

DEtime: Inferring the perturbation time from biological time course data

Jing Yang

2017-01-14

Abstract

Time course data is often used to study the dynamics in a biological process after perturbation at certain time. Inferring the perturbation time under different scenarios in a biological process allows us to identify these critical moments and focus on any following activities in the process, which is of critical importance in understanding likely causal relationships. In DEtime package, we propose a Bayesian method to infer the perturbation time from a control and perturbed system. A non-parametric Gaussian Process regression is applied in deriving the posterior distribution of the perturbation point. This vignette explains how to use the package. For further information of the algorithm, please refer to our paper:

Jing Yang, Christopher A. Penfold, Murray R. Grant and Magnus Rattray, Inferring the perturbation time from biological time course data, Bioinformatics, 32(19): pp 2956-2964, 2016

Description

This package implements the Gaussian regression framework for perturbation time point inference in a two sample case. The package contains two main functions: **DEtime_infer**, which is used to find out perturbation point of genes, and **DEtime_rank**, which is used to filter these silent genes before carrying out perturbation point inference by **DEtime_infer** function.

The package works on the time course data from a wild-type and a perturbed system. Acting upon pre-defined testing perturbation time, the package goes over these perturbation time candidates and derives their likelihoods. From Bayes' theory, under a uniform prior assumption, the posterior distribution of the tested perturbation time is derived from their corresponding likelihoods. *Maximum a posterior (MAP)*, *mean* or *median* of the posterior distribution can be taken as the solution to the estimated perturbation time point.

Details

Package: DEtime
Type: Package
Version: 1.1
Date: 2017-01-14
License: GPL-3

Author(s)

Jing Yang

Maintainer

Jing Yang ynnjing@gmail.com

Functions

DEtime_infer - Perturbation time inference

Description

DEtime_infer is the main function in DEtime Package, which applies a mixedGP kernel to time course data under control and perturbed conditions. It returns the posterior distribution of these predefined perturbation time candidates and relevant statistical estimations of the inferred perturbation time point.

Usage

```
DEtime_infer(ControlTimes, ControlData, PerturbedTimes, PerturbedData, TestTimes=NULL,
gene_ID=NULL, bound.lengthscale=NULL)
```

Arguments

- **ControlTimes**: experimental time points at which the time course data for the control condition are measured. They can either be ordered by time, for instance t1,t1,t2,t2,... or ordered by replicated time, for instance t1,t2,...,t1,t2,...
- **ControlData**: The measured time course data under control condition. The data is a matrix where each row represents the time course data for one particular gene. The measurements have to match the time points in **ControlTimes**.
- **PerturbedTimes**: experimental time points at which the time course data for the perturbed condition are measured. They can either be ordered by time, for instance t1,t1,t2,t2,... or ordered by replicated time, for instance t1,t2,...,t1,t2,... The replicates do not have to be the same everywhere. **ControlTimes** and **PerturbedTimes** can differ from each other.
- **PerturbedData**: The measured time course data under perturbed condition. The data is a matrix where each row represents the time course data for one particular gene. The measurements have to match the time points in **PerturbedTimes**.
- **TestTimes**: perturbation time points which will be evaluated by **DEtime_infer** function. **TestTimes** has to be in the range of times and evenly spaced. If this input is missing, **TestTimes** is set to 50 time points evenly spaced between the minimum of **ControlTimes** and **PerturbedTimes** and the maximum of **ControlTimes** and **PerturbedTimes**.
- **gene_ID**: The IDs of genes investigated in the algorithm. If this value is missing, '1', '2', '3', ... will be used instead.
- **bound.lengthscale**: the bounds used for the lengthscale parameter in the RBF kernel used in the model. We recommend you not to change this parameter unless necessary.

Returns

The function will return a **DEtimeOutput** object which contains:

- **result**: statistical estimations for the inferred perturbation time, which includes:
 - **\$MAP**: *maximum a posterior* solution to the inferred perturbation time
 - **\$mean**: mean of the posterior distribution of the inferred perturbation time
 - **\$median**: median of the posterior distribution of the inferred perturbation time
 - **\$ptl5**: 5 percentile of the posterior distribution of the inferred perturbation time

