Supplementary Tables

Supplementary Table 1. Mechanistic models used for growth modeling in *growMod*, modified from ref¹.

	Model	Differential equation †	${\rm Analytical\ solution}^{\dagger}$	Linearized form ^{\dagger}
	Exponential	$\frac{\mathrm{d}y}{\mathrm{d}t} = ry(t)$	$y = y_0 e^{rt}$	$\ln(y) = \ln(y_0) + rt$
Control plants	Monomolecular	$\frac{\mathrm{d}y}{\mathrm{d}t} = r\left(K - y(t)\right)$	$y = K - (K - y_0) e^{-rt}$	$\ln \frac{1}{K - y} = \ln \frac{1}{K - y_0} + rt$
	Gompertz	$\frac{\mathrm{d}y}{\mathrm{d}t} = ry(t)\ln\frac{K}{y(t)}$	$y = K \left(\frac{y_0}{K}\right)^{e^{-rt}}$	$-\ln\left(-\ln\frac{y}{K}\right) = -\ln\left(-\ln\frac{y_0}{K}\right) + rt$
	Logistic [§]	$\frac{\mathrm{d}y}{\mathrm{d}t} = ry(t)\left(1 - \frac{y(t)}{K}\right)$	$y = \frac{Ky_0}{y_0 + (K - y_0) e^{-rt}}$	$\ln\frac{y}{K-y} = \ln\frac{y_0}{K-y_0} + rt$
	Weibull¶	$\frac{\mathrm{d}y}{\mathrm{d}t} = rmt^{m-1} \left(K - y(t) \right)$	$y = K - (K - y_0) e^{-rt^m}$	$\ln\left(\ln\frac{K-y_0}{K-y}\right) = \ln(r) + m\ln(t)$
	Quadratic	$\frac{\mathrm{d}y}{\mathrm{d}t} = b - 2at$	$y = c + bt - at^2$	$y = c + bt - at^2$
Stressed plants	Bell-shaped 1	$\frac{\mathrm{d}y}{\mathrm{d}t} = 2Aa\left(t - t_{max}\right)e^{a\left(t - t_{max}\right)^2}$	$y = Ae^{a(t - t_{max})^2}$	$\ln(y) = \ln(A) + a \left(t - t_{max}\right)^2$
	Bell-shaped 2	$\frac{\mathrm{d}y}{\mathrm{d}t} = A \left(b/t - a \right) t^b e^{-at}$	$y = At^b e^{-at}$	$\ln(y) = \ln(A) + b\ln(t) - at$
	Bell-shaped 3	$\frac{\mathrm{d}y}{\mathrm{d}t} = A(b - 2at)e^{bt - at^2}$	$y = Ae^{bt - at^2}$	$\ln(y) = \ln(A) + bt - at^2$
	$Linear^{\ddagger}$	$\frac{\mathrm{d}y}{\mathrm{d}t} = r$	$y = y_0 + rt$	$y = y_0 + rt$

[†] y is biomass; t denotes time; r is intrinsic growth rate for control plants or re-growth rate for stressed plants; K is upper asymptote of biomass for control plants in monomolecular, Gompertz, logistic and Weibull models; m determines the slope of growth in Weibull model; $t_{max} = \frac{b}{2a}$ is the time point (the center of the peak in bell-shaped curves) at which plant under stress shows the asymptotic maximum biomass (determined by A). Other parameters are constants.

 \S Only the three-parameter version of logistic model was considered. In this model, the lower asymptote is fixed at 0 and the inflection point falls strictly at y = K/2.

[¶]Weibull model with three parameters was considered, where $y_0 = 0$. The model can thus be simplified as $y = K\left(1 - e^{-rt^m}\right)$. It is reasonable in most cases. For example, at planting, the plant biomass is very close to zero (Archontoulis and Miguez, 2013).

 ‡ Linear growth for stressed plants is only modeled in the recovery phase.

Supplementary Table 2. Various machine-learning approaches integrated into HTPmod.

