Change Point Detection For Clustered Expression Data In Short Time Series.

Miriam Sieg (miriam.sieg@charite.de)
Institute of Biometry and Clinical Epidemiology, Charité - Universitätsmedizin Berlin
Lina Katrin Sciesielski
Department of Neonatology, Charité - Universitätsmedizin Berlin
Karin Kirschner
Institute of Translational Physiology, Charité - Universitätsmedizin Berlin
Jochen Kruppa
Institute of Medical Informatics, Charité - Universitätsmedizin Berlin

Research Article

Keywords: time series, change point detection, multiple contrast tests, (generalized) linear mixed models, expression analysis

Posted Date: June 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-605289/v1

METHODOLOGY ARTICLE

Change point detection for clustered expression data in short time series.

Miriam Sieg[1*], Lina Katrin Sciesielski[2], Karin Kirschner[3] and Jochen Kruppa[4]

[*]Correspondence:
miriam.sieg@charite.de
[1]Institute of Biometry and Clinical Epidemiology, Charité - University Medicine, Charitéplatz 1, 10117 Berlin, Germany
Full list of author information is available at the end of the article

Abstract

Background: In longitudinal studies, observations are made over time. Hence, the single observations at each time point are dependent, making them a repeated measurement. In this work, we explore a different, counterintuitive setting: At each developmental time point, a lethal observation is performed on the pregnant or nursing mother. Therefore, the single time points are independent. Furthermore, the observation in the offspring at each time point is correlated with each other because each litter consists of several (genetically linked) littermates. In addition, the observed time series is short from a statistical perspective as animal ethics prevent killing more mother mice than absolutely necessary, and murine development is short anyway. We solve these challenges by using multiple contrast tests and visualizing the change point by the use of confidence intervals.

Results: We used linear mixed models to model the variability of the mother. The estimates from the linear mixed model are then used in multiple contrast tests. There are a variety of contrasts and intuitively, we would use the Changepoint method. However, it does not deliver satisfying results. Interestingly, we found two other contrasts, both capable of answering different research questions in change point detection: i) Should a single point with change direction be found, or ii) Should the overall progression be determined? The Sequen contrast answers the first, the McDermott the second. Confidence intervals deliver effect estimates for the strength of the potential change point. Therefore, the scientist can define a biologically relevant limit of change depending on the research question.

Conclusion: We present a solution with effect estimates for short independent time series with observations nested at a given time point. Multiple contrast tests produce confidence intervals, which allow determining the position of change points or to visualize the expression course over time. We suggest to use McDermott's method to determine if there is an overall significant change within the time frame, while Sequen is better in determining specific change points. In addition, we offer a short formula for the estimation of the maximal length of the time series.

Keywords: time series; change point detection; multiple contrast tests; (generalized) linear mixed models; expression analysis

Background

Independent time points are counterintuitive. If we observe samples at different time points, we would assume a dependent data structure. In our work in contrast, time points are defined stages during gestation and later life. Further, the intervention is lethal. In addition, at each developmental stage, littermates and non-littermates are observed. Hence, we have a data setting with independent time points but a

dependent and independent data structure at each time point. The described setting is not common but can be observed in development studies in small mammals. Due to the nature of animal experiments, the number of time points is small from the perspective of a common time series. In our work, the endpoint is the gene expression. An ongoing aim of scientists is a better understanding of the underlying fundamental mechanisms that control organisms development. Scientists have investigated many genes, transcripts, proteins, etc. and their corresponding roles and have introduced models of connecting these networks. Therefore, we want to model the gene expression course to find abrupt changes in clustered expression data, to detect change points.

A high amount of gene expression data analysis is based on snapshots of the transcriptome of many individuals at one time point. Although the expression of genes is to some extent fluctuating, huge differences between individuals are usually not expected. It is, nevertheless, possible that the expression level of certain genes changes considerably during the lifetime of an individual, particularly during specific developmental stages. Changes in expression could be due to maturation of certain organs or at time of birth [1, 2, 3]. The change could be gradual over time or very abrupt. The time point of an abrupt major change in gene expression is called a change point (or breakpoint). The definition of change points, however, is not restricted to time series of expression, but can probably be found in every type of time series data.

Sudden expression changes in such time series can be analyzed by so-called change point analyses. Algorithms for this type of analysis find abrupt major changes in usually slightly fluctuating time series data. The algorithm returns points in the examined time series data at which these changes took place, i.e. the change points. These algorithms are based on Bayesian approaches [4] or binary segmentation approaches [5, 6]. The general literature on methods in change point detection is overwhelming [7]. We refer to Aminikhanghahi and Cook (2016) [8] for a broad overview on methods, definition and application of change point detection in medicine, finance, business, meteorology, entertainment or overall data science. As a standard book for classical change point detection, please be referred to Tartakovsky et al. (2014) [9]. Sturludottir et al. (2017) [10] present short time series in environmental studies. These short time series are modeled assuming autocorrelation. Therefore, the time points are dependent. Menne and Williams (2005) (2005) [11] present detection of change points using multiple test statistics.

Usually, when detecting change points, the time series data consists of hundreds or thousands of time points. A time series consisting of 30 time points is already considered a short time series. The autoregressive integrated moving average (ARIMA) model is widely used for forecasting [12] until today [13]. As a rule of thumb, the minimum number of ARIMA modeling is often given as 30. However, the real number is based on the model parameters to be estimated and the amount of randomness in the given data set. In the case of gene expression across developmental stages, e.g. in mice, the collection time points must be as few as possible but as many as necessary [14]. To assess relevant gene expression changes throughout the lifetime of relatively short-lived organisms like mice, one has to acquire data at specific, predefined time points during all developmental stages like embryonic, fetal, postnatal and adult.

Predefined time points can be based on the day of the conception (embryonic day: E) or the day of birth (postnatal day: P) [15]. Time series in those cases consists of around 12-15 time points, which could lead to problems when applying common change point detected methods.

Additionally, at certain developmental stages and with certain data acquisition techniques, the examination is lethal and an individual can only be tested once. The gene expression information gathered at various time points would therefore originate from different independent individuals. However, when lethal data acquisition is performed, ethical reasons demand examination of all pups in a litter [16]. To reduce the bias from one mother mouse and increase the sample size, pups from at least three mother mice are examined at each time point. The nesting leads to so-called mother effects and therefore dependency inbetween certain data points. As each litter introduces its on variance, this information has to be taken into account when analyzing the data. Ernst and Bar-Joseph (2006)[17] provided an algorithm to cluster short time series gene expression data. However, the clustering combines data from genes with similar time series patterns to reduce the number of tested genes. Linear (mixed effect) regression models with segmentation have already been used for change point detection, but with this method, one does not necessarily find abrupt changes, but rather changes of the slope [18].

Therefore, a change point algorithm is required to analyze very short time series with data consisting of dependent and independent data points. In this work, we applied generalized hypothesis testing by using a linear mixed effect model as a possible change point detection method. We selected three potential contrast matrices for the generalized hypothesis testing. When using a linear regression model, one can decide between effect parameterization and mean parameterization. In case of effect parameterization, one fits a model where the intercept is determined during the fitting process and all β-coefficients are dependent on and compared to the intercept. In case of mean parameterization, the intercept is set to zero and the calculated β-coefficients represent the mean of the corresponding variable. As we wanted to calculate the adjusted mean value for every time point, we decided to use mean parameterization. A linear mixed effect model with mean parameterization allows inclusion of the mix of dependent and independent data, while leaving the focus on the predictor of interest, the developmental time point in our case. Generalized hypothesis testing offers the possibility to include multiple contrast scenarios. To our knowledge, this combination of methods has not been used on data with the goal of change point detection with an interpretable effect estimate. We present three different types of contrast matrices to provide an overview on their applicability for this specific data setting.