- **\$ptl95**: 95 percentile of the posterior distribution of the inferred perturbation time
- **\$posterior**: posterior distribution of the tested perturbation time points
- **\$model**: optimized GP model which will be used for later GP regression work
- **\$best_param** : optimized hyperparameter for the optimized GP model
- **\$ControlTimes**: experimental time points for the control condtion which will be used for future print or plot functions
- **\$ControlData**: experimental time measurements under the control condtion which will be used for future print or plot functions
- **\$PerturbedTimes**: experimental time points for the perturbed condtion which will be used for future print or plot functions
- **\$PerturbedData**: experimental time measurements under the perturbed condtion which will be used for future print or plot functions
- **\$TestTimes**: tested perturbation time points
- **\$gene_ID**: the ID of genes for the data

Details

Control and perturbed data can be measured at different time points with differnt numbers of replicates. However, it would be reasonable to have control and perturbed data measured at roughly the same region. to facilitate the estimation of perturbation point.

Examples

```
## read simulated example data
library("DEtime")
data(SimulatedData)
res <- DEtime_infer(ControlTimes = ControlTimes, ControlData = ControlData,
                    PerturbedTimes = PerturbedTimes, PerturbedData=PerturbedData)
```

DEtime_rank - Rank time course data by log-likelihood ratio

Description

DEtime_rank is the function used for filtering silent genes in DEtime Package. In this function, an independent GP and an integrated GP are applied to model the time course data under control and perturbed conditions, respectively. The log-likelihood ratio of the GP modeling result is used as an indication of the differential expression of the studied gene. A higher rank generally indicates better differential expression.

Usage

```
DEtime_rank(ControlTimes, ControlData, PerturbedTimes, PerturbedData, gene_ID=NULL, save-
file=TRUE)
```

Arguments

- **ControlTimes**: experimental time points at which the time course data for the control condition are

measured. They can either be ordered by time, for instance t1,t1,t2,t2,... or ordered by replicated time, for instance t1,t2,...,t1,t2,...

- **ControlData**: The measured time course data under control condition. The data is a matrix where each row represents the time course data for one particular gene. The measurements have to match the time points in **ControlTimes**.
- **PerturbedTimes**: experimental time points at which the time course data for the perturbed condition are measured. They can either be ordered by time, for instance t1,t1,t2,t2,... or ordered by replicated time, for instance t1,t2,...,t1,t2,... The replicates do not have to be the same everywhere. And **ControlTimes** and **PerturbedTimes** do not have to be exactly the same.
- **PerturbedData**: The measured time course data under perturbed condition. The data is a matrix where each row represents the time course data for one particular gene. The measurements have to match the time points in **PerturbedTimes**.
- **gene_ID**: The IDs of genes investigated in the algorithm. If this value is missing, '1', '2', '3', ... will be used as the gene IDs instead.
- **savefile**: A BOOLEAN parameter used to indicate if the ranking list will be saved in a file or not. If set to TRUE, the result will be saved in Detime_rank.txt

Returns

The function will return a table which contains the gene_IDs as the first column and the associated loglikelihood ratio as the second column.

Details

Control and perturbed data can be measured at different time points with different numbers of replicates. However, it would be reasonable to have control and perturbed data measured at roughly the same region. to facilitate the estimation of perturbation point.

Examples

```
## read simulated example data
library("DEtime")
data(SimulatedData)
res <- DEtime_rank(ControlTimes = ControlTimes, ControlData = ControlData,
                   PerturbedTimes = PerturbedTimes, PerturbedData=PerturbedData, savefile=TRUE)
```

print_DEtime - print the results from DEtime function

Description

print_DEtime prints the results returned from **DEtime_infer** function, which will show the **gene_ID** associated with **MAP**, **mean**, **median**, **ptl5** (lower 5 percentile) and **ptl95** (upper 5 percentile) of the posterior distribution of inferred perturbation time points.

Usage

```
print_DEtime(DEtimeOutput)
```

Argument

- **DEtimeOutput**: the result from **DEtime_infer** function

Example

```
library("DEtime")
## read simulated example data
data(SimulatedData)
res <- DEtime_infer(ControlTimes = ControlTimes, ControlData = ControlData,
                    PerturbedTimes = PerturbedTimes, PerturbedData=PerturbedData)

## ControlData is accepted
## PerturbedData is accepted
## gene IDs are not provided. Numbers are used instead.
## Testing perturbation time points are not provided. Default one is used.
## gene 1 is done
## gene 2 is done
## gene 3 is done
## gene 4 is done
## gene 5 is done
## DEtime inference is done.
## Please use print_DEtime or plot_DEtime to view the results.
```

```
print_DEtime(res)
```

```
## Perturbation point inference results from DEtime package:
## =====
##   gene_ID   MAP   mean median   pt15 pt195
## 1         1  9.796  9.855  9.796  9.388 10.61
## 2         2 20.000 17.367 17.959 10.612 20.00
## 3         3  9.388  9.165  9.388  8.163 10.20
## 4         4 13.469 11.477 13.061  5.714 13.88
## 5         5  9.388  7.749  8.980  2.857 10.20
## =====
```