ID	Short name	Full name	R package [*]			
Regres	Regression Models					
1	BGLM	Bayesian generalized linear model	arm			
2	BLASSO	Bayesian Lasso	monomvn			
3	BRNN	Bayesian regularized neural networks	brnn			
4	GBM	Stochastic gradient boosting	gbm & plyr			
5	GLM	Generalized linear model				
6	GLMNET	Lasso and elastic-net regularized generalized linear models	glmnet & Matrix			
7	GP-Poly	Gaussian process with polynomial kernel	kernlab			
8	GP-Radial	Gaussian process with radial kernel	kernlab			
9	KNN	k-nearest neighbors	kknn			
10	LASSO	Lasso model	elasticnet			
11	MARS	Multivariate adaptive regression spline	earth			
12	MLR	Multivariate linear regression				
13	RF	Random forest	randomForest			
14	RIDGE	Ridge regression	elasticnet			
15	SVM-Radial	Support vector machines with linear kernel	kernlab			
16	SVM-Linear	Support vector machines with radial kernel	e1071			

Classification Models

1	CART	Classification and regression trees	rpart
2	GBM	Stochastic gradient boosting	gbm & plyr
3	GLMNET	Lasso and elastic-net regularized generalized linear models	glmnet & Matrix
4	KNN	k-nearest neighbors	kknn
5	LDA	Linear discriminant analysis	MASS
6	LLDA	Localized linear discriminant analysis	klaR
7	MARS	Multivariate adaptive regression spline	earth
8	MDA	Mixture discriminant analysis	mda
9	NBC	Naive Bayes	naivebayes
10	NNET	Neural network	nnet
11	PDA	Penalized discriminant analysis	mda
12	PLS	Partial least squares	pls
13	RDA	Regularized discriminant analysis	klaR
14	RF	Random forest	randomForest
15	SVM-Linear	Support vector machines with linear kernel	e1071
16	SVM-Radial	Support vector machines with radial kernel	kernlab

Note: some models (e.g., support vector machines and random forest) can be used for both regression and classification purpose.

Supplementary Table 3. Example data sets used to test the functionality of HTPmod.

ID	Study / Reference	Data for modeling or visualization	Tested in	Supplementary Figures [*]
1	(Chen et al., 2014) ²	HTP: using time-series image data to investigate plant growth	growMod,	Supplementary Figs. 3, 5
		and phenotypic components of drought responses in barley	htpdVis	and 11
		(Hordeum vulgare).		
2	(Chen et al., 2018) ³	HTP: using image-derived parameters to prediction plant	predMod,	Supplementary Figs. 6-8
		biomass accumulation in three consecutive barley	htpdVis	and 10
		experiments.		
3	(Jiao and Meyerowitz,	HTS: using ChIP-seq data to predict plant organ-specific gene	predMod	Supplementary Fig. 9;
	2010) ⁴ and (Chen et al.)	expression patterns.		
4	(Fahlgren et al., 2015)⁵	HTP: using image data to predict plant growth under different	growMod	Supplementary Figs. 4 and
		water conditions in Setaria.		5
5	(Jiao and Meyerowitz,	HTS: using TRAP-seq data to study flower organ-specific gene	htpdVis	Fig. 4
	2010)4	expression patterns in Arabidopsis.		
6	(Smaczniak et al., 2017) ⁶	HTS: using SELEX-seq data to reflect the difference of DNA	htpdVis	Supplementary Fig. 12
		binding specificity of floral homeotic protein complexes and to		
		further predict organ-specific target genes.		
7	(Song et al., 2016) ⁷	HTS: combining RNA-seq and ChIP-seq data to illuminate the	predMod	Figs. 2 and 3 ;
		relationship of differential gene expression patterns and the		Supplementary Fig. 7
		combinatorial regulation by multiple ABA-related TFs.		

8	(Wang et al., 2015) ⁸	HTS: use Hi-C and ChIP-seq data to describe chromatin states	htpdVis	Supplementary Fig. 13
		in Arabidopsis		
9	(Zhu et al., 2018) ⁹	HTP: profiling of 980 metabolites in 442 tomato (Solanum	htpdvis	Supplementary Fig. 13
		lycopersicum) accessions.		

* Note: reanalysis of published data by HTPmod (see online document for how data were collected). Results are shown in the Supplementary Figures as below. Although the examples were selected from studies in plants, HTPmod can broadly be used for studies in any other organisms. HTP: high-throughput phenotyping; HTS: high-throughput sequencing.

Supplementary Figures

HTPmod 🕋 Home 🚍 Apps - 🖉 About

HTPmod

A Web-based Platform for High-Throughput Data Modeling and Visualization

A Shiny Application for Modeling and Visualizing Large-Scale, High-Dimensional Datasets



Supplementary Figure 1. Screenshot of the homepage of HTPmod.



