation, in which the regulatory interactions between A and B are repressive.Two alleles of each gene, with bursty transcription of gene products at each allele. A mutated reference gene, A, regulates downstream effector gene B. When mutated, nonsense copies of A product stochastically upregulate a paralog of A, called A'. A' can also regulate B, with different parameters than A. In one set of simulations, A' does not have any basal expression (without the effects of NITC); in another set of simulations, A' has variable amounts of basal expression. Hill functions are used in propensities for regulatory relationships between gene products and target alleles. See Methods for full model specification.
B. Prevalence of distribution classes for all gene products across the sampled parameter space for 10000 sampled parameter sets. No basal expression of A'. Sankey plot demonstrates classes of each gene in the wildtype A genotype and, after mutation, in the heterozygous A genotype, and after a second mutation, in the homozygous mutant A genotype.
C. Prevalence of distribution classes for each gene product across the sampled parameter space for 10000 sampled parameter sets. Variable amounts of basal expression of A'. Sankey plot demonstrates classes of gene B in the wildtype A genotype and, after mutation, in the heterozygous A genotype.

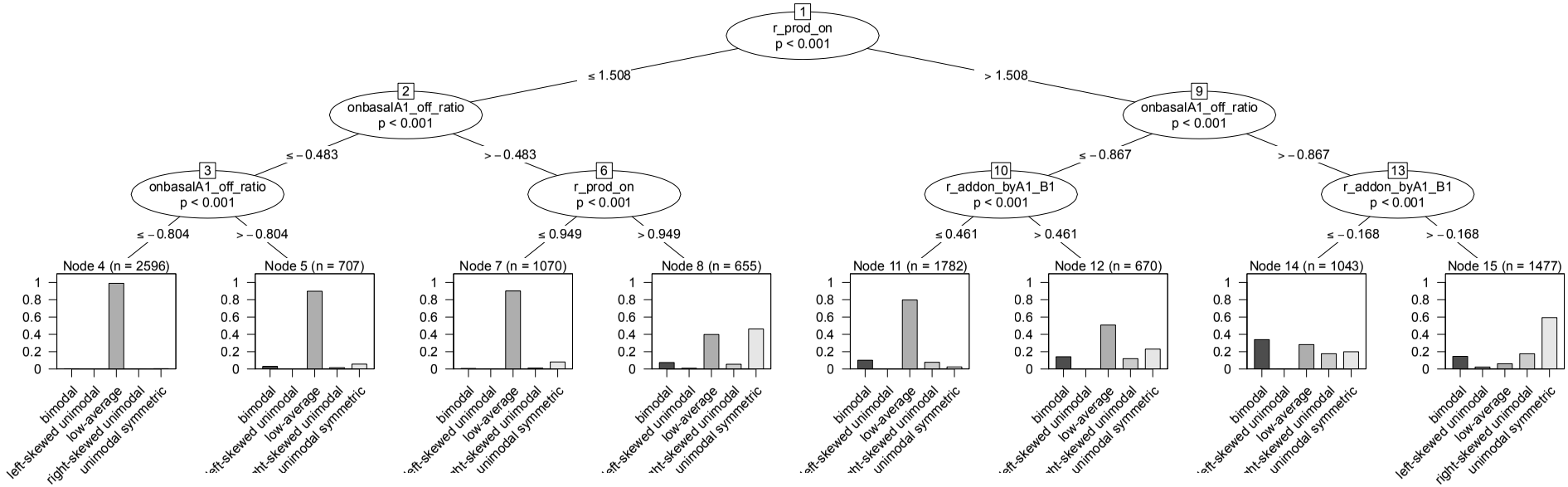


Supplementary Figure 11

S11: 2-dimensional projections of distribution shapes onto parameter ratio coordinates

A. Contour plot colored by distribution shape on NITC ratio vs. production ratio. Kernel density estimation using geom_density_2d() in R.
B. Contour plot colored by distribution shape on activation ratio vs. NITC ratio. Kernel density estimation using geom_density_2d() in R.
C. Contour plot colored by distribution shape on activation ratio vs. production ratio. Kernel density estimation using geom_density_2d() in R.

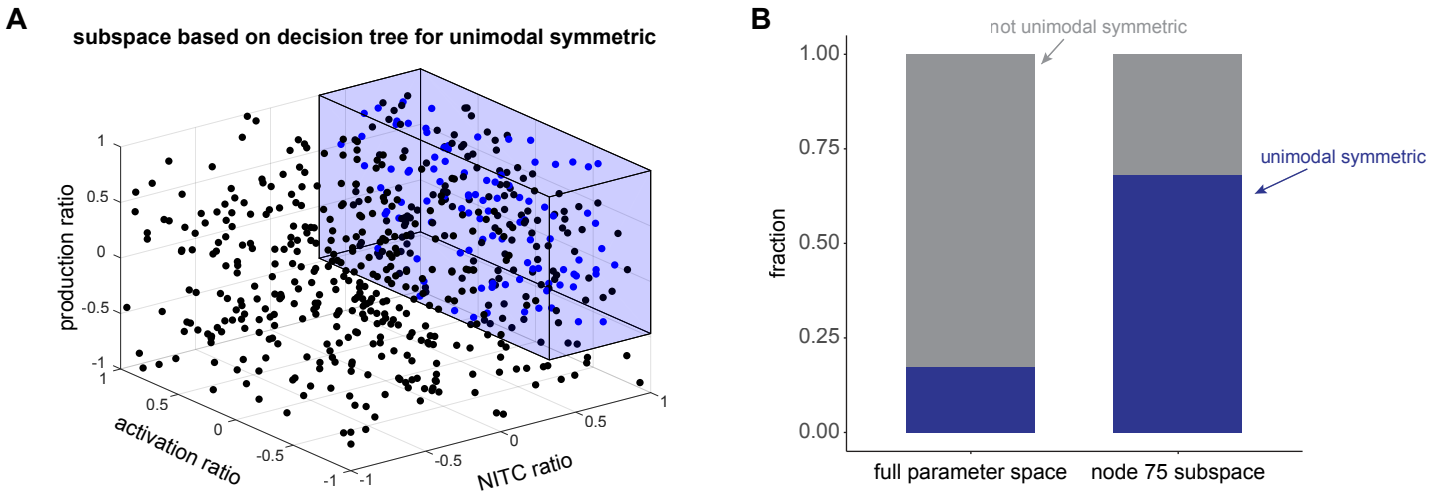
A



Supplementary Figure 12

S12: Multi-class decision tree for five distribution shapes of gene B in the heterozygous genotype