plot_DEtime - plot the results of DEtime function

Description

plot_DEtime plots the results returned from **DEtime_infer** function. The produced figures show the the posterior distribution of inferred perturbation time points on the upper panel and Gaussian Regression of the original data on the lower panel. Please note that by default the MAP solution of the perturbation point is taken as the optimized estimate to the perturbation point and Gaussian Regression is derived based upon this estimated perturbation point.

Usage

```
plot_DEtime(DEtimeOutput, BestPerturbPos=NULL, plot_gene_ID=NULL)
```

Argument

- **DEtimeOutput**: the result from **DEtime_infer** function
- **BestPerturbPos**: to choose which statistical inference of the posterior distribution of the perturbation points to be used as the optimized estimate to the final perturbation point. You can set this parameter to “mean”, “median” or “MAP”, so that the corresponding statistical results from the posterior distribution of the perturbation points will be used in Gaussian regression plotting. If not given, MAP solution will be used.

- **plot_gene_ID**: the gene_IDs of those genes whose GP regression and posterior distribution of the perturbation time points will be plotted. If not supplied, all the genes will be plotted.

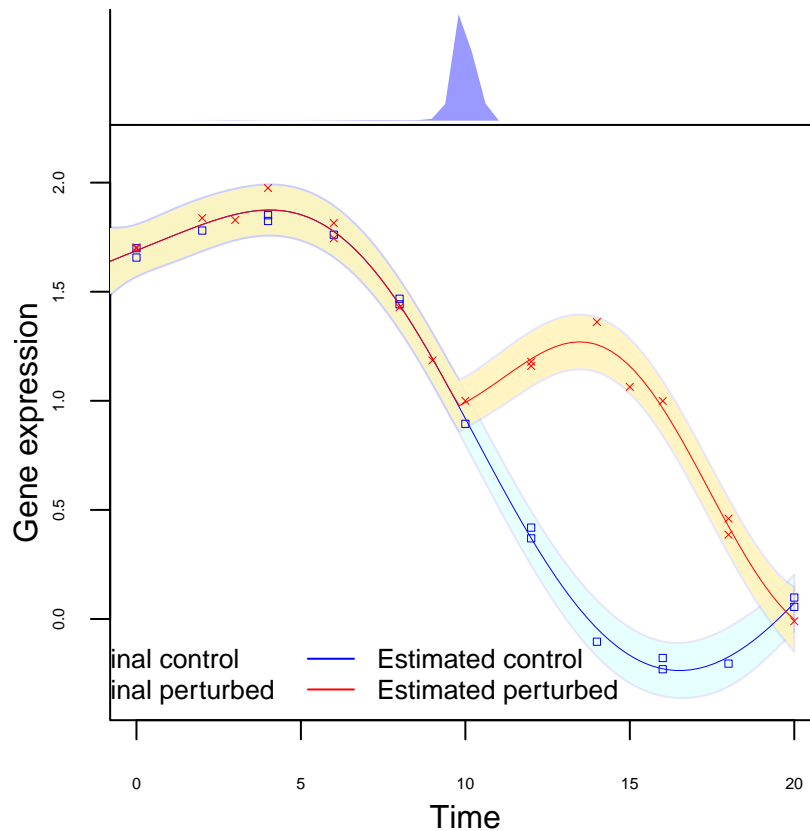
Example

```
library("DEtime")
## read simulated example data
data(SimulatedData)
res <- DEtime_infer(ControlTimes = ControlTimes, ControlData = ControlData,
                    PerturbedTimes=PerturbedTimes, PerturbedData=PerturbedData)
```