Methods

In the following, we present a combination of model fitting and multiple contrast testing for the detection of change points in data which consists of both, independent and dependent data points. However, dependence is not between data points at different but at the same time points. Our observations are nested in each time point. As example, we use development data set. The pups are nested in the mothers. At each time point, there are three new mother animals. Measurement of the

expression levels is lethal for both, mother mice and their offspring. The aim was to find change points in short time series of gene expression data. In more detail, we want to find time points where the expression level of a gene majorly changed compared to the expression levels measured before, incorporating the underlying data characteristics. We tested our method on four sets of short time series from biological gene expression data and eleven simulated data sets. The simulation settings were designed by (basic research) scientists to ensure applicability.

Change point detection with linear mixed models and multiple contrast tests

To determine change points in our specific time series data, we first fit a simple linear mixed effects model with mean parametrization. The expression data for one gene was set as the response. The different measurement time points were set as the fixed effects. The random effects part of the model were the mothers of the mouse pups. Therefore, the litter effect is accounted for and possible overdispersion is reduced. Our simple linear mixed model can be written as:

$$expression_i \sim 0 + \beta_i \times timepoint_i + \gamma_i \times mother_i + \epsilon_i \quad (1)$$

where $expression_i$ is the expression measured for one gene from one pup and $i = 1, \ldots, n$ with n being the number of expression values measured. The coefficients β and γ correspond to the fixed effects covariate *timepoint* and the random effects covariate *mother*, respectively. As a result, the β-coefficients represent the estimated mean values of the respective time points without the random effects variance introduced by the mothers. Using this approach, even more complex models with more confounders would be possible. Here, we concentrate on a simple model. The aim of this work is to demonstrate the general framework. The effects of the time points can be adjusted as in any other multiple linear regression analysis.

We used the `lme4` package [19] in R to fit the linear mixed models using the function `lmer()`. The function `lmer()` uses restricted maximum likelihood estimation by default to fit models that include varying random effects. The functionality determines the variances introduced by the random effects, here the mother effects. With respect to the variances, the rest of the model is fitted and the mean of each time point estimated. In the next step, change points are determined applying generalized linear hypotheses testing. Generalized linear hypotheses testing utilizes contrast matrices and directly performs multiple testing adjustment by applying a multivariate t-distribution. We test different contrast matrices on the data to compare biologically relevant scenarios. In general, other endpoint distributions are possible by modifying the proposed linear regression model. The function `glmer()` allows to fit the full range of the exponential distribution family.

Tables 1, 2, and 3 show different contrast matrices. In the context of our work, the columns in a contrast matrix represent each existing time point and the rows represent the scenarios. The scenarios can be considered as weighted comparisons between the time points. Each cell contains an assigned weight for the corresponding time point at the respective contrast. The sum of the weights equals zero for each row. There are different methods to calculate the respective weights depending on the type of a contrast matrix. In the context of this study, the following three types

of contrast matrices were tested to detect change points: Changepoint, Sequen, and McDermott [20] from the R `multcomp` package [21]. Constructions of the contrast matrices to represent each of these types can be found in the supplementary material 6.

Table 1 Changepoint contrast for five time points and the resulting four contrasts. In C1 the first time point t_1 is compared to the average of the other time points. In C2 the average of $t_1$2 and t_2 is compared to the average of t_3, t_4, and t_5.

	t1	t2	t3	t4	t5
C 1	-1.00	0.25	0.25	0.25	0.25
C 2	-0.50	-0.50	0.33	0.33	0.33
C 3	-0.33	-0.33	-0.33	0.50	0.50
C 4	-0.25	-0.25	-0.25	-0.25	1.00

Table 2 Sequen contrast for five time points and the resulting four contrasts. In C1 the first time point t_1 is compared to the time point t_2. In C2 the timepoint t_2 is compared to t_3 and so on. A zero indicates, that the time point is ignored for this specific contrast.

	t_1	t_2	t_3	t_4	t_5
1-2	-1.00	1.00	0.00	0.00	0.00
2-3	0.00	-1.00	1.00	0.00	0.00
3-4	0.00	0.00	-1.00	1.00	0.00
4-5	0.00	0.00	0.00	-1.00	1.00

Table 3 McDermott contrast for five time points and the resulting four contrasts. In C1 the first time point t_1 is compared to the second time point t_2. In C2 the average of $t_1$2 and t_2 is compared to t_3. In comparison to the Sequen contrast the average of on increasing number of time points is compared to a single time point. Therefore, in the last contrast C5 the average of t_1 to t_4 is compared to t_5.

	t1	t2	t3	t4	t5
C 1	-1.00	1.00	0.00	0.00	0.00
C 2	-0.50	-0.50	1.00	0.00	0.00
C 3	-0.33	-0.33	-0.33	1.00	0.00
C 4	-0.25	-0.25	-0.25	-0.25	1.00

We designed the contrast matrices in our work as follows: Each row of a contrast matrix consists of one possible single change point scenario with respect to the selected construction method. Hence, the contrast matrix represents all possible single change point scenarios for the respective time series and selected method. Table 1 shows an example of the Changepoint contrast. If the Changepoint contrast is selected, the data is first divided into two groups for each row of a contrast matrix. One group contains the time points before the potential change point, the other group the time points at and after the potential change point. Then, the relative weight for each time point with respect to its group is calculated. Basically, the sample sizes from all time points of a group are summed and the sample size of each time point is divided by the respective sum. The sum of the weights from each group therefore adds up to one and the sum of the weights of both groups equals zero. The weights belonging to the time points before and at the possible change point are negated. If the Sequen contrast method is selected, only the time point directly before and at the possible change point are considered. All other time points are set to 0. The time point directly before the possible change point is set to -1 and the possible change point is set to 1. Table 2 shows an numerical example.

Lastly, the McDermott contrast is a mixture between the Changepoint and the Sequen contrasts. Table 3 presents a numeric example. The weights of the time points of the time series before the possible change point are calculated the same

way as for the Changepoint contrast. The sample sizes of each time point in this part of the time series are divided by summed sample sizes of this group. The possible change point itself is set to 1 and the rest of the time series is set to 0. The McDermott contrast matrix was originally invented for ordered means. A significant contrast in our setting would therefore suggest an overall significant change of the time series, especially since our means are not ordered. In summary, Changepoint considers all data points in the time series, Sequen considers data points at and just preceding the potential change point, and McDermott only the data points at the time points before and at each potential change point.

Taken together, we fitted linear mixed effect models for different biologically relevant time courses and for each of the four in vivo expression short time series. To each fitted model, we applied three varying generalized hypotheses testing contrasts. The contrasts returned effect estimates for each scenario and respective 95% confidence intervals. The contrasts were evaluated on the basis of whether the respective contrast could be used to determine change points and whether it would potentially return the positions and directions of change points.

Maximal number of usable time points

The presented approach has a theoretical limitation in the number of significant time point differences. If many time points are included, we will correct the comparison of each time point for the type I error. Therefore, at a given number of time points depending on the maximal observed effect size δ_{max} and the corresponding standard deviation s, no significant change point will be detected. In addition, the approximation also depends on the chosen contrast matrix. In the following, we will examine an approximation of how many time points can be analyzed. The scientist must estimate a δ_{max} and the corresponding s from the literature or the observed data. Then, we can calculate the Z score:

$$z = \frac{\delta_{max}}{s} \tag{2}$$

The absolute value of the Z-score can be used by the probability density function of the normal distribution to calculate a p-value. In R, this can be achieved by the function `pnorm()`, which returns the integral from $-\infty$ to z of the probability density function of the normal distribution. We multiply the result by two to account for a two-sided test resulting in the p_{max}. We simplify by assuming a Bonferroni adjustment. Dividing 0.05 by p_{max} will determine the maximal number of theoretically possible detectable change points. The emphasis is on theoretical, because if we are not able to find any significant p-value, we will also not find any significant confidence intervals. Therefore, all confidence intervals will be widely scattered and therefore include zero. We demonstrate here only an approximation, see the discussion section for further considerations. A small numeric example is given from Figure 3.a) which shows a δ_{max} of 3 between the two plateaus. If we assume a standard deviation of 1, we can calculate a z of $\frac{3}{1}$ equal 3. Using the function `pnorm(-3)` we get a p-value of 0.00135. Hence, we are able to use approximately 37 time points in our analysis. More time points will make each confidence interval wider to a point, where differences will no longer be detectable. Should significance be the only criterion, further research will be required to circumvent these limitations.