Features (columns)

Supplementary Figure 2. HTPmod accepts the simplest tables files (with rows as individuals and columns as features; the header line is required) as the only input. In the modules of *predMod* and *htpdVis*, HTPmod automatically determine the column types according their contents: columns containing non numeric values are considered as either annotation for individuals (in grey) or categorical features (in orange). Categorical class columns can be either used for color schema purpose in plots or the output feature in the classification analysis. Users need to convert

numeric categorical features into non-numeric values if input files containing numeric values for categorical purposes.

In the subsections, we demonstrated how to use HTPmod functionalities to explore highthroughput datasets from the selected studies in **Supplementary Table 3**.

HTPmod 🗠 growMod			
▶ 1. Load Data ● Example data	Show 10 ventries	Modeling 0° Group-based Modeling	Performance
O Upload your data Example Data	Data summary at plant level	# pointe	1121KN001
Chen et al., 2014; Normal-growth plants	plant 🔶 g	roup 🔶 valid / total 🗘	Models
The example file is from Chen et al., 2014 the Plant Cell. You can download it here	1 1121KN001 Ap	ex 46 / 48	•• $K - (K - y_0)e^{-T}$, Webult, $R^2 = 0.9860$ •• $Ky_0 (y_0 + (K - y_0)e^{-T})$, Logititic, $R^2 = 0.9834$ •• $K(y_0 / K^2)^*$, Compate, $R^2 = 0.874$ •• $K - (K - y_0)e^{-T}$, Monomolecular, $R^2 = 0.8568$
▶ 2. Input Data	2 1121KN002 Ap	ex 46/48	S → y ₀ e ⁻ : Exponential, R ⁻ = 0.9188 → y ₀ + n, Linear, R ² = 0.8840
Which measurement for modeling?	3 1121KN003 Ap	ex 46/48	
volume.fluo.iap	4 1121KN004 Ap	ex 46 / 48	Auo.lar
▶ 3. Select Models	5 1121KN005 Ap	ex 13/48	ohme
3.1. Which type of growth model?	6 1121KN006 Ap	ex 46 / 48	
Normal	7 1121KN007 Ap	ex 46 / 48	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
3.2. Models (empty to select all)	8 1121KN008 Ap	ex 46 / 48	and the second se
Please select models below	9 1121KN009 Ap	ex 38 / 48	00
 ↓ 4. Customization ☑ Check data completeness before modeling 	10 1121KN145 Ba Showing 1 to 10 of 156 entries	:keCS 45 / 47	8
Show data points in the plot	Previous 1 2	3 4 5 16 Next	
Minimum data points for modeling	▲ Save as tab delimited .bxt		▲ Save as png ▲ Save as pdf ▲ Save as svg ▲ Dowload parameters
x label of growth curve			
y label of growth curve			

1. Growth modeling with the *growMod* module

Supplementary Figure 3. Screenshot of the *growMod* module. Example shows the growth modeling of barley plants under normal growth conditions².

1. Load Data 🕜	S My	Data 0° Plant-	based Modeling	C Group-based Modeling	Per Per	rformance
 Example data Upload your data 	Show 1	10 v entries	Search:		Growth	Curves (plant)
Example Data	Data au	minary at plant level				Dp1AA000001
Fahlgren et al., 2015; Group by treatment		plant	group	# points valid / total	40	Models
The example file is from Fahlgren et al., 2015 Molecular Plant. You can download it here	1	Dp1AA000001	100% FC	11 / 11		
▶ 2. Input Data	2	Dp1AA000004	100% FC	11 / 11	8	 K -(K γ₀)e²⁷, Monomolecular, R² = 0.9485 γ₀e²⁷, Exponential, R² = 0.7713
Which measurement for modeling?	3	Dp1AA000007	100% FC	11 / 11		-• y ₀ +rt, Linear, R*= 0.9628
fw_biomass ·	4	Dp1AA000008	100% FC	11 / 11	SSE	
2 Salact Madala	5	Dp1AA000012	100% FC	11 / 11	20	
3.1. Which type of growth model?	6	Dp1AA000017	100% FC	11 / 11	é	
Normal	7	Dp1AA000018	100% FC	11 / 11	ę.	- ////
3.2. Models (empty to select all)	8	Dp1AA000019	100% FC	11 / 11		
Please select models below	9	Dp1AA000020	100% FC	11 / 11	0	
A Customization	10	Dp1AA01003	100% FC	11 / 11		0 10 20 30
Customization Check data completeness before modeling	Showing	g 1 to 10 of 478 entri	es			Day
		Previous 1	2 3 4	5 48 Next		

Supplementary Figure 4. Growth modeling of Setaria plants².