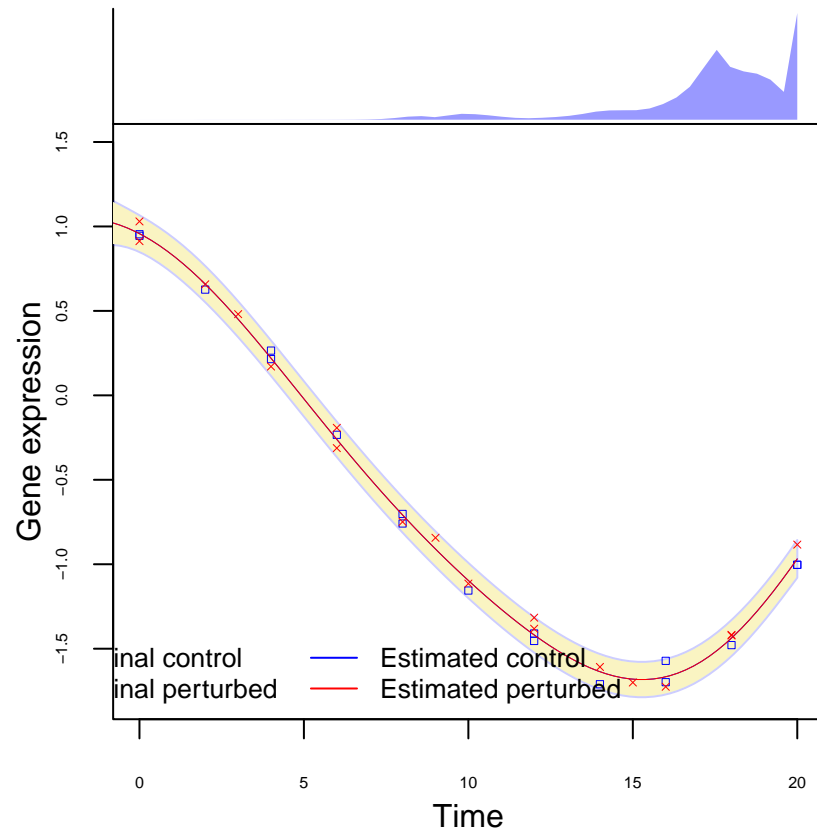
```
## ControlData is accepted
## PerturbedData is accepted
## gene IDs are not provided. Numbers are used instead.
## Testing perturbation time points are not provided. Default one is used.
## gene 1 is done
## gene 2 is done
## gene 3 is done
## gene 4 is done
## gene 5 is done
## DEtime inference is done.
## Please use print_DEtime or plot_DEtime to view the results.
```

```
plot_DEtime(res)
```

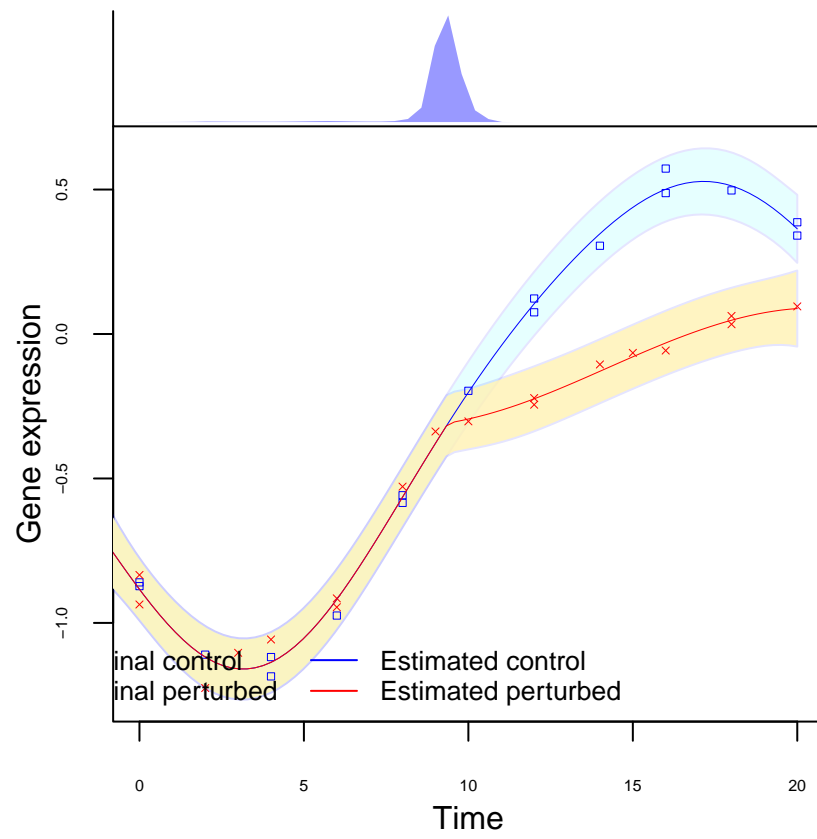
```
## All genes will be plotted
## 1 is plotted
```