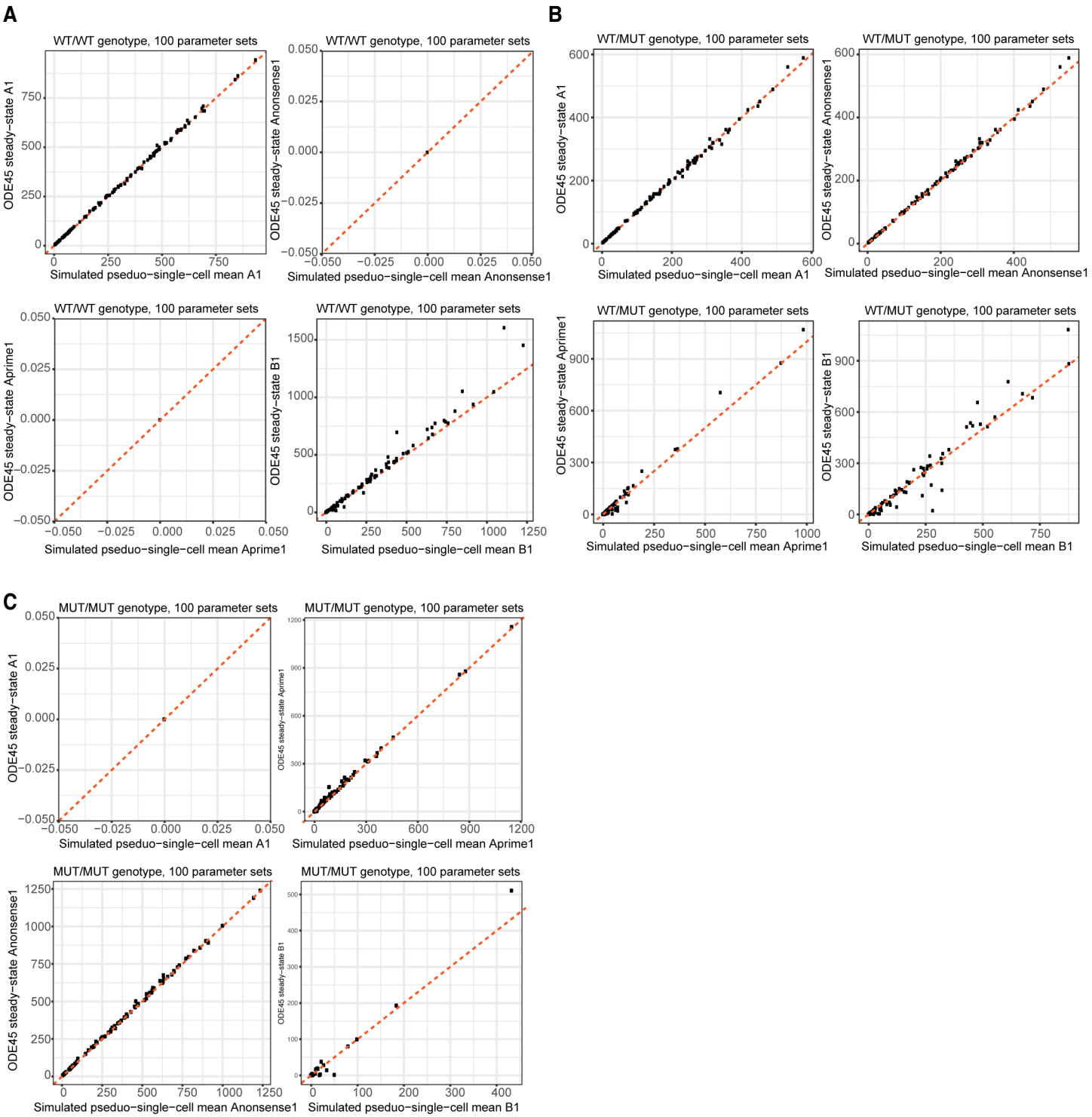
A. Decision tree trained on all parameter sets, on whether gene B is classified as one of {low-average, left-skewed, right-skewed, unimodal symmetric, bimodal} in the heterozygous genotype.



Supplementary Figure 13

S13: Resampling to identify parameter subspaces enriched for unimodal symmetric distributions

A. Parameter subspace corresponding to node 75 in the unimodal symmetric shape classifier tree in Supplementary Note, projected onto only the dimensions of NITC ratio, activation ratio, and production ratio..
B. Enrichment of unimodal symmetric gene expression distribution shapes after sampling parameter sets from the subspace in (A) vs. from the full parameter space.



Supplementary Figure 14

S14: Steady-state solutions of gene regulatory network systems of differential equations

A. For each gene product, steady-state solutions of the full system of ordinary differential equations representing the network in Fig. 2A, see Methods, in the wildtype genotype, compared against simulation average expression levels.
B. Equivalent plots in the heterozygous genotype.
C. Equivalent plots in the homozygous mutant genotype.

Supplementary Note

Prevalence of and gene regulatory constraints on transcriptional adaptation in single cells

Ian A. Mellis[1,2,*], Madeline E. Melzer[3,4,5], Nicholas Bodkin[3,4,5], and Yogesh Goyal[3,4,5,*,†]

[1]Department of Pathology and Cell Biology, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA
[2]Aaron Diamond AIDS Research Center, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA
[3]Department of Cell and Developmental Biology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[4]Center for Synthetic Biology, Northwestern University, Chicago, IL, USA
[5]Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

*Authors for correspondence: yogesh.goyal@northwestern.edu and im2613@cumc.columbia.edu
†Lead contact

Stochastic Model of compensatory gene regulatory networks

Parameter scanning: We scanned the following independent variables: the basal on-rate of A, the production rate when active (applied to all genes), the Hill coefficient n (applied to all genes), the r_{add} of A on transcription of B, and four parameter ratios of interest (Figure S4A). Four independent parameter ratios were: the ratio of $r_{add,NITC}$ to the on-rate of A (how strong is NITC relative to basal regulation of A), the ratio of the on-rate of A to the off-rate (how rapidly does A's expression burst cycle), the ratio of r_{add} of A on B to r_{add} of A' on B (what are the relative strengths of A and A' causing B burst activation), and the ratio of r_{prod} of B in the A-directed on state to the r_{prod} of B in the A'-directed on state (the relative activities of B on states) (Figure S4A). We used the Gillespie algorithm to numerically simulate network output under different parameter values.

Parameter ranges: Parameter search ranges were based on previous studies on transcriptional burst kinetics [56,60–62]. A recent study of transcriptome-wide allele-specific expression in single cells inferred telegraph model parameters relative to transcript degradation rates [60]. We therefore fixed the degradation rate at 1, concordant with Larsson and colleagues' analysis, and picked ranges of the burst-on rate of A, the on-off rate ratio, and production rate based on the range of parameters inferred by Larsson and colleagues [60] (Figure S4B-D). Other parameter ratios were similarly picked to span two orders of magnitude while also ensuring at least partial concordance with the parameter ranges observed for simple telegraph models [56,60–62].

Simulation quality control and sanity checks: We simulated our stochastic transcriptional model to capture and quantify the emergent single-cell variability within the minimal gene regulatory networks

(Figure 3B). Since the simulations are expected to be ergodic, we reasoned that we could condense long-timescale traces into “single-cell-like sub-simulations” that are independent of each other (Figure S5A-C, Methods) [56]. Indeed, we found no significant autocorrelation in the single-cell-like sub-simulations, which, in turn, enabled us to get population-level distributions of counts of each gene per parameter set. Briefly, with static snapshots taken every 300 timesteps, we simulated hundreds of single-cell-like gene expression profiles in wildtype, heterozygous mutant, and homozygous mutant genotypes. We summarized the expression levels per gene per genotype in each simulation with manually annotated empirical distribution classes (Figure 3C). We confirmed that average gene expression levels matched steady state approximations for the corresponding system of differential equations using the ode45 solver in MATLAB (Figure S6A-C).

Summary Statistics: We used summary statistics to describe the moments and other features of expression distributions including mean, coefficient of variation, skewness, bimodality coefficient (a statistic correlated with multimodality), gini coefficient, and entropy. For expressed genes, the values of statistics varied widely across parameter sets (Figure S7). We explored distribution shapes and their respective summary statistics to arrive at classification rules predicting distribution shape. The rules included mean, skewness, bimodality coefficient, and bimodality coefficient normalized to mean expression level using LOESS (since there was systematic variation in bimodality coefficient with the mean) (see Methods).

Analysis on model with basal paralog expression: We wanted to check if paralogs with some basal expression, like many observed paralogs in real world datasets, display similar diversity in distribution shapes and robustness of distribution shape to mutation [63]. We ran 10,000 more simulations, keeping the original 8 independently varying parameters and ratios in their original ranges. We added a 9th independent variable: the ratio of the basal burst-on rate of A’ relative to that of A. The second set of simulations produced all 5 distribution classes observed in the first set. Notably, more parameter sets preserve the distribution shape of the downstream target, B, after mutation, than in the first set of simulations (Figure S9). A larger fraction of robustness-to-mutation parameter sets suggests that having non-zero basal paralog expression, representative of gene A-A’ redundancy, may help ensure robustness to mutation.