Supplementary Figure 5. Performance of growth models in barley² (top panel; using the example data "Chen et al., 2014; Normal-growth plants") and Setaria⁵ (bottom panel; using the example data "Fahlgren et al., 2015; Group by genotype"). Results showed that Weibull model shows the best performance for modeling growth of barley plants while Gompetz model does for Setaria plants. Default parameter settings under the "Performance" page were used in the two analyses.

2. Prediction models: the predMod module



Supplementary Figure 6A. Classification of three consecutive HTP experiments³ based on imagederived features (the example data "Chen et al., 2018; Regression or Classification"). Note that some models were failed to run on this dataset; all other models perfectly separated the experiments. Default parameter settings for all the classification models were used in the analysis.



Supplementary Figure 6B. Prediction of plant biomass accumulation with image-derived parameters, using HTP data from ref³ (the example data "Chen et al., 2018; Regression or Classification"). Each panel shows the prediction result for a regression model. See **Supplementary Table 2** for the full model names. Default parameter settings for all the regression models were used in the analysis.

Supplementary Figure 6. (A-B) Apply various prediction models on HTP data.



Supplementary Figure 7. Evaluation of the model performance in the prediction of plant biomass based on image-derived paramters³ (left; the example data "Chen et al., 2018; Regression or Classification") or gene expression changes based on transcription factor binding data⁷ (right; the example data "Song et al., 2016; Regression"). All parameters were set in default in the analysis.



Supplementary Figure 8. Calculate the relative importance of features in regression models. Random forest (RF) model was used for prediction of barley biomass accumulation, using imagederived feature data in three consecutive HTP experiments³ (the example data "Chen et al., 2018; Regression or Classification"). The results are consistent to that in the original study.



Supplementary Figure 9A. Confusion matrixes for measuring the classification accuracy of different models. Note that some models were failed to run on this dataset.



Supplementary Figure 9B. Evaluation of the performance of different classification models. Results show that RF (random forest) and KNN (k-nearest neighbors) outperform other models.

Prediction of Cluster, based on KNN





Supplementary Figure 9C. Influence of feature selection on classification performance. Results from both models show that at least four features are required to retain good classification performance. This suggests that (1) plant organ differentiation is not controlled by a single factor, but rather depends on the joint behavior of multiple factors and (2) that there might be redundancy between factors.

Supplementary Figure 9. **(A-C)** Apply classification models to predict plant organ-specific gene expression patterns (using the example data "Chen et al., 2018; Classification"). All parameters were set as default in the analysis.

3. High-throughput data visualization with the *htpdVis* module

HTPmod III htpdVis		
 1. Load Data Example data Upload your data Example Data Chen et al., 2018; Consecutive HTP experiments	My Data PCA \$ 1-SNE MDS SOM Number of components In the PCA plot: Show component scores Show component scores 1 2 4 5 4 State Show component scores 2 4 5 4 5 Show component scores Show factor loadings PCA: Principal Component Analysis Show factor loadings Show factor loadings	C K-MC d. HCA EX Misc Azimuthal direction 500 0 00 00 00 00 00 00 00 00 00 00 00 0
The example tile is from Chen et al., 2018 Gigascience, You can download it here > 2. Data Summary All input data, with size of: [1] 929 41 Numeric data matrix for visualization, with size of: [1] 929 37 > 3. Customization Point colors by: Experiment	PC1 % % variance explained 0	SD PCA
Point symbols by: Treatment Cata preprocessing: Center Scale	Color FLUO Geometric NIR	Pot Pot Pot Pot Pot Pot Pot Pot Pot Pot

Supplementary Figure 10A. Principal component analysis (PCA).

HTPmod III htpdVis		
I. Load Data	SMy Data & PCA ★t-SNE ♥ MDS ♥ SOM	C K-MC _L HCA S Misc
Example Data		
Chen et al., 2018; Consecutive HTP experiments	t-SNE: t-Distributed Stochastic Neighbor Embedding	t-SNE
The example file is from Chen et al., 2018 GigaScience. You can download it here	t-SNE Transmoot	3D t-SNE
▶ 2. Data Summary	20 -10 -5 -0 -5 10 15	
All input data, with size of:	g - A stress	
[1] 929 41	° t-SNE1	
Numeric data matrix for visualization, with size of:	₽ - -	
[1] 929 37		
► 3. Customization	Experiment • KN1121	
Point colors by:	t-SNE2 • KN1130	
Experiment		
Point symbols by:		1900
Treatment		Treatment • KN1121
Data preprocessing:		A stress • KN1130
✓ Center ✓ Scale	8	

Supplementary Figure 10B. t-distributed stochastic neighbor embedding (t-SNE).