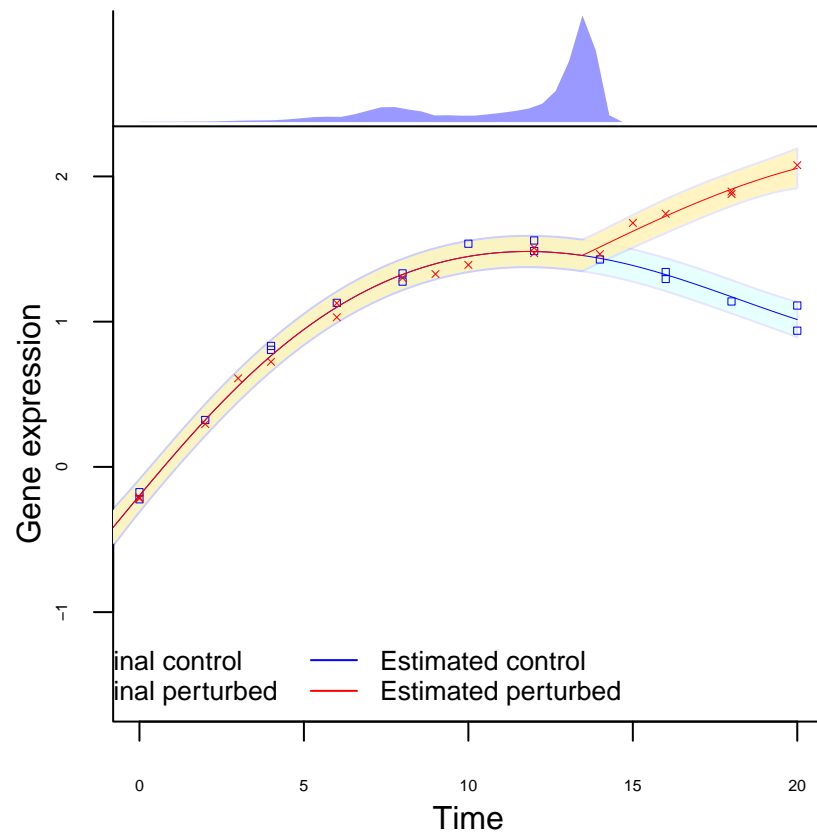
2 is plotted



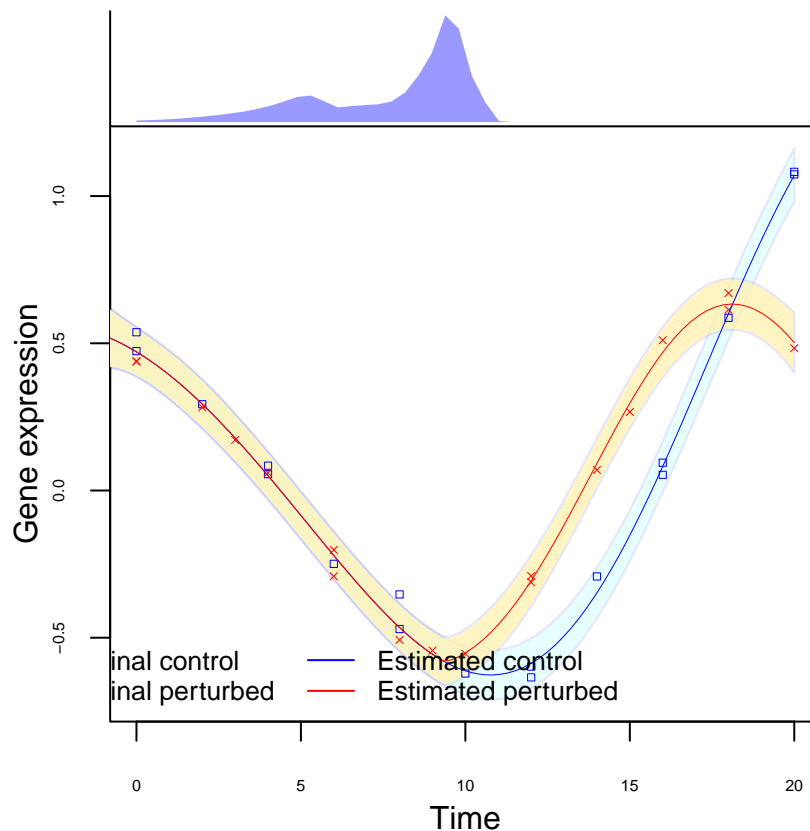
3 is plotted



4 is plotted



5 is plotted



Run the package on the real data used in our paper and plot the one with top loglikelihood ratio