Multiple paralogs: Many human and mouse genes have multiple annotated paralogs that could in principle contribute to transcriptional adaptation upon mutation [41]. We therefore wondered whether an expanded gene regulatory network model, with multiple paralogs of the ancestral regulator A, would demonstrate similar diversity of resulting expression distribution shapes and robustness to mutation of A (see Methods, Figure S11A). Indeed, we observed a diversity of distribution shapes across the full parameter space, with all 5 previously observed classes represented (Figure S11B,C), and robustness results were qualitatively concordant with single paralog network simulations (Figure S12, S13).

Repressive networks: Many gene regulatory networks also include transcription factors with repressive effects on downstream targets [64]. We wondered whether transcriptional adaptation could be important for preserving network output after mutation of a repressor (see Methods; Figure S14A). Here again, we observed similarly diverse distribution shapes across the full parameter space (except left-skewed distributions) (Figure S14B,C), and a subset of parameters (albeit a narrower set than activator network) enabled robustness (Figure S15,16). Further work may elucidate regulatory constraint differences between activator- and repressor-based networks with compensation.

Unimodal Symmetric robustness: We chose to focus on unimodal symmetric, as this distribution shape is most reflective of a homogeneous, unskewed population, with non-zero expression. Therefore, we trained a decision tree classifier on the sampled parameter sets (see Methods) to identify the model parameter and ratio combinations most predictive of whether gene B in the heterozygous genotype would be unimodal symmetric class (see Methods). We found a total of 31 significant decision rules up to 6 layers deep per combination, resulting in 33 groupings of parameter sets (Figure S10). All included independent parameters and ratios that were present in at least one decision rule. The combination with the highest fraction of unimodal symmetric gene B distributions in the heterozygous state had 6 decision rules, including requiring sufficiently strong $r_{add,NITC}$ (i.e., a low ratio of $r_{on,basal}$ of A to $r_{add,NITC}$ acting on A'). Consistent with bimodality coefficient correlation results, other subspaces enriched for unimodal symmetric shapes included rules restricting the ratio of A- and A'-directed B production rates to be closer to 1.

Trimodality additional discussion: Trimodality was unexpected because *a priori* the expected possible regulation regimes for gene B are A-driven, A'-driven, and neither-driven, with the last having no basal expression. Examination of the gene expression traces suggested that the rate of B mRNA degradation was too slow to reach zero before an allele was activated by A or A' (Figure 3E). This result raises the plausibility of the existence of a triad of phenotypic outcomes, and underscores the emergent complexity inherent to even relatively simple gene regulatory networks. To verify whether this distribution was associated with the specific parameter combination, i.e., subspace within parameter space for this parameter set, we sampled directly around the parameter set with defined ranges smaller than the original search space ranges (Figure 3F, see Methods). We chose subspace ranges significantly smaller in each dimension than the full parameter space, and consistent with the size of some parameter ranges bounding subspaces enriched for other distribution shapes (see additional decision tree point above). Indeed, we found that essentially all (98%) parameter sets sampled within the narrow subspace ranges displayed similar trimodality, while a new random sample of the full original parameter space resulted in only 1% of sampled simulations displaying trimodality of B (Figure 3G). We supplemented our analysis with a more systematic approach as discussed in the main text (Fig. S6). Taken together, this analysis further established that we can use parameter set combinations to predict distribution shapes.

Supplementary Note Figures

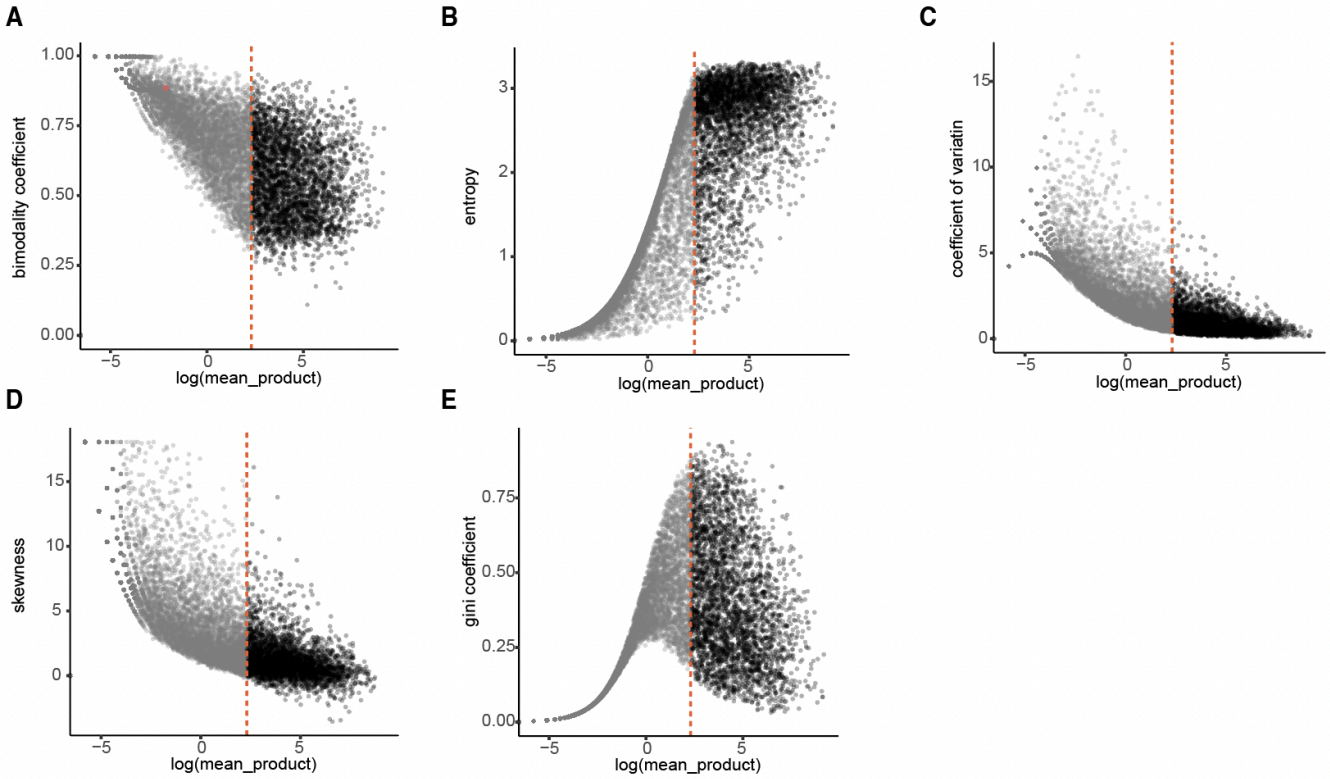


Fig. SN1: Summary statistic covariation with mean

A. Bimodality coefficient against log(mean) for gene product B in the heterozygous genotype, Red dashed line marks mean = 10.
B. Entropy against log(mean) for gene product B in the heterozygous genotype, Red dashed line marks mean = 10.
C. Coefficient of variation against log(mean) for gene product B in the heterozygous genotype, Red dashed line marks mean = 10.
D. Skewness against log(mean) for gene product B in the heterozygous genotype, Red dashed line marks mean = 10.
E. Gini coefficient against log(mean) for gene product B in the heterozygous genotype, Red dashed line marks mean = 10.