Supplementary Figure 10C. Multidimensional scaling (MDS).

HTPmod III htpdVis				
N1 Load Data 0	🛢 My Data 🛭 🗞 PCA 🕸 t-S		K-MC 🚠 HCA 🖾 Misc	
Example data Upload your data Example Data Chan et al. 2018: Consecutive HTP experiments	Dimensions of the grid x-dimension	y-dimension	Topology (Hexago Rectang	of the grid nal gular
The example file is from Chen et al., 2018 GigaScience. You can download it here	SOM: Self-Organizing Map	9 10 1 2 3 4 5	SOM + PCA	
▶ 2. Data Summary		Figure legend		Cluster colors
All input data, with size of: [1] 929 41 Numeric data matrix for visualization, with size of:	SOM	Treatment Experiment • KN1121 • KN1120 • KN1127	Mapping plot	Cluster 2 Cluster 3 (151) Cluster 1 (172)
[1] 929 37 > 3. Customization				(170) (141) (169)
Point colors by:				
Experiment			PCA with SOM clustering	#observations in clusters
Point symbols by:	Counto plat	Cluster colored by 'Experiment'		KN1137 163 124 3 19 0 0
Treatment	170			and the second sec
Data preprocessing:	100 155 140 130 130			N1121 0 0 0 189 141 1 2 3 4 5 6
			PC1 (35.87%)	Cluster

Supplementary Figure 10D. Self-organizing map (SOM).



Supplementary Figure 10E. K-means clustering (K-MC).



Supplementary Figure 10F. Hierarchical cluster analysis (HCA) with heatmaps.

Supplementary Figure 10. **(A-F)** Apply different visualization tools to the same dataset obtained from ref³ (i.e., the example data "Chen et al., 2018; Consecutive HTP experiments").



Supplementary Figure 11. Reanalysis of the HTP data from ref² (i.e., the example data "Chen et al., 2014; Barley HTP data"). Plants in the same genotypes under the same growth condition show similar phenotypic profile, as show in SOM (left; 6x6; colored by "Genotype" and shaped by "Treatment") and HCA (right; bars colored by "Genotype").



Supplementary Figure 12. Visualize the DNA binding specificity (based on SELEX-seq data) of floral homeotic protein complexes⁶ (the example data "Smaczniak et al., 2017; DNA binding specificity of MADS proteins") using t-SNE (left; dimension=3, data points colored by "Cluster") and HCA (right; left bars colored by "Cluster").



Supplementary Figure 13. PCA on chromatin states identified by 16 epigenetic data sets over the Arabidopsis epigenome at 400-bp resolution⁸ (left; the example data "Wang et al., 2015; Chromatin states") or a metabolome dataset with profiling data of 980 metabolites in 442 tomato accessions⁹ (right; the example data "Zhu et al., 2018; Tomato metabolomes"). Results are similar to the original studies.

Supplementary References

- Chen, D. Dissecting and Modeling the Phenotypic Components of Plant Growth and Drought Responses Based on High-throughput Image Analysis. (Doctoral Diss. Martin-Luther-Universität Halle-Wittenberg) (2017).
- 2. Chen, D. *et al.* Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell* **26**, 4636–4655 (2014).
- Chen, D. *et al.* Predicting plant biomass accumulation from image-derived parameters.
 Gigascience (2018). doi:10.1093/gigascience/giy001
- 4. Jiao, Y. & Meyerowitz, E. M. Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol. Syst. Biol.* **6**, (2010).

- 5. Fahlgren, N. *et al.* A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in Setaria. *Mol. Plant* **8**, 1520–1535 (2015).
- Smaczniak, C., Muiño, J. M., Chen, D., Angenent, G. C. & Kaufmann, K. Differences in DNA-binding specificity of floral homeotic protein complexes predict organ-specific target genes. *Plant Cell* 29, tpc.00145.2017 (2017).
- 7. Song, L. *et al.* A transcription factor hierarchy defines an environmental stress response network. *Science (80-.).* **354**, aag1550-aag1550 (2016).
- Wang, C. *et al.* Genome-wide analysis of local chromatin packing in Arabidopsis thaliana.
 Genome Res. 25, 246–256 (2015).
- Zhu, G. *et al.* Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell* **172**, 249–261.e12 (2018).