Biological expression data

We present a biological data set as motivation example. The expression data set is an extraction of a so-far unpublished study. We used the biological data as received (full course, not cleaned) and demonstrate our proposed method. It is on the researcher to decide which developmental stages should be included depending on the research question. In detail, our example data consists of two mouse organs. We analyzed mouse livers and kidneys from different developmental stages (embryonic to adult) for *glucose transporter 1 (Glut1)* and *carbonic anhydrase 9 (Car9)* expression by probe-based qPCR against a standard curve. The expression levels are displayed as *Glut1* or *Car9* molecules per $10^6 \beta - Actin(Actb)$ molecules. We used log-transformed expression values for our analysis to meet assumptions of the linear mixed model.

The data structure was the same for all four data sets. Each data set consisted of gene expression data from multiple samples determined at 12 fixed time points plus the adult stage. The time points represent different developmental stages. The gene expression information was constrained to one gene measured in one organ per time series. Expression of a specific gene in a specific organ was measured in multiple mouse pups by multiple mothers from twelve days after coitus (E12, Theiler Stage TS20) onwards. The 12 fixed time points contained two embryonic, four fetal, six postnatal and the adult stage(s). No pup was included twice and each mother only had one litter, i.e. at each time point, the litters originated from different mothers. The variance introduced from a varying litter is called the litter effect. Not including this information in the final model could lead to overdispersion [22]. From a statistical point of view, this means expression data gained from the same litter was dependent, but was independent between the litters. Hence, at each time point the data consisted of both dependent and independent data points. Additionally, expression information between different time points was independent.

Finally, the four data sets were chosen because both genes showed a stable basal expression and a change of expression in only one of the organs, so that the other organ could serve as a "no change point" control. Expression changes from high-to-low (liver *Glut1*) and low-to-high (kidney *Car9*) were used to visualize our approach.

Artificial expression data

The physicians in the study defined four hypothetical time series of gene expression data, representing biologically realistic and interesting scenarios. We simulated data with respect to the described data structure shown in figure 1. In detail, theoretical curves of the mean of the measured expression values for the respective time points in a time series were acquired. On the theoretical courses, we were able to determine the properties of the different contrast tests. In total, four overall relevant courses of the means of the gene expression in the time series were defined and are as follows: a) no change, b) steady change, c) stepwise change and d) partly dropped. In addition, we also simulated both directions (increase and decrease), if possible, simulating a linear increase as well as a linear decrease and so on.

We would not expect the detection of change points in the time series representing scenarios a) and b). Therefore, both scenarios are our control or null models. However, for scenarios c) and d), we would expect detection of at least one change point. In addition, the confidence intervals should also provide more details on our

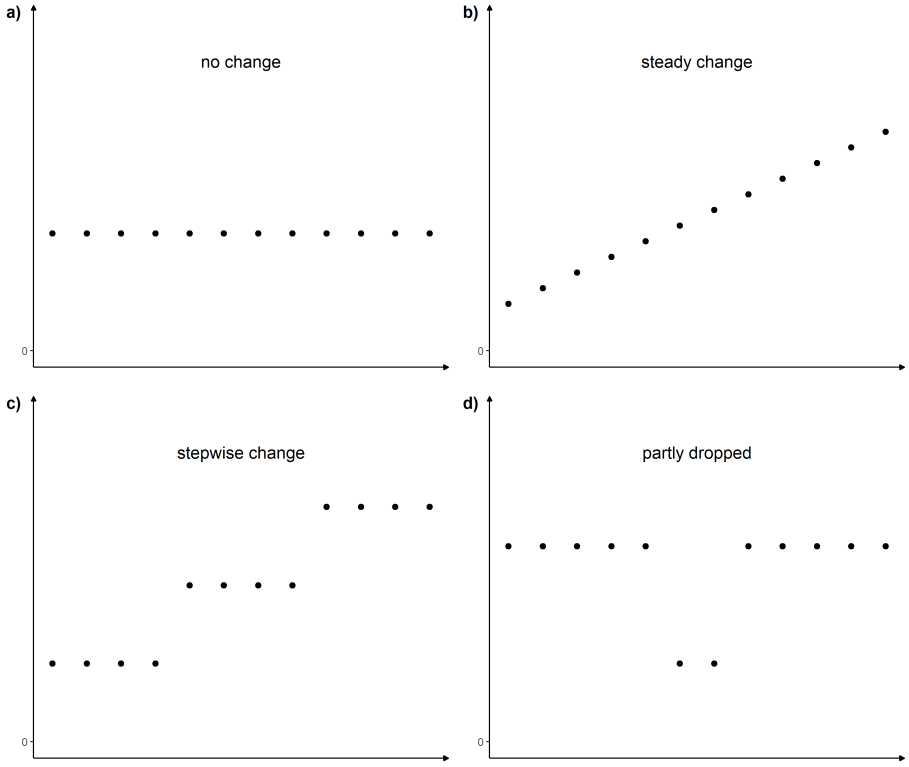

Figure 1 Possible courses of included short time series. Each subplot a) to d) represents one group of scenarios of courses within a time series. Time points are on the x-axis outcome on the y-axis. Scenario a) shows a time series with no change. Steady changes b) and stepwise changes c) each include increase and decrease of outcome values within the time series. Scenario d) represents a partly dropped course which readjusts to previous outcomes after a while. The values may drop down to zero. These hypothetical time courses were provided. For scenarios a) and b), one would not expect any change points. In contrast, one would predict finding change points for scenarios c) and d).

findings. For each of the defined time series, gene expression data for 12 distinct time points were simulated. As our biological example data had 13 developmental stages, we removed the adult stage to generate congruent data sets. The number has also good properties for the generation of the time points. For simulation of the expression data, we used the statistical programming language R 3.6 and the R package `simstudy` [23]. For each time point, we first generated three data points sampled from a normal distribution with a mean of zero and a variance of 5, the mother effects. These simulated mother values represented the individual effects each of the selected mother mice introduced on their respective litters. We do expect some mother effect, but no drastic differences at the same time point. We have chosen a high mother variance, to achieve a more drastic setting. A very low variance would generate very distinct expression values. We do not believe that this is a very realistic setting. The amount of pups per litter were sampled from a Zero-truncated Poisson distribution with a lambda of 10. Therefore, each mother has an average of roughly 10 pups. The expression values of the mouse pups from the different litters were then generated by sampling from a normal distribution. The mean was based on the respective intercept and sampled mother effect. The vari-

ance was set to 2 since we expected only small differences between the expression values of the pups. In consequence, we had simulated expression values for pups from three different mothers for each of the 12 time points per defined course. For the more programming-oriented reader, we present the R code on a GitHub repository (https://github.com/msieg08/clustered_data_changepoint_detection) and code chunks in the supplementary material section 6.

We did not run different simulations with different sample sizes because the properties of the estimates from a linear mixed model in multiple contrast test is already well known. A general tutorial on linear mixed models using contrasts in R and the theoretical background can be found in Schad et al. (2020) [24]. Also Bretz et al. (2011) [25] and Hothorn et al. (2008) [21] deliver the theoretical background. Linear mixed models used in multiple contrast test will deliver unbiased estimates and will produce simultaneous confidence intervals on a 95% significance level. The properties are checked for heterogeneity [26], complex data models [27], and even under overdispersion and small sample sizes [22]. Therefore, we consider the use of linear mixed models a valid and unbiased way to determine the estimates for the multiple contrast testing.