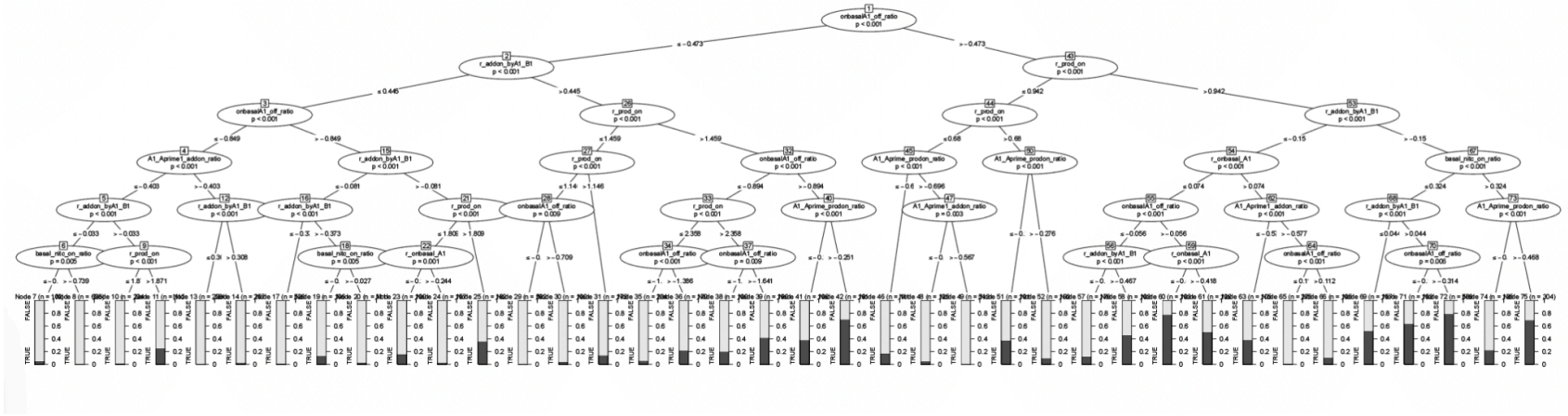

1
onbasalA1_off_ratio
p < 0.001
≤ −0.473
> −0.473
2
r_addon_byA1_B1
p < 0.001
43
r_prod_on
p < 0.001
≤ 0.445
> 0.445
≤ 0.942
> 0.942
3
onbasalA1_off_ratio
p < 0.001
26
r_prod_on
p < 0.001
44
r_prod_on
p < 0.001
53
r_addon_byA1_B1
p < 0.001
≤ −0.849
> −0.849
≤1.459
>1.459
≤0.68
>0.68
≤ −0.15
> −0.15
4
A1_Aprime1_addon_ratio
p < 0.001
15
r_addon_byA1_B1
p < 0.001
27
r_prod_on
p < 0.001
32
onbasalA1_off_ratio
p < 0.001
45
A1_Aprime_prodon_ratio
p < 0.001
50
A1_Aprime_prodon_ratio
p < 0.001
54
r_onbasal_A1
p < 0.001
67
basal_nlc_on_ratio
p < 0.001
≤ −0.403
> −0.403
≤ −0.081
> −0.081
≤1.14
>1.146
≤ −0.894
> −0.894
≤ −0.6
> −0.696
≤ 0.074
> 0.074
≤0.324
>0.324
5
r_addon_byA1_B1
p < 0.001
12
r_addon_byA1_B1
p < 0.001
16
r_addon_byA1_B1
p < 0.001
21
r_prod_on
p < 0.001
28
onbasalA1_off_ratio
p = 0.009
33
r_prod_on
p < 0.001
40
A1_Aprime_prodon_ratio
p < 0.001
47
A1_Aprime1_addon_ratio
p = 0.003
60
onbasalA1_off_ratio
p < 0.001
62
A1_Aprime1_addon_ratio
p < 0.001
68
r_addon_byA1_B1
p < 0.001
73
A1_Aprime_prodon_ratio
p < 0.001
≤ −0.033
> −0.033
≤ −0.3
> −0.373
≤ −0.894
> −0.894
≤ −0.696
> −0.896
≤ −0.056
> −0.056
≤ −0.577
> −0.577
≤0.044
>0.044
≤ −0.709
> −0.709
≤ −0.251
> −0.251
≤ −0.276
> −0.276
≤ −0.567
> −0.567
≤ −0.467
> −0.467
≤ −0.418
> −0.418
≤ −0.468
> −0.468
6
basal_nlc_on_ratio
p = 0.005
9
r_prod_on
p < 0.001
18
basal_nlc_on_ratio
p = 0.005
22
r_onbasal_A1
p = 0.001
34
onbasalA1_off_ratio
p < 0.001
37
onbasalA1_off_ratio
p = 0.009
41
A1_Aprime_prodon_ratio
p < 0.001
48
A1_Aprime1_addon_ratio
p < 0.001
51
A1_Aprime_prodon_ratio
p < 0.001
56
r_addon_byA1_B1
p < 0.001
59
r_onbasal_A1
p < 0.001
64
onbasalA1_off_ratio
p < 0.001
70
onbasalA1_off_ratio
p = 0.006
75
A1_Aprime_prodon_ratio
p < 0.001
≤ −0.739
> −0.739
≤1.8
>1.871
≤ −0.027
> −0.027
≤ −0.244
> −0.244
≤ −1.386
> −1.386
≤ −1.641
> −1.641
≤ 0.112
> 0.112
≤ −0.314
> −0.314
Node 7 (n = 100)
Node 8 (n = 286)
Node 10 (n = 204)
Node 11 (n = 100)
Node 13 (n = 286)
Node 14 (n = 286)
Node 17 (n = 286)
Node 19 (n = 286)
Node 20 (n = 286)
Node 23 (n = 286)
Node 24 (n = 286)
Node 25 (n = 286)
Node 29 (n = 286)
Node 30 (n = 286)
Node 31 (n = 286)
Node 35 (n = 286)
Node 36 (n = 286)
Node 38 (n = 286)
Node 39 (n = 286)
Node 41 (n = 286)
Node 42 (n = 286)
Node 46 (n = 286)
Node 49 (n = 286)
Node 52 (n = 286)
Node 57 (n = 286)
Node 58 (n = 286)
Node 60 (n = 286)
Node 61 (n = 286)
Node 63 (n = 286)
Node 65 (n = 286)
Node 66 (n = 286)
Node 69 (n = 286)
Node 71 (n = 286)
Node 72 (n = 286)
Node 74 (n = 286)
Node 75 (n = 204)
FALSE
TRUE
0
0.2
0.4
0.6
0.8
1

Fig. SN2: Decision tree analysis of unimodal symmetric distribution shape in the heterozygous genotype

A. Decision tree trained on all parameter sets, on whether gene B is classified as unimodal symmetric shape in the heterozygous genotype.