Descriptions of the real data

In this experiment, the aim is to study the transcriptional change occurring in Arabidopsis following inoculation with *P. syringae* pv. tomato DC3000 (PtoDC3000) versus the disarmed strain Pto DC3000hrpA

The data contain two different time series:

- infection of Arabidopsis with virulent *Pseudomonas syringae* pv. tomato DC3000, which leads to disease development (perturbed condition 1), referred as ControlData in the dataset
- infection of Arabidopsis with the disarmed strain DC3000hrpA (perturbed condition 2), referred as PerturbedData in the dataset

In this example, the perturbation time between perturbed condition 1 and perturbed condition 2 is inferred.

```
library("DEtime")
## import data
data(RealData)
## calculate the loglikelihood ratio for each gene
res_rank <- DEtime_rank(ControlTimes = ControlTimes, ControlData = ControlData,
                        PerturbedTimes = PerturbedTimes, PerturbedData = PerturbedData)
```

```
## [1] "ControlData is accepted"
## [1] "PerturbedData is accepted"
```

```

## [1] "gene IDs are not provided. Numbers are used instead"
## rank list saved in Detime_rank.txt

## inferring the perturbation point by Detime_infer
res <- Detime_infer(ControlTimes = ControlTimes, ControlData = ControlData,
                    PerturbedTimes = PerturbedTimes, PerturbedData=PerturbedData)

## ControlData is accepted
## PerturbedData is accepted
## gene IDs are not provided. Numbers are used instead.
## Testing perturbation time points are not provided. Default one is used.
## gene 1 is done
## gene 2 is done
## gene 3 is done
## gene 4 is done
## gene 5 is done
## gene 6 is done
## gene 7 is done
## gene 8 is done
## gene 9 is done
## gene 10 is done
## gene 11 is done
## gene 12 is done
## gene 13 is done
## gene 14 is done
## gene 15 is done
## gene 16 is done
## gene 17 is done
## gene 18 is done
## gene 19 is done
## gene 20 is done
## Detime inference is done.
## Please use print_Detime or plot_Detime to view the results.

## Print a summary of the results
print_Detime(res)

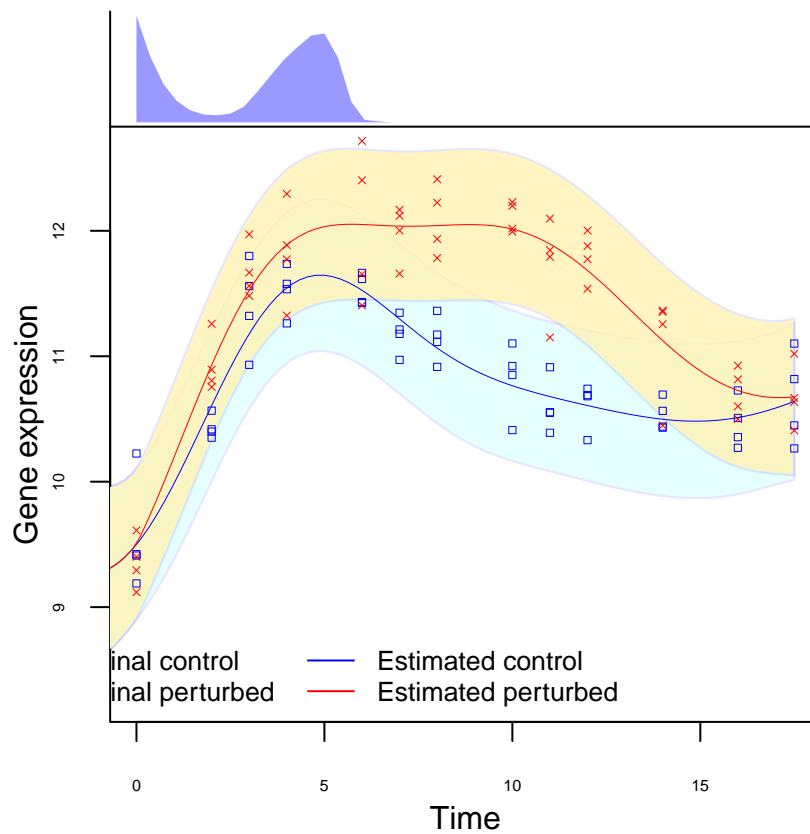
## Perturbation point inference results from Detime package:
## =====
##      gene_ID    MAP    mean median    pt15    pt195
## 1         1  0.000  3.082  3.929 0.0000  5.357
## 2         2  1.429  4.852  3.929 0.3571 13.214
## 3         3  6.786  5.008  5.357 0.7143  8.214
## 4         4  8.214  7.582  7.500 2.1429 14.643
## 5         5  8.929  7.923  8.214 2.8571 14.286
## 6         6  7.857  6.757  7.143 3.2143  8.929
## 7         7 17.500 10.949 12.143 2.8571 17.500
## 8         8 10.357  8.958  9.643 5.3571 10.714
## 9         9 11.071  9.055  9.643 0.7143 16.786
## 10        10  2.143  2.341  2.143 0.3571  7.143
## 11        11  1.429  6.915  2.143 0.3571 17.143
## 12        12  1.429  8.312  8.571 0.7143 16.786
## 13        13  2.500  9.184  8.571 1.7857 17.500