Results

The following section is divided into two parts. First, we present four biological motivation data examples, two of which can be found in the supplementary material. The mouse development data set underlines the biological necessity of our approach. Second, we simulate different course settings inspired by the biological data. We show the resulting confidence interval plots for each simulation and contrast and separately report the effect estimates.

In all presented plots, subplot a) shows the respective data with time points on the x-axis and the observed expression values on the y-axis. We assume here to have at least a log-normal distributed outcome. Each dot in the plot represents one observed value. The colors represent the data dependencies, meaning that dots with the same color belong to the same cluster, e.g. pups from the same mother. Subplots b) to d) show the estimated mean difference including the 95%-confidence interval (x-axis) for each respective change point scenario (y-axis).

Biological gene expression data

Biological data of the *Glut1* gene expression in the kidney and liver are presented in figure 2 and figure 3, respectively. Supplementary figure 1 shows the biological data of the *Car9* expression in the kidney. The estimation of the model parameters shown in supplementary figure 1 caused converting problems. We observed singular fits. Supplementary figure 2 presents the *Car9* expression data from liver. Table 4 shows the numerical effect estimates for the *Glut1* data from kidney and table 5 the *Glut1* values from liver. Both plots have the same structure and consist of the same subplots. The subplot a) shows the biological data separated into three developmental stages. Each dot represents a single pup nested into a single mother which is indicated by the same (litter) color. Please note that the expression data is log-transformed. The other subplots show the results of the different contrast tests: b) Changepoint, c) Sequen, and d) McDermott. The scattered line indicates the

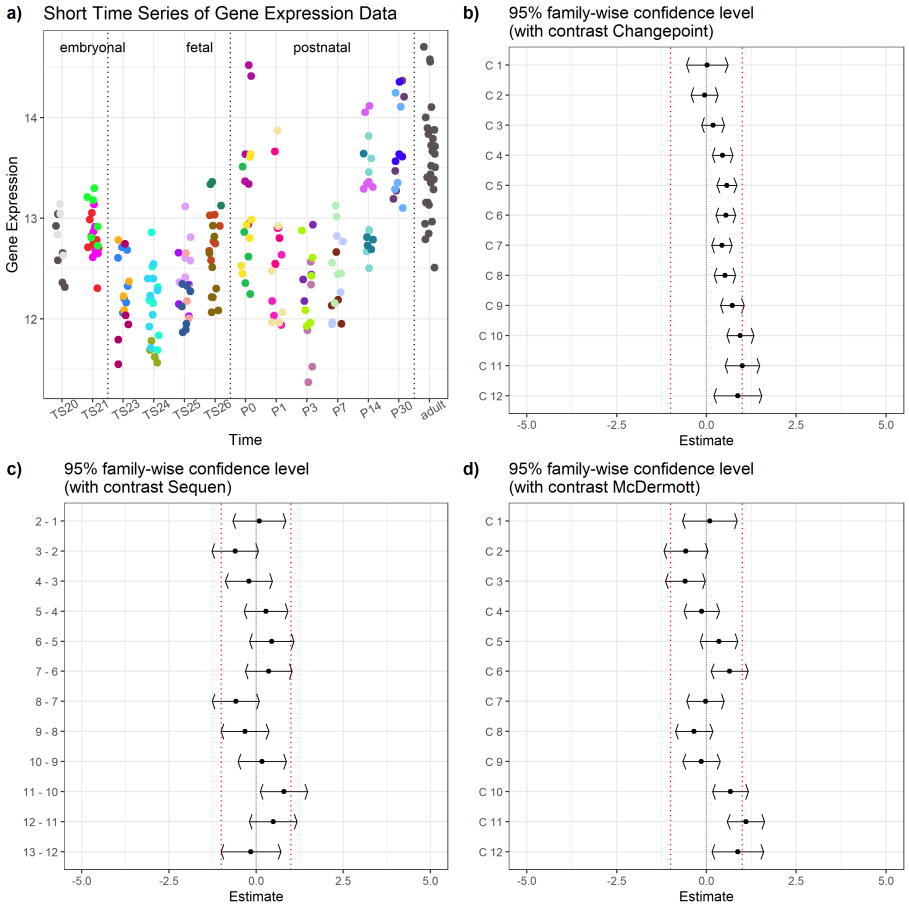

Figure 2 Biological example data for Glut1 expression in the developing kidney. Subplot a) shows the biological example data set (log-transformed). Each time point on the x-axis represents an independent developmental stage. Each data point represents a pup and each color a mother animal. The pups are nested into the mothers. We added three broader development stages (embryonal, fetal, postnatal) for easier reference. The subplots show the confidence intervals of the Changepoint contrast (b), Sequen contrast (c), and McDermott contrast (d). The red scattered line indicates the chosen limits of biological relevance.

biological relevance limits. The limits are user-specific and depend on the research question. We decided to choose ± 1.

In figure 2, no obvious change point or expression shift can be detected. The overall tendency of the point estimates is increasing. The corresponding numeric values are presented in table 4. All three contrast tests deliver significant confidence intervals below the biological relevance threshold. In comparison to the Changepoint contrast, however, the McDermott contrast is able to visualize the plateau of the last three developmental time points with effect estimates of 0.67, 1.10, and 0.87. Biologically, the points where the sign of the effect changes, i.e. the confidence interval jumps from -x to +x, are also interesting. We observe this change of sign at sequences (e.g. 7-6 to 8-7) as well as at McDermott (e.g. C9 to C10).

Figure 3 shows an example of a visually obvious change point with severe expression changes after birth (from P0). The change point is indicated by a gray line in table 5. The Changepoint contrast visualizes the overall course of the time

Table 4 Contrasts and estimates of figure 2. The table shows the numeric values form the *Glut1* example data from liver. The C column indicates the contrast, the Δ the log mean change of the corresponding contrast C. The gray row indicates a possible change point by visual inspection of figure 2. A significant confidence interval does not include zero.

Changepoint				Sequen				McDermott			
		95% CI				95% CI				95% CI	
C†	Δ‡	Low	Upp	C†	Δ‡	Low	Upp	C†	Δ‡	Low	Upp
C 1	0.02	-0.56	0.60	2 - 1	0.09	-0.66	0.85	C 1	0.09	-0.67	0.85
C 2	-0.05	-0.43	0.32	3 - 2	-0.60	-1.27	0.07	C 2	-0.57	-1.19	0.04
C 3	0.18	-0.14	0.50	4 - 3	-0.21	-0.88	0.47	C 3	-0.59	-1.15	-0.03
C 4	0.45	0.16	0.74	5 - 4	0.28	-0.35	0.91	C 4	-0.13	-0.63	0.36
C 5	0.57	0.29	0.84	6 - 5	0.45	-0.19	1.09	C 5	0.35	-0.18	0.88
C 6	0.54	0.26	0.82	7 - 6	0.36	-0.31	1.03	C 6	0.65	0.13	1.16
C 7	0.43	0.14	0.71	8 - 7	-0.58	-1.26	0.09	C 7	-0.03	-0.55	0.49
C 8	0.51	0.21	0.82	9 - 8	-0.32	-1.01	0.37	C 8	-0.35	-0.87	0.18
C 9	0.72	0.38	1.05	10 - 9	0.17	-0.52	0.86	C 9	-0.14	-0.66	0.38
C 10	0.94	0.56	1.32	11 - 10	0.80	0.12	1.48	C 10	0.67	0.17	1.17
C 11	1.00	0.51	1.48	12 - 11	0.49	-0.19	1.17	C 11	1.10	0.58	1.62
C 12	0.87	0.20	1.54	13 - 12	-0.15	-1.01	0.71	C 12	0.87	0.15	1.59

† Given contrast. See Eq.2 for Sequen, Eq.1 for Changepoint, and Eq. 3.
‡ Point estimator of the confidence interval i.e. mean difference given the contrast.

points more than the rapid decrease from TS26 to P3 and it does not deliver a clear interpretable position of the change. The averaging over all time points concealed the linear increase between the TS17 and TS21 developmental stages because the decrease at the end of the time points is too severe. In contrast, the Sequen contrast detects the change point at the 9-8 position (P0-P1) with an effect of -1.82 [-3.03; -0.61]. Due to the mixed modeling, we were able to account for the high variance of developmental stage P1. However, no confidence interval falls below the lower relevance limit. The McDermott contrast shows confidence intervals below the relevance limit with an effect of -2.42 [-3.34; -1.50] at birth. In the following, the confidence intervals have a point estimate around -3.2. In addition, the slight increase in the beginning is also pictured in the course of the confidence intervals with an effect around 1.