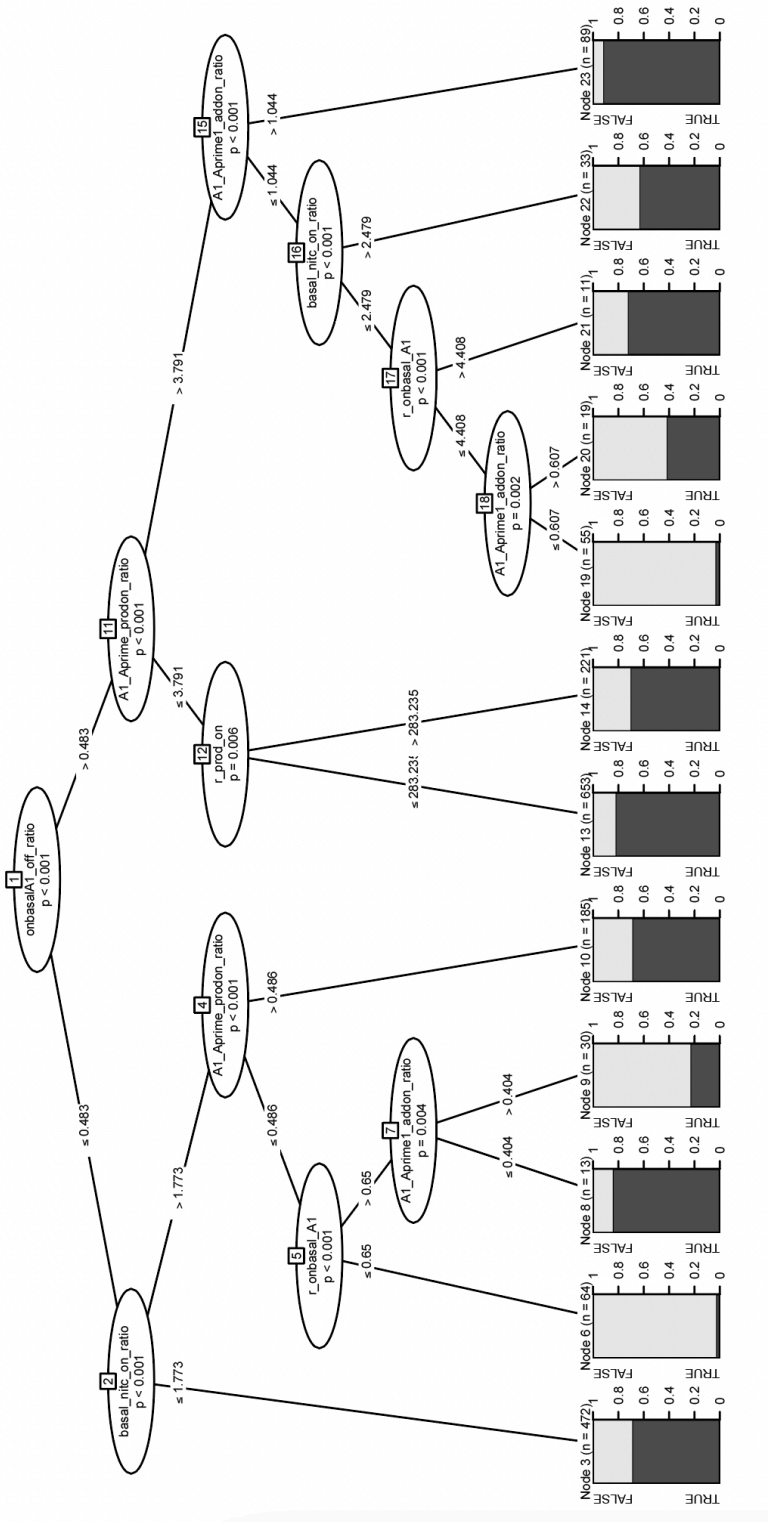


Fig. SN3: Decision tree analysis for robustness of unimodal symmetric B expression in a multi-paralog model with equal paralog parameters

A. Decision tree trained on parameter sets in which gene B is classified as unimodal symmetric in the wildtype phenotype, on whether gene B is classified as having unimodal symmetric shape in the heterozygous genotype.

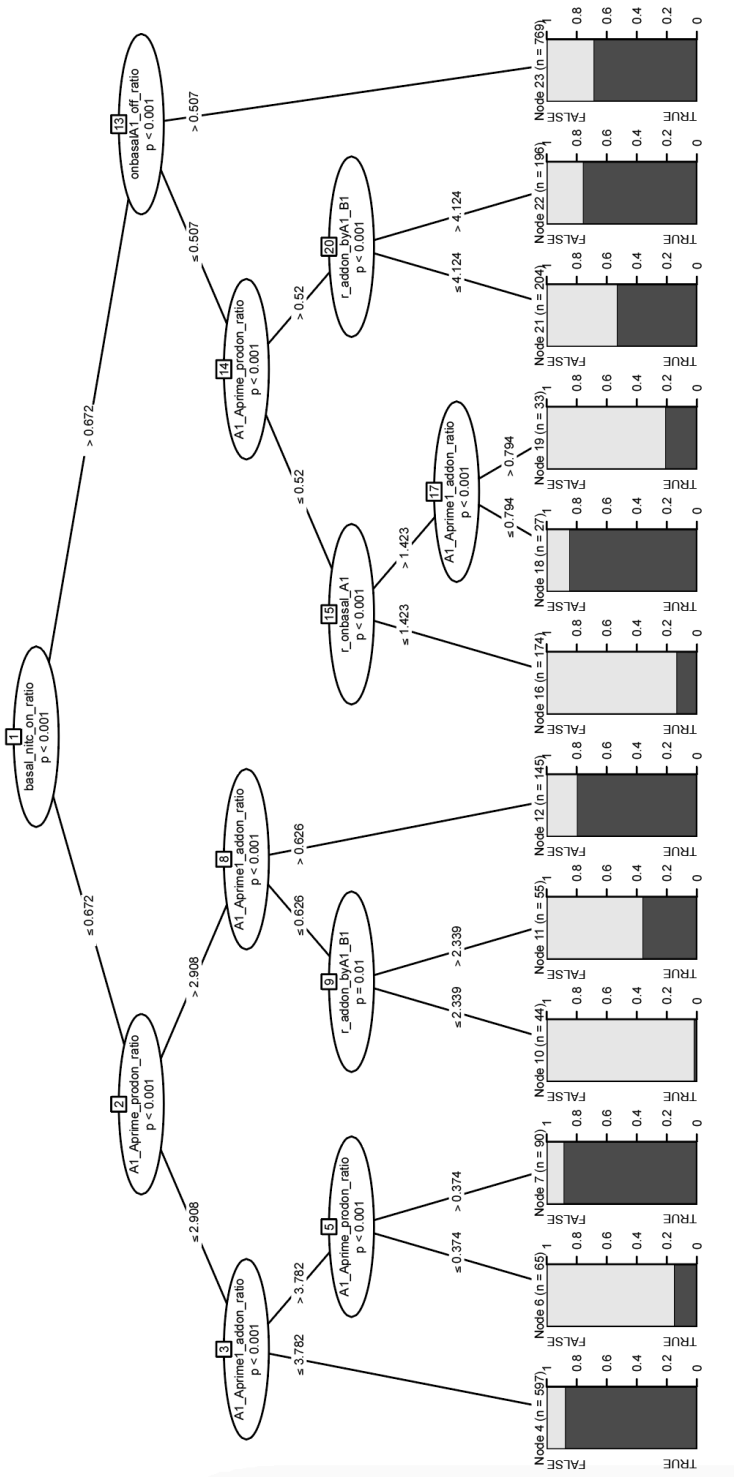


Fig. SN4: Decision tree analysis for robustness of unimodal symmetric B expression in a multi-paralog model with different paralog parameters

A. Decision tree trained on parameter sets in which gene B is classified as unimodal symmetric in the wildtype phenotype, on whether gene B is classified as having unimodal symmetric shape in the heterozygous genotype.