```

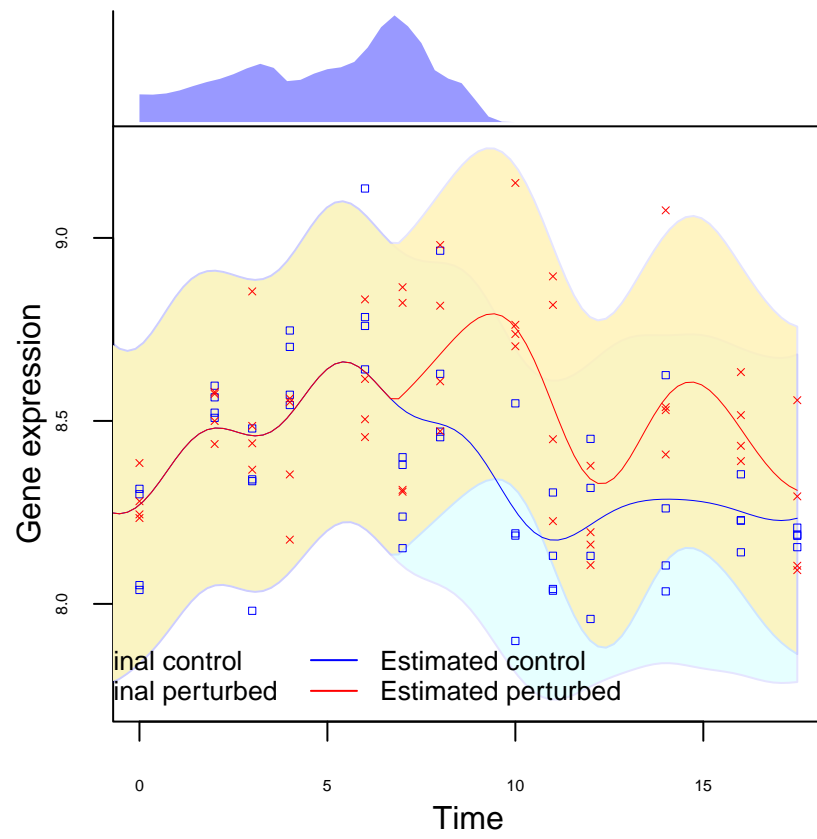
```
## 14      14  0.000  2.275  1.071  0.0000  8.571
## 15      15 10.000 10.122 10.357  1.4286 17.143
## 16      16  2.143  1.952  2.143  1.4286  2.500
## 17      17 13.571 12.163 12.857  7.1429 15.000
## 18      18  3.214  2.970  3.214  1.7857  3.571
## 19      19 17.500 14.052 15.000  7.5000 17.500
## 20      20  1.429  8.623  8.929  0.3571 16.786
## =====
```

```
## plot the gene with loglikelihood ratio > 5
plot_Detime(res, plot_gene_ID=as.character(which(res_rank[,2]>5)))
```