Supplementary figure 1 shows the biological data of the *Car9* gene from kidney. The numerical values can be found in supplementary table 1. The estimation of the model parameters caused converting problems. We achieve singular fits, therefore estimated variance-covariance matrices with less than full rank. The warning indicates that one or more variances are very close to zero. Therefore, a careful consideration of the results is required. We are sure to avoid the fitting of overly complex models [28] and assure consistency of the model with the experimental design [29]. Therefore, we believe that the mean estimates and the variance /covariance matrices are valid, even if mixed models can show converting problems. The biological data shows a plateau from TS20 to P7 with a high increase of the expression at P14. The Changepoint contrast again delivers a biased visualization. The change point might be recognized, but the overall trend is flawed. Therefore, the Changepoint contrast cannot be recommended. The Sequen contrast detects the change point significantly and above the relevance limit. The lower limit of the confidence interval exceeds the upper relevance limit with 2.15 [1.64; 2.66]. Finally, the McDermott contrast visualizes the plateau in conjunction with the rise of expression with an point estimate of 2.01 [1.64; 2.37]. The last three confidence intervals are all above the relevance limit with an effect of 2.01, 2.93, and 2.63. In

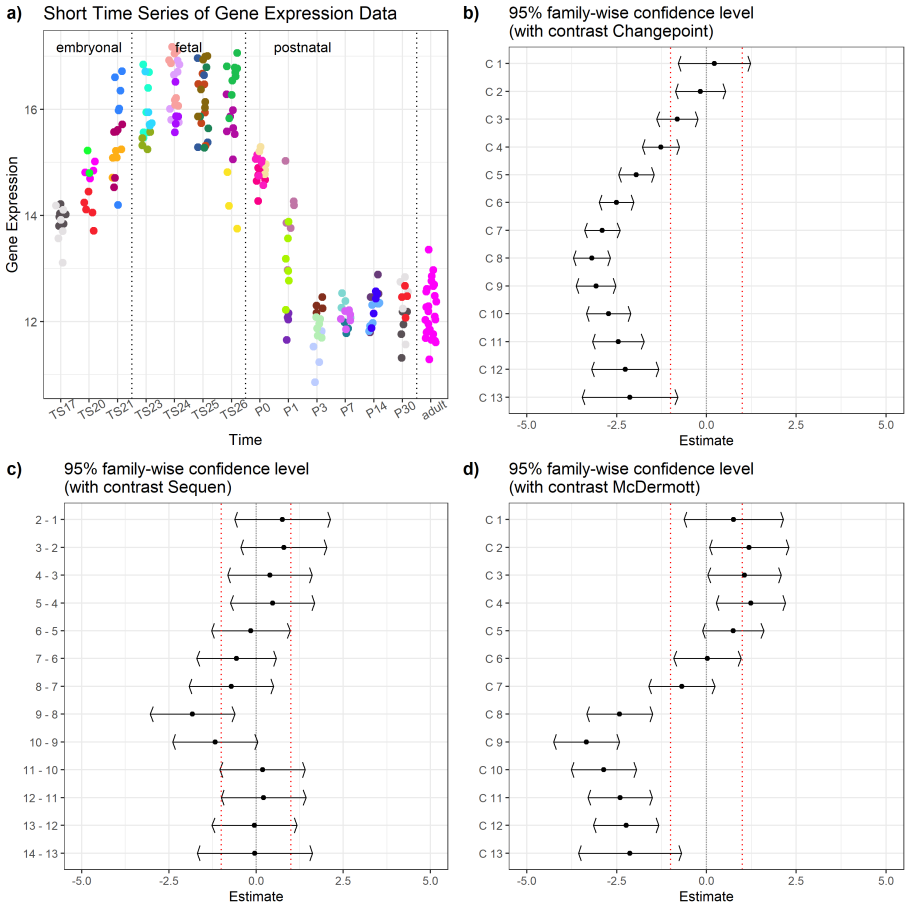

Figure 3 Biological example data for Glut1 expression in the developing liver. Subplot a) shows the biological example data set (log-transformed). Each time point on the x-axis represents an independent developmental stage. Each data point represents a pup and each color a mother animal. The pups are nested into the mothers. We added three broader development stages (embryonal, fetal, postnatal) for easier reference. The subplots show the confidence intervals of the Changepoint contrast (b), Sequen contrast (c), and McDermott contrast (d). The red scattered line indicates the chosen limits of biological relevance

the last biological example (Car9 expression in the developing kidney), there is no obvious expression change apart from some higher variance in P14 expression. All three contrasts stay inbetween the relevance limits (Supplementary Table 2). This example illustrates that both, biological visualisation and confidence intervals, are required.

Simulation data

We simulated eleven simulation settings according to figure 1. We fitted one linear mixed effect model on each of the simulated times series. These fitted models were then used for generalized linear hypothesis testing with three different contrast matrices. The results of interest were the mean difference and associated 95% confidence intervals. Depending on the used contrast matrix, the output suggested the presence or absence of change points. We present here two out of the eleven simulated settings. Please be referred to the supplementary material for all simu-

Table 5 Contrasts and estimates of figure 3. The table shows the numeric values for the *Glut1* example data from liver. The C column indicates the contrast, the Δ the log mean change of the corresponding contrast C. The gray row indicates a possible change point by visual inspection of figure 3. A significant confidence interval does not include zero.

Changepoint				Sequen				McDermott			
		95% CI				95% CI				95% CI	
C†	Δ‡	Low	Upp	C†	Δ‡	Low	Upp	C†	Δ‡	Low	Upp
C 1	0.22	-0.79	1.23	2 - 1	0.76	-0.62	2.13	C 1	0.76	-0.63	2.14
C 2	-0.17	-0.87	0.53	3 - 2	0.79	-0.44	2.03	C 2	1.19	0.08	2.29
C 3	-0.81	-1.39	-0.23	4 - 3	0.39	-0.82	1.61	C 3	1.06	0.04	2.08
C 4	-1.27	-1.79	-0.75	5 - 4	0.47	-0.73	1.68	C 4	1.24	0.27	2.20
C 5	-1.95	-2.45	-1.46	6 - 5	-0.15	-1.28	0.98	C 5	0.74	-0.11	1.60
C 6	-2.51	-2.99	-2.02	7 - 6	-0.56	-1.70	0.58	C 6	0.03	-0.91	0.96
C 7	-2.90	-3.40	-2.41	8 - 7	-0.71	-1.92	0.50	C 7	-0.69	-1.61	0.23
C 8	-3.19	-3.71	-2.67	9 - 8	-1.82	-3.03	-0.61	C 8	-2.42	-3.34	-1.50
C 9	-3.08	-3.63	-2.52	10 - 9	-1.17	-2.39	0.04	C 9	-3.34	-4.26	-2.42
C 10	-2.73	-3.34	-2.11	11 - 10	0.18	-1.04	1.41	C 10	-2.86	-3.77	-1.95
C 11	-2.46	-3.18	-1.74	12 - 11	0.22	-1.00	1.43	C 11	-2.41	-3.31	-1.51
C 12	-2.26	-3.20	-1.33	13 - 12	-0.05	-1.26	1.17	C 12	-2.24	-3.15	-1.33
C 13	-2.13	-3.46	-0.80	14 - 13	-0.04	-1.69	1.61	C 13	-2.13	-3.57	-0.70

† Given contrast. See Eq.2 for Sequen, Eq.1 for Changepoint, and Eq. 3.
‡ Point estimator of the confidence interval i.e. mean difference given the contrast.

lation results. Figures 4 and 5 show the course in figure 1 c) and d). Table 6 and 7 present the numeric values. We indicated the simulated change point by a gray row. In particular, the number of simulations was increased by the fact that when expression increased, we modeled the decrease separately.