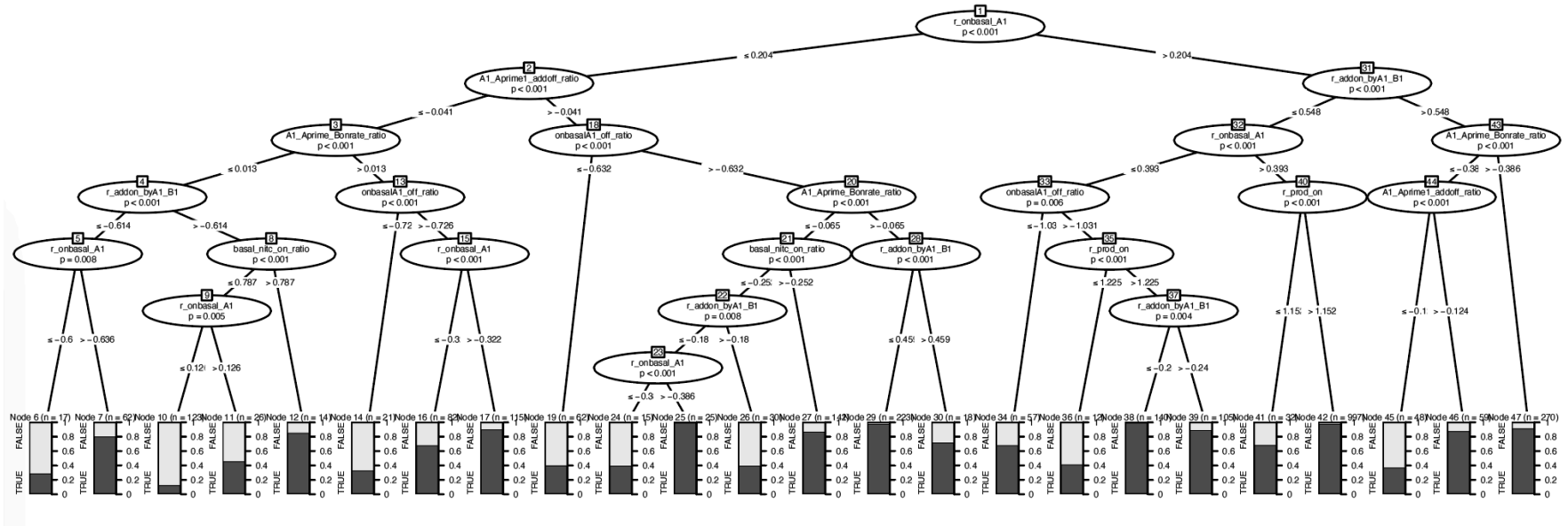

1
r_onbasal_A1
p < 0.001
≤ 0.204
> 0.204
2
A1_Aprime1_addoff_ratio
p < 0.001
31
r_addon_byA1_B1
p < 0.001
≤ −0.041
> −0.041
≤ 0.548
> 0.548
3
A1_Aprime_Bonrate_ratio
p < 0.001
16
onbasalA1_off_ratio
p < 0.001
32
r_onbasal_A1
p < 0.001
43
A1_Aprime_Bonrate_ratio
p < 0.001
≤ 0.013
> 0.013
≤ −0.632
> −0.632
≤ 0.393
> 0.393
≤ −0.38
> −0.386
4
r_addon_byA1_B1
p < 0.001
13
onbasalA1_off_ratio
p < 0.001
20
A1_Aprime_Bonrate_ratio
p < 0.001
33
onbasalA1_off_ratio
p = 0.006
40
r_prod_on
p < 0.001
44
A1_Aprime1_addoff_ratio
p < 0.001
≤ −0.614
> −0.614
≤ −0.72
> −0.726
≤ −0.065
> −0.065
≤ −1.03
> −1.031
≤ 1.15
> 1.152
≤ −0.1
> −0.124
5
r_onbasal_A1
p = 0.008
8
basal_nitc_on_ratio
p < 0.001
15
r_onbasal_A1
p < 0.001
21
basal_nitc_on_ratio
p < 0.001
28
r_addon_byA1_B1
p < 0.001
35
r_prod_on
p < 0.001
37
r_addon_byA1_B1
p = 0.004
≤ 0.787
> 0.787
≤ −0.252
≤ 1.225
> 1.225
≤ −0.2
> −0.24
9
r_onbasal_A1
p = 0.005
22
r_addon_byA1_B1
p = 0.008
≤ −0.18
> −0.18
23
r_onbasal_A1
p < 0.001
≤ −0.3
> −0.386
≤ 0.126
> 0.126
≤ −0.3
> −0.322
Node 6 (n = 17)
Node 7 (n = 62)
Node 10 (n = 123)
Node 11 (n = 26)
Node 12 (n = 14)
Node 14 (n = 21)
Node 16 (n = 82)
Node 17 (n = 115)
Node 19 (n = 62)
Node 24 (n = 15)
Node 25 (n = 25)
Node 26 (n = 30)
Node 27 (n = 142)
Node 29 (n = 223)
Node 30 (n = 18)
Node 34 (n = 57)
Node 36 (n = 12)
Node 38 (n = 14)
Node 39 (n = 105)
Node 41 (n = 32)
Node 42 (n = 997)
Node 45 (n = 48)
Node 46 (n = 59)
Node 47 (n = 270)
FALSE
TRUE
0
0.2
0.4
0.6
0.8

Fig. SN5: Decision tree analysis for robustness of unimodal symmetric B expression in a network model of a repressor, without basal paralog expression

A. Decision tree trained on parameter sets in which gene B is classified as unimodal symmetric in the wildtype phenotype, on whether gene B is classified as having unimodal symmetric shape in the heterozygous genotype.