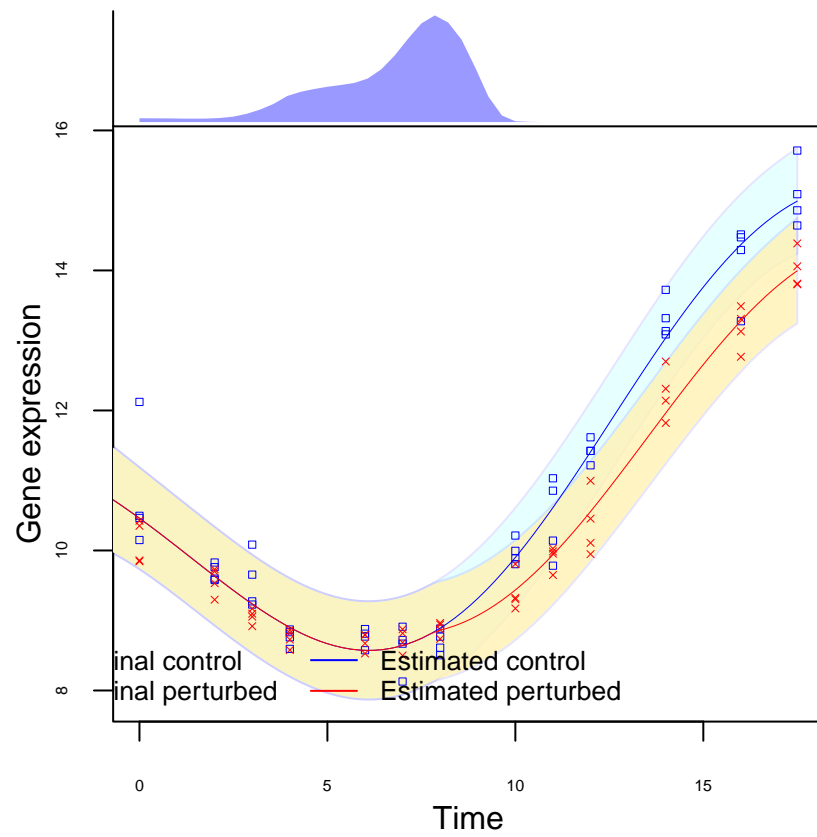
1 is plotted



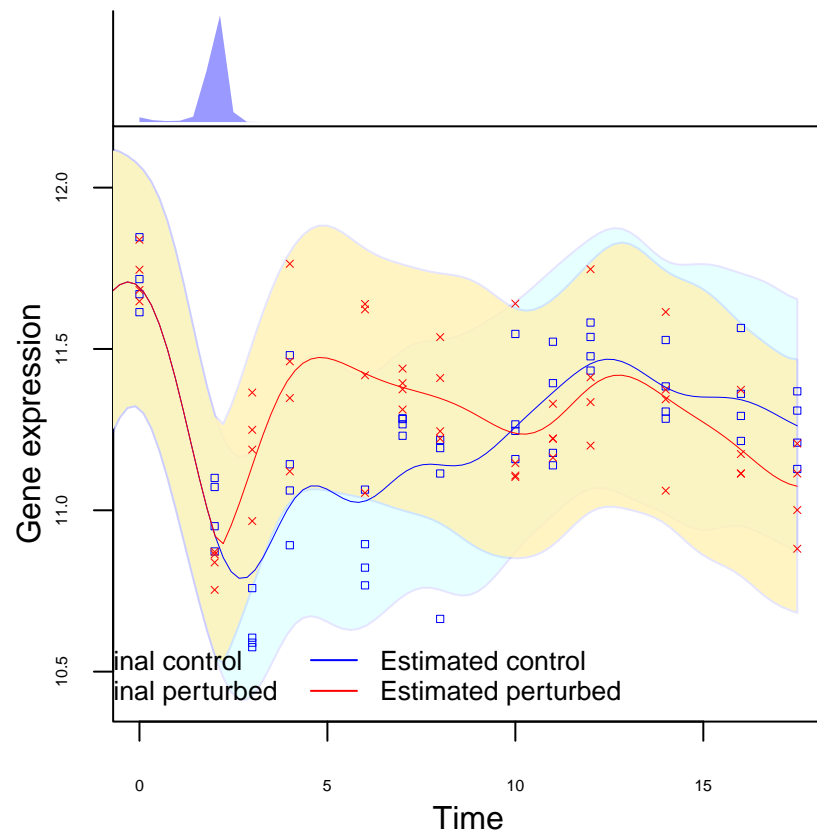
3 is plotted



6 is plotted



16 is plotted



18 is plotted