Figures 4 shows a stepwise increase of expression, table 6 the corresponding numeric values. We observe two distinct change points. For demonstration purposes, we simulated the variance in such a way that a slight overlap of the observations occurred. The simulated effect was 10. Therefore, each rise/expression change increased the average expression by 10, resulting in the gene expression course shown in subplot a). In contrast to our assumption, the Changepoint contrast does not detect a change point by looking at subplot b) and the confidence intervals in table 6. Hence, the name of the contrast is misleading - as is the position of all significant confidence intervals. The Sequen contrast delivers the change points correctly at contrasts 5-4 and 9-8. We were able to detect the change by the significant confidence intervals or visually by exceeding of the intervals. The direction of the change is also represented correctly. In addition, there is a slightly lower effect of 7.77 [3.36; 12.17] at the second compared to the first change point with 10.63 [6.22; 15.05] as in the visualization in subplot a). Hence, the Sequen contrast delivers the correct direction in conjunction with the correct effect estimates. Finally, the McDermott contrast mimics the steps of the simulated data. Each rise at C4 and C8 can be observed by a stronger shift of the confidence intervals to the right with an effect of 9.53 [6.02; 13.05] and 13.05 [9.70; 16.40], respectively. Hence, position and the direction of the change point are both correct. The confidence interval itself is not on the same level because the single time points have slightly different means. These findings are also true for two positive change points shown in supplementary figure 5 as well as four positive change points presented in supplementary figure 7. The decreasing setting is presented in supplementary figure 9 for two change points, in supplementary figure 10 for three change points, and in supplementary figure 11 for four change points. The findings for the decreasing setting are the same as for the positive one. In summary, the Sequen and McDermott contrasts are able to detect

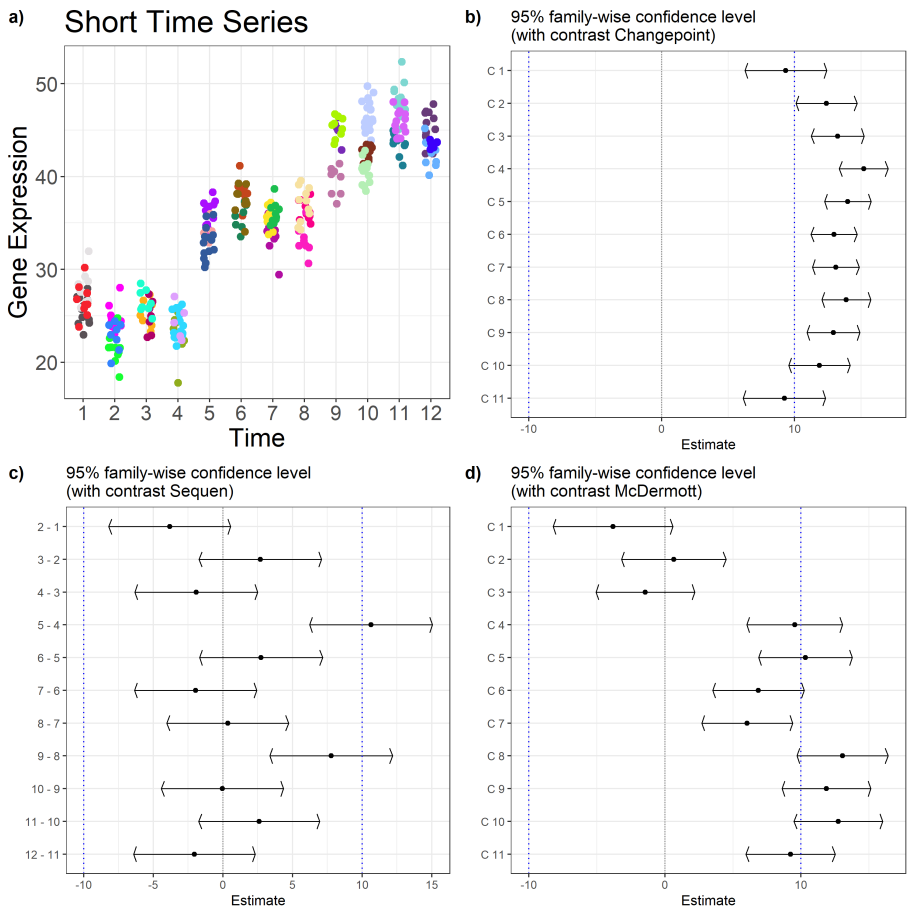

Figure 4 Confidence intervals of estimates from linear mixed model coupled with contrast matrix for short time series data with three change points. Figure a) shows the longitudinal time points (x-axis) of the sampled short time series in association with gene expression (y-axis) with two expected change points. Each color is related to one mother mouse. Subfigures b), c) and d) show the estimates (x-axis) including confidence intervals for the observed contrasts (y-axis) with methods Changepoint, Sequen and McDermott, respectively. The blue line indicates the simulated effect.

the position and direction (Sequen) or the overall course (McDermott) of predefined change points.

Figure 5 presents a "partly dropped" change point. The corresponding numeric values are shown in table 7. The expression is reduced at two time points before it is restored to the original values. In supplementary figure 12 we show a total expression shot down with an expression of zero over four time points. In figure 5 c), the Changepoint contrast confidence intervals are shown. In contrast to figure 4, the Changepoint contrast does deliver a change in the confidence interval plot. However, the indicated change of 2.16 [0.40; 3.92] at C7 does not mimic the simulated data. Again, the Changepoint contrast does not help to indicate the correct position or effect directions as it indicates a positive change instead of a negative one (decreased expression). The Sequen contrast indicates both change points at the correct position. The 6-5 and 8-7 contrasts are significant with an effect of -7.26 [-11.67; -2.85] and 10.34 [5.98; 14.71]. The direction is also correct. The first signif-

Table 6 Contrasts and estimates to figure 4. The table shows the numeric values form the simulation for three change points. The C column indicates the contrast, the Δ the log mean change of the corresponding contrast C. The gray row indicates the predefined change point(s). A significant confidence interval does not include zero.

Changepoint				Sequen				McDermott			
		95% CI				95% CI				95% CI	
C†	Δ‡	Low	Upp	C†	Δ‡	Low	Upp	C†	Δ‡	Low	Upp
C 1	9.34	6.25	12.42	2-1	-3.82	-8.19	0.55	C 1	-3.82	-8.23	0.58
C 2	12.40	10.10	14.70	3-2	2.68	-1.71	7.08	C 2	0.65	-3.19	4.49
C 3	13.25	11.26	15.25	4-3	-1.91	-6.32	2.50	C 3	-1.44	-5.07	2.19
C 4	15.20	13.37	17.04	5-4	10.63	6.22	15.05	C 4	9.53	6.02	13.05
C 5	14.01	12.25	15.77	6-5	2.74	-1.67	7.15	C 5	10.33	6.89	13.77
C 6	12.97	11.24	14.70	7-6	-1.96	-6.34	2.42	C 6	6.87	3.51	10.24
C 7	13.10	11.34	14.86	8-7	0.34	-4.03	4.71	C 7	6.05	2.71	9.39
C 8	13.90	12.05	15.74	9-8	7.77	3.36	12.17	C 8	13.05	9.70	16.40
C 9	12.92	10.93	14.91	10-9	-0.03	-4.42	4.35	C 9	11.87	8.61	15.14
C 10	11.87	9.56	14.18	11-10	2.61	-1.73	6.95	C 10	12.74	9.49	16.00
C 11	9.23	6.13	12.34	12-11	-2.05	-6.42	2.32	C 11	9.23	5.96	12.51