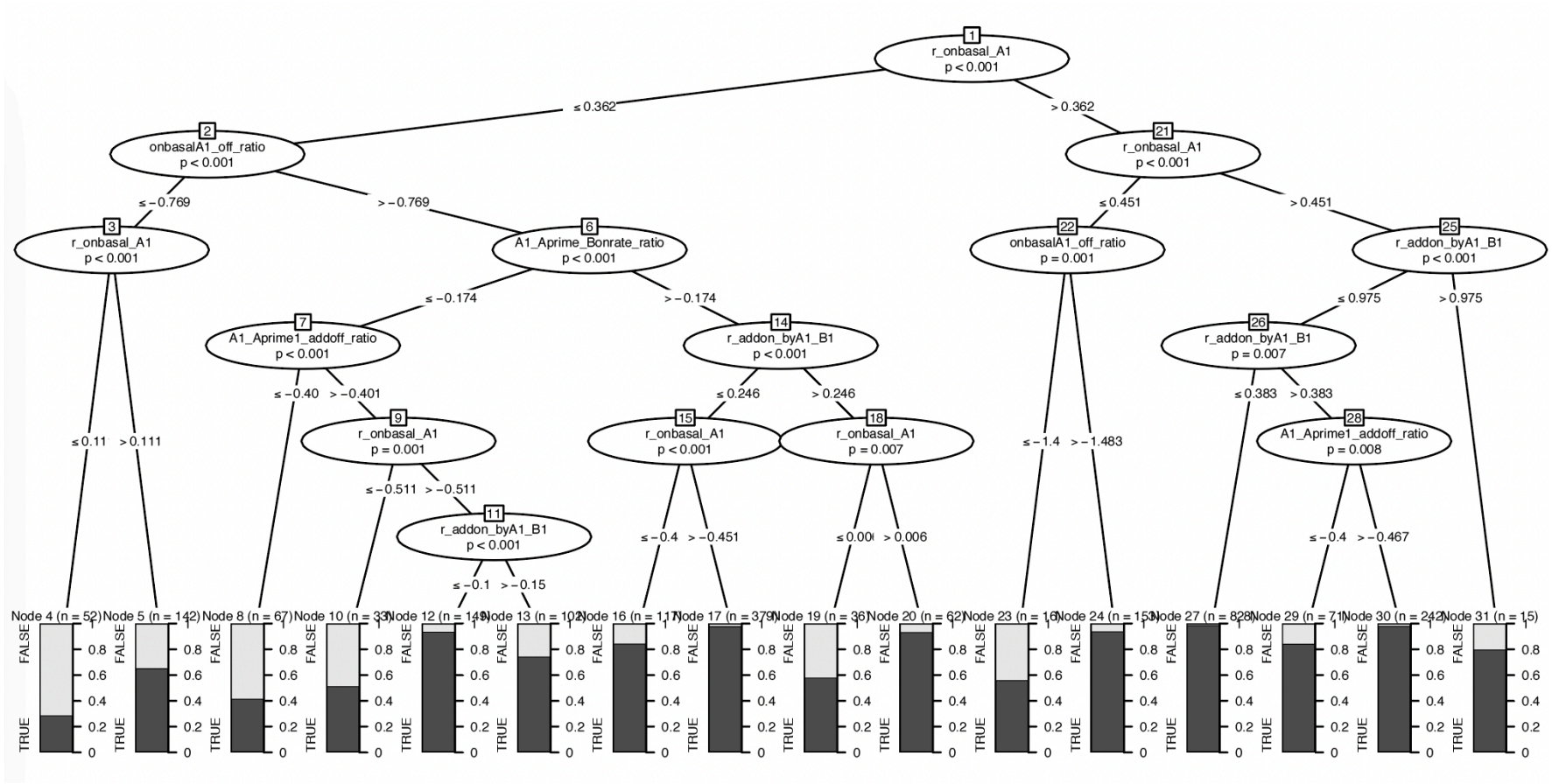

1
r_onbasal_A1
p < 0.001
≤ 0.362
> 0.362
2
onbasalA1_off_ratio
p < 0.001
21
r_onbasal_A1
p < 0.001
≤ −0.769
> −0.769
≤ 0.451
> 0.451
3
r_onbasal_A1
p < 0.001
6
A1_Aprime_Bonrate_ratio
p < 0.001
22
onbasalA1_off_ratio
p = 0.001
25
r_addon_byA1_B1
p < 0.001
≤ −0.174
> −0.174
≤ 0.975
> 0.975
7
A1_Aprime1_addoff_ratio
p < 0.001
14
r_addon_byA1_B1
p < 0.001
26
r_addon_byA1_B1
p = 0.007
≤ 0.11
> 0.111
≤ −0.40
> −0.401
≤ 0.246
> 0.246
≤ 0.383
> 0.383
9
r_onbasal_A1
p = 0.001
15
r_onbasal_A1
p < 0.001
18
r_onbasal_A1
p = 0.007
≤ −1.4
> −1.483
28
A1_Aprime1_addoff_ratio
p = 0.008
≤ −0.511
> −0.511
11
r_addon_byA1_B1
p < 0.001
≤ −0.4
> −0.451
≤ −0.4
> −0.467
≤ −0.1
> −0.15
Node 4 (n = 52)
Node 5 (n = 142)
Node 8 (n = 67)
Node 10 (n = 33)
Node 12 (n = 149)
Node 13 (n = 102)
Node 16 (n = 117)
Node 17 (n = 379)
Node 19 (n = 36)
Node 20 (n = 62)
Node 23 (n = 16)
Node 24 (n = 153)
Node 27 (n = 828)
Node 29 (n = 71)
Node 30 (n = 242)
Node 31 (n = 15)
FALSE
TRUE
0
0.2
0.4
0.6
0.8
1

Fig. SN6: Decision tree analysis for robustness of unimodal symmetric B expression in a network model of a repressor, with basal paralog expression

A. Decision tree trained on parameter sets in which gene B is classified as unimodal symmetric in the wildtype phenotype, on whether gene B is classified as having unimodal symmetric shape in the heterozygous genotype.

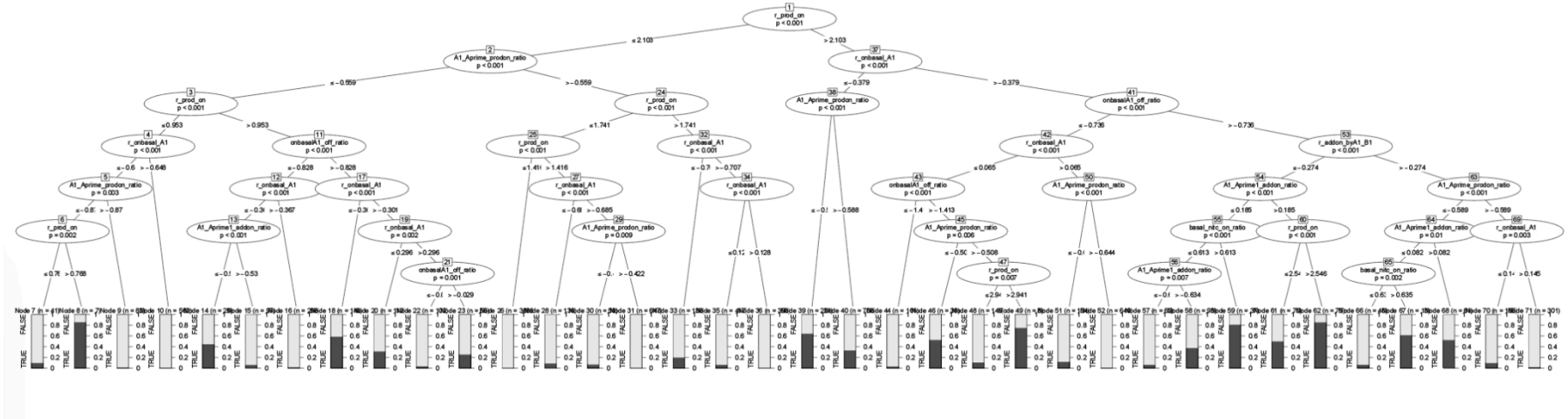

1
r_prod_on
p < 0.001
≤ 2.103
> 2.103
2
A1_Aprime_prodon_ratio
p < 0.001
39
r_onbasal_A1
p < 0.001
≤ −0.559
> −0.559
≤ −0.379
> −0.379
3
r_prod_on
p < 0.001
24
r_prod_on
p < 0.001
38
A1_Aprime_prodon_ratio
p < 0.001
41
onbasalA1_off_ratio
p < 0.001
≤ 0.953
> 0.953
≤ 1.741
> 1.741
≤ −0.736
> −0.736
4
r_onbasal_A1
p < 0.001
11
onbasalA1_off_ratio
p < 0.001
25
r_prod_on
p < 0.001
32
r_onbasal_A1
p < 0.001
42
r_onbasal_A1
p < 0.001
53
r_addon_bvA1_B1
p < 0.001
≤ −0.6
> −0.646
≤ −0.828
> −0.828
≤ 1.416
> 1.416
≤ −0.7
> −0.707
≤ 0.065
> 0.065
≤ −0.274
> −0.274
5
A1_Aprime_prodon_ratio
p = 0.003
12
r_onbasal_A1
p < 0.001
17
r_onbasal_A1
p < 0.001
27
r_onbasal_A1
p < 0.001
34
r_onbasal_A1
p < 0.001
43
onbasalA1_off_ratio
p < 0.001
50
A1_Aprime_prodon_ratio
p < 0.001
64
A1_Aprime1_addon_ratio
p < 0.001
63
A1_Aprime_prodon_ratio
p < 0.001
≤ −0.85
> −0.87
≤ −0.3
> −0.367
≤ −0.3
> −0.301
≤ −0.68
> −0.685
≤ −0.1
> −0.588
≤ −1.4
> −1.413
≤ 0.185
> 0.185
≤ −0.589
> −0.589
6
r_prod_on
p = 0.002
13
A1_Aprime1_addon_ratio
p < 0.001
19
r_onbasal_A1
p = 0.002
29
A1_Aprime_prodon_ratio
p = 0.009
45
A1_Aprime_prodon_ratio
p = 0.036
55
basal_nlc_on_ratio
p < 0.001
60
r_prod_on
p < 0.001
64
A1_Aprime1_addon_ratio
p = 0.01
69
r_onbasal_A1
p = 0.003
≤ 0.296
> 0.296
≤ −0.5
> −0.422
≤ 0.12
> 0.128
≤ −0.5
> −0.508
≤ −0.1
> −0.644
≤ 0.613
> 0.613
≤ 2.54
> 2.546
≤ 0.082
> 0.082
≤ 0.14
> 0.145
21
onbasalA1_off_ratio
p = 0.001
47
r_prod_on
p = 0.007
66
A1_Aprime1_addon_ratio
p = 0.007
65
basal_nlc_on_ratio
p = 0.002
≤ −0.1
> −0.029
≤ 2.94
> 2.941
≤ −0.1
> −0.634
≤ 0.6
> 0.635
Node 7 (n = 41)
Node 8 (n = 7
Node 9 (n = 5
Node 10 (n = 5
Node 14 (n = 2
Node 15 (n = 3
Node 16 (n = 2
Node 18 (n = 1
Node 20 (n = 1
Node 22 (n = 1
Node 23 (n = 7
Node 26 (n = 3
Node 28 (n = 1
Node 30 (n = 3
Node 31 (n = 9
Node 33 (n = 1
Node 35 (n = 4
Node 36 (n = 2
Node 39 (n = 2
Node 40 (n = 7
Node 44 (n = 1
Node 46 (n = 3
Node 48 (n = 1
Node 49 (n = 1
Node 51 (n = 1
Node 52 (n = 6
Node 57 (n = 2
Node 58 (n = 2
Node 59 (n = 2
Node 61 (n = 7
Node 62 (n = 7
Node 66 (n = 6
Node 67 (n = 1
Node 68 (n = 3
Node 70 (n = 1
Node 71 (n = 301)
FALSE
TRUE
0
0.2
0.4
0.6
0.8

Fig. SN7: Decision tree analysis of bimodal distribution shape in the heterozygous genotype

A. Decision tree trained on all parameter sets, on whether gene B is classified as bimodal shape in the heterozygous genotype.

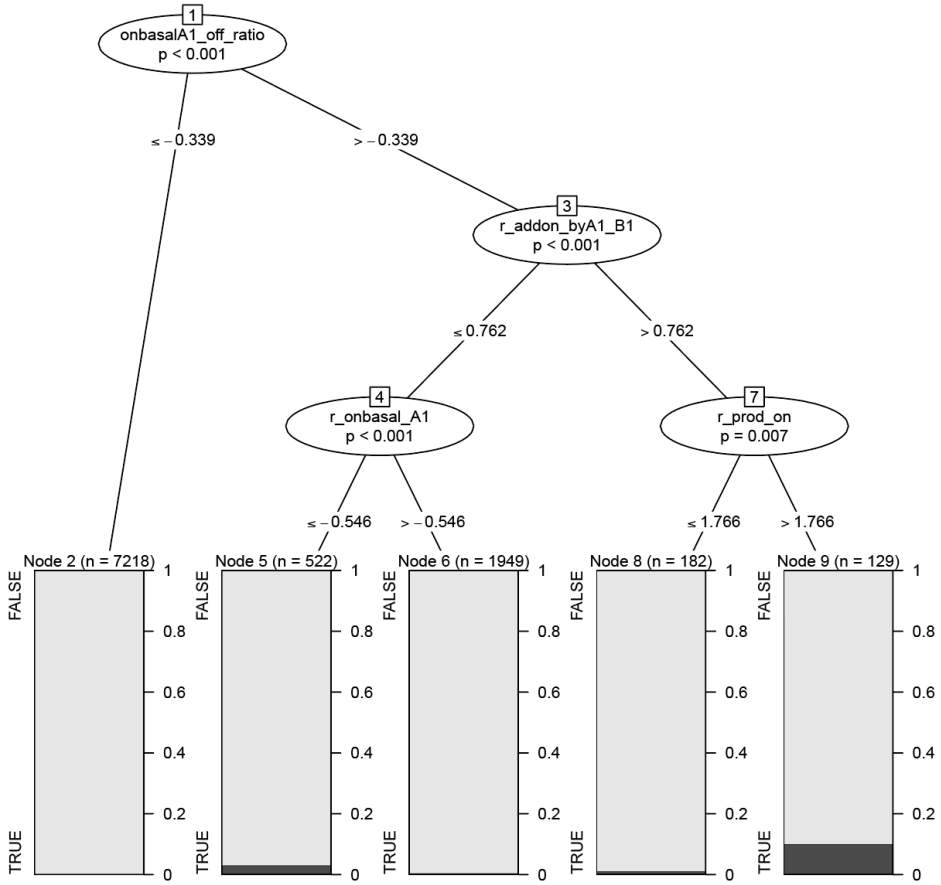


Fig. SN8: Decision tree analysis of left-skewed distribution shape in the heterozygous genotype

A. Decision tree trained on all parameter sets, on whether gene B is classified as left-skewed shape in the heterozygous genotype.

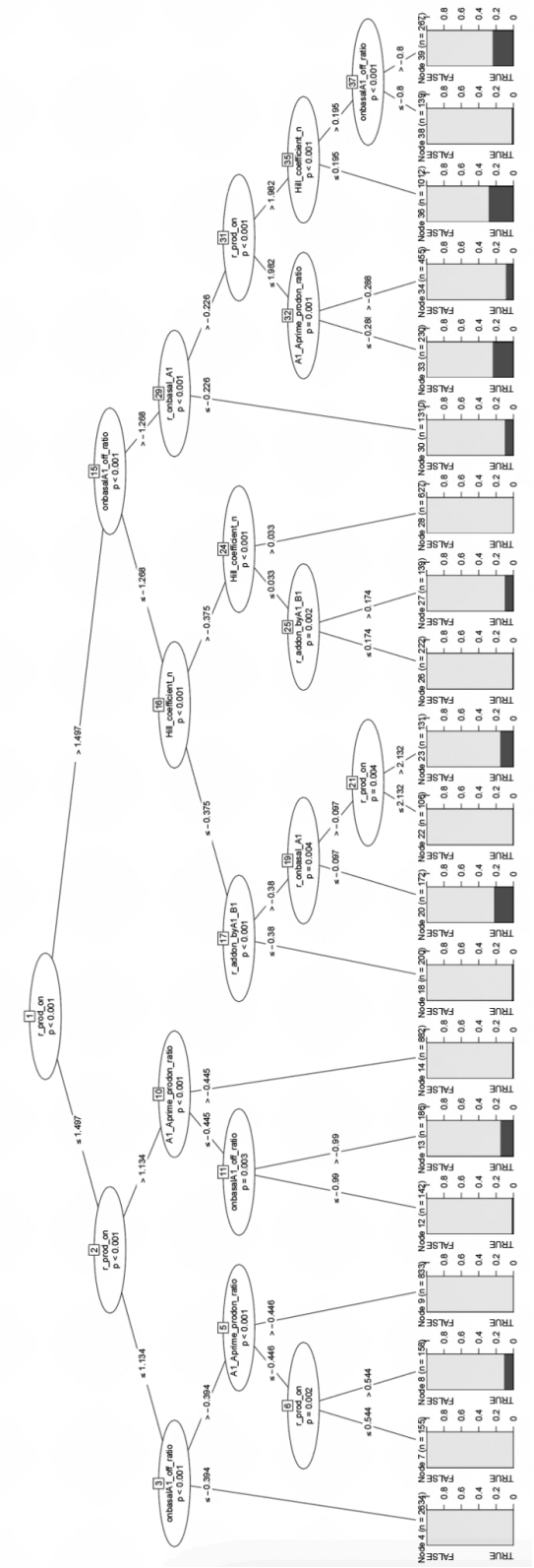


Fig. SN9: Decision tree analysis of right-skewed distribution shape in the heterozygous genotype

A. Decision tree trained on all parameter sets, on whether gene B is classified as right-skewed shape in the heterozygous genotype.

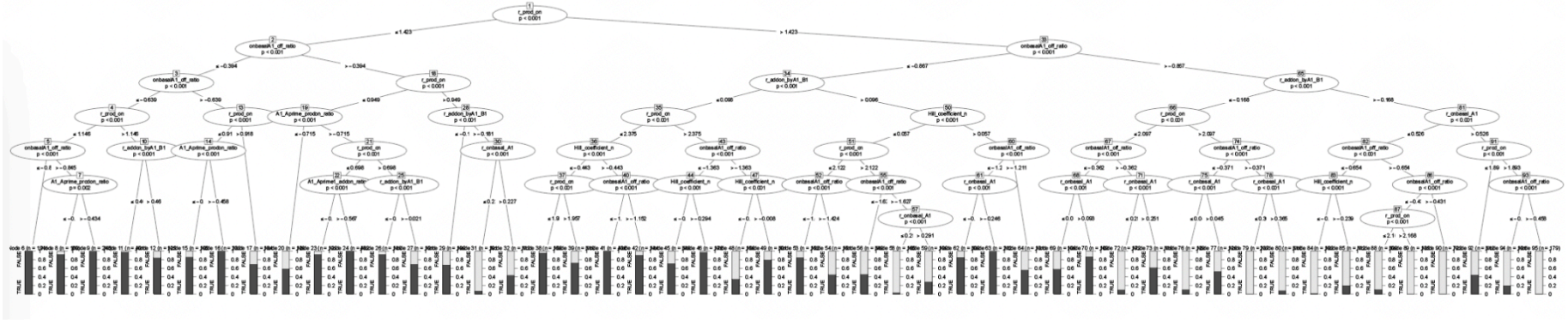


Fig. SN10: Decision tree analysis of low average symmetric distribution shape in the heterozygous genotype

A. Decision tree trained on all parameter sets, on whether gene B is classified as low average shape in the heterozygous genotype.