† Given contrast. ‡ Point estimator of the confidence interval i.e. mean difference given the contrast.

icant confidence interval has a negative effect, indicating the drop and the second significant confidence interval has a positive effect indicating the rise in expression. In comparison to the Sequen contrast, the McDermott contrast must be interpreted differently. Again, the two significant confidence intervals are indicating the area of change with two significant confidence intervals at C5 and C6 with an effect of -7.63 [-11.07; -4.19] and -8.49 [-11.85; -5.13]. However, the direction of the change must be calculated by the researcher. The McDermott contrast rather visualizes the course than giving the concrete direction of the decrease/increase. Depending on the research question, Sequen or McDermott might be preferred. Supplementary figure 12 shows the extreme event of four time points with no expression and therefore no variance at those. In this extreme scenario, all three contrasts deliver confidence intervals. Again, the Changepoint contrast pictures highly misleading directions and effects. We observe a lower plateau with a linear increase to another plateau. This does not emulate the course of the expression data at all. The Sequen contrast correctly delivers the change point positions and directions at 5-4 and 9-8 with the effects of -8.76 [-12.62; -4.90] and 8.28 [4.43; 12.14]. The McDermott contrast has more biased confidence intervals. The drop is visualized by the contrast but the last confidence intervals falsely indicate a higher plateau of expression than at the beginning of the time course. In addition, the significant confidence intervals indicating the drop also show a false steady decrease of the effect. Please see supplementary table 12 for the numeric values of the confidence intervals.

Finally, we simulated no change, linear increase, and linear decrease. Supplementary figure 3 shows the results of the no change simulation. All contrasts did not detect any change points, presenting non-significant, overlapping confidence intervals. The supplementary figures 4 and 8 show a linear increase and a linear decrease, respectively. The overall tendencies of the confidence intervals are the same in both settings. Supplementary figure 4 is a mirror of supplementary figure 8. The Changepoint contrast is significant for all confidence intervals with a strong effect. The point estimates are the same for nearly all confidence intervals. The Sequen contrast has some slightly significant confidence intervals. However, all confidence intervals overlap, indicating no change in expression. The McDermott contrast mim-

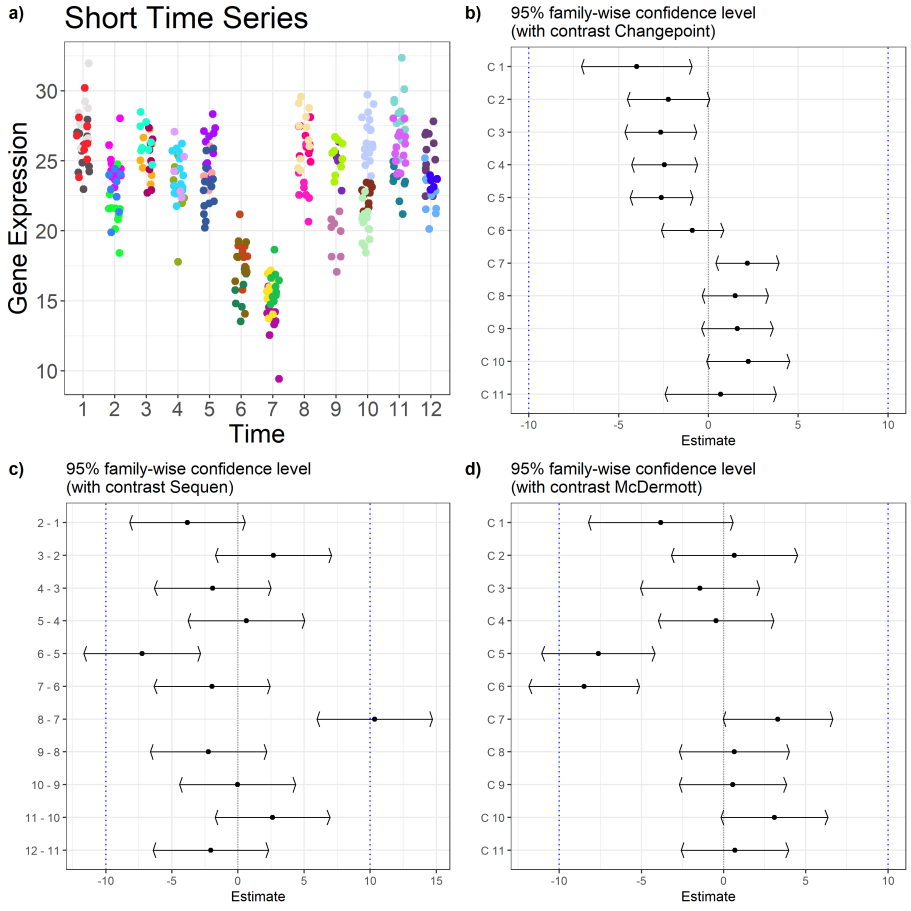

Figure 5 Confidence intervals of estimates from linear mixed model coupled with contrast matrix for short time series data with a "partly dropped" change points. Figure a) shows the increasing time points (x-axis) of the sampled short time series in association with gene expression (y-axis) with two expected change points. Each color is related to one mother mouse. Subfigures b), c) and d) show the estimates (x-axis) including confidence intervals for the observed contrasts (y-axis) with methods Changepoint, Sequen and McDermott, respectively. The blue line indicates the simulated effect.

ics the linear tendency of the expression data with its positive and negative trends. As all confidence intervals overlap, we conclude that no change point is present.

A word of caution about the estimated effects and the direction of the effect. Our approach allows determining the point estimate of the difference between time points. Depending on the contrast, different effects will be reported. The preferred contrast is therefore highly dependent on the research question. While the Sequen contrast provides the point of change, the McDermott contrast visualizes the overall course of the change. However, we cannot recommend the original Changepoint contrast for detection or assessment of the change point as its effect estimates are biased.

In summary, if Sequen or McDermott were applied as contrast matrices and an actual change point was present in the simulated data, the confidence interval from the respective contrast was significant and no (or only a small) overlap with the confidence interval of the preceding contrast occurred. When there was no change

Table 7 Contrasts and estimates to figure 5. The table shows the numeric values form the simulation for a "partly dropped" change point. The C column indicates the contrast, the Δ the log mean change of the corresponding contrast C. The gray row indicates the predefined change point(s). A significant confidence interval does not include zero.

Changepoint				Sequen				McDermott			
		95% CI				95% CI				95% CI	
C†	Δ‡	Low	Upp	C†	Δ‡	Low	Upp	C†	Δ‡	Low	Upp
C 1	-4.00	-7.08	-0.92	2-1	-3.82	-8.19	0.55	C 1	-3.82	-8.23	0.58
C 2	-2.23	-4.52	0.07	3-2	2.68	-1.71	7.07	C 2	0.65	-3.19	4.49
C 3	-2.66	-4.65	-0.67	4-3	-1.91	-6.32	2.50	C 3	-1.44	-5.07	2.19
C 4	-2.45	-4.28	-0.62	5-4	0.63	-3.78	5.05	C 4	-0.47	-3.98	3.05
C 5	-2.62	-4.37	-0.86	6-5	-7.26	-11.67	-2.85	C 5	-7.63	-11.07	-4.19
C 6	-0.89	-2.62	0.84	7-6	-1.96	-6.34	2.42	C 6	-8.49	-11.85	-5.13
C 7	2.16	0.40	3.92	8-7	10.34	5.98	14.71	C 7	3.30	-0.04	6.64
C 8	1.49	-0.35	3.34	9-8	-2.23	-6.64	2.17	C 8	0.65	-2.70	4.00
C 9	1.61	-0.38	3.60	10-9	-0.03	-4.41	4.35	C 9	0.56	-2.71	3.82
C 10	2.21	-0.10	4.52	11-10	2.61	-1.73	6.95	C 10	3.08	-0.17	6.34
C 11	0.68	-2.42	3.78	12-11	-2.05	-6.42	2.32	C 11	0.68	-2.60	3.95

† Given contrast. ‡ Point estimator of the confidence interval i.e. mean difference given the contrast.

point, the 95% confidence intervals for each contrast were either not significant or they overlapped with the confidence interval of the preceding contrast. The respective patterns can be observed in a more or less defined way on all simulated data from the Sequen and McDermott contrasts. The Changepoint contrast cannot be recommended for the detection of a change point in any simulation setting. Overall, we suggest using McDermott's method to determine if there is a significant change within the time frame, while Sequen could be applied to determine the specific change point(s) and their direction.

Discussion

In a classical longitudinal design, each patient is observed at each inter-dependent time point. In this study, we examine a different non-intuitive setting: The time points are independent as the intervention on the pregnant mice is lethal and the observations, gene expression in the litter organs, at each time point are correlated, resulting in a mixture of dependent and independent data structures at one time point. In addition, from a statistical point of view, the time series is short. We solve the research question looking for change points in this experimental setting by using multiple contrast tests and by visualizing the change point with simultaneous confidence intervals. We have investigated three contrasts which differ in the research questions they can answer: Should a single change point be found, or should the overall course rather be pictured? The Sequen contrast answers the first, the McDermott the second. The Changepoint contrast gives a clearly biased visualization and is unable to correctly determine change points in our setting. To summarize, we used generalized hypothesis testing with linear mixed effect models using various contrast matrices to detect change points in short time series of gene expression levels with independent and dependent data points.

Usually, a short time series still has around 30 time points. This amount is an established rule of thumb, which is also often generally criticized. However, in our case, the number of time points is limited to at most 13 and cannot be increased dramatically for animal-ethical reasons. First, due to the three R principle [30, 14], any animal experiment will only have a relatively small, limited number of

observations. Second, we need some distinct distance between the developmental stages to detect a rise or a fall. More traditional change point algorithms were not designed to find change points in very short time series as they assumed fluctuations over certain time frames. They were also not implemented for our specific data structure. Our approach presented here is able to visualize complex, inter-dependent data structures for further decision making, closing the gap for statistical assessment of such short time series data.

A connected question is how long such a time line can be to still be able to detect differences. As generalized hypothesis testing is applied, it automatically adjusts locally for multiple testing. Therefore, for each model, the respective significance level is met. The number of time points minus one comparison was evaluated for all the contrasting methods we chose. The higher the number of time points, the more contrasts are tested, leading to a stricter change point selection but also higher run times. In our method section, we only give an approximation of the theoretical maximal length of the time series because the main aim of our work was to identify the most informative contrast test for detecting a given change point pattern. We found the Sequen and McDermott contrasts which both are not intuitively the first choice. Furthermore, our approximation is based on the Bonferroni adjustment. This is not the correct one for multiple contrast tests. In future work, the borders of the number of maximal time points and multiplicity adjustment approaches [31, 32] will be examined in more detail.

We have discussed the possible length of time series under the consideration of significance. Hence, if a confidence interval is significant, we would assume a change point. In the biological example data, however, we were also able to define a relevance threshold moving from (just) significant to biologically relevant in our decision making. In our view, the significance is not as important as the relevance [33]. Therefore, the focus on the point estimator and the overall course of the confidence interval is more important. The shift to informative effect estimates is therefore required to make sure that findings can be reproduced on the way from basic research to clinical trials [34, 35]. Our approach allows estimating the effect of the change point. In our work we used a log-normal transformation of the expression values. Depending on the outcome, the linear mixed models also have a generalized implementation to model the full range of the exponential family [22].

Many multiple contrast tests are well described in the literature as well as the application in statistical inference [25]. The most common contrast might be the all-pairs contrast (also known as the Tukey contrast), or the many-to-one contrast (also known as the Dunnett contrast). Other types of contrasts are not so widespread and known. Interestingly, the so-called Changepoint contrast does not deliver any change point in the context of our experimental design. We do not criticize its general approach but for our data, it does not deliver the best interpretable change point(s) in the context of confidence intervals. The Sequen and McDermott contrasts are both able to detect change points while answering slightly different questions. Sequen visualizes the point and direction of change, while McDermott visualizes the course of the change. Of note, if the mean differences in sequential contrasts seem to be significant but switch between plus and minus, one should evaluate whether there are multiple change points or just high fluctuations. Consequently, although

change points were detected by these methods, one should still check for validity and relevance visually. Using generalized hypothesis testing may be a prefilter but the final decision should still be made by an expert of the respective field based on the context of the study.

If we would use a simple linear model without taking the nested litter/mother effects into account, the linear model would cause some type of overdispersion. In addition, our model would not reflect our true data structure. The results would include a high amount of false positives. In our case, this would mean that non-existing change points would be detected. Especially, if we would focus only on significance for decision making. As a drawback, the `lme` package sometimes has convergence or model fitting problems with small sample sizes. In some cases, the `lmer()` function displays a is singular warning that the estimated variance-covariance matrix has some entries of zero. Therefore, the matrix does not have a full rank. In these cases, it is possible that some standard errors are underestimated and should be considered with care.

In our study, we present a very special type of time series as we observe independent time points. From a statistical point of view, each time point is independent as different animals are observed at each of them. However, the experiment has an inherent order of developmental steps. Therefore, the experiment could be considered an independent time series with a limited number of observations. In addition, the time points describe different states of the same organ's development. As our aim is not to forecast any time point in the future, many applications of change point detection and analysis of time series cannot be applied to our specific experimental design. Nevertheless, we believe our suggested design might perfectly answer the typical questions in developmental studies. Of note, the time points can have different distances. The interpretation on how close the time points should be is up to the scientist. Hence, our approach is able to close even wider gaps without expression information. There is no need for the same distance between each time point or developmental stage.

We presented four in vivo expression data sets of developmental stages in mice. We decided to present different biological courses to provide evidence for its practical application: Two of the data sets did not show any abrupt changes, one first showed a steady increase over three time points, stayed at that level for some time and then increased again. The fourth data set showed no changes apart from two time points with a drastic drop in expression. The respective R code can be found in in the supplementary as well on our GitHub repository. Therefore, the presented application should easily be replicated by the interested scientist. In our work, however, we present a solution for short time series with a limited number of observed genes. If the number of genes goes into the hundreds, a visual inspection will not be feasible any longer. Hence, the scientist must sort the potential change points by effect strength in comparison to the respective relevance limits and only perform a visualization of the top relevance hits. A pattern recognition on confidence intervals is open to further research.

Conclusion

In summary, we show that multiple contrast tests can be used for change point detection in short time series. Our application is special in the sense that the indi-

vidual time points are independent of each other. Nevertheless, there is a dependent data structure within the individual time points. We showed that generalized hypothesis testing with linear mixed-effect models can be used to detect change points in short time series of clustered expression data. We deliver an approximation of the maximal length of the time series usable with our approach. The researcher can define relevance boundaries to guide decision making by the effect estimators. The usage of our algorithm is easy to apply in R. We tested three different contrast matrices and found Sequen to be the best to detect a concrete change point at a given time point. It delivers a good visualization of the position of the change point as well as an interpretable estimator of the strength and direction of the change. To determine if there is an overall significant change within the time frame, we suggest using McDermott's method as it is good at detecting change points in the time series' course. Both methods might also be used subsequently to screen results from many different time series: McDermott first and Sequen for a more detailed inspection of interesting time series second.

List of abbreviations Actb: β-Actin; Car9: carbonic anhydrase 9; Changepoint: Multiple contrast name see table 1 Glut1: glucose transporter; McDermott: Multiple contrast name see table 3 Sequen: Multiple contrast name see table 2

Declarations

Ethics approval and consent to participate All procedures were authorized by the Local Animal Care Committee (T0018/17, T0046/20, T0063/20) and performed in accordance with the guidelines and regulations of the German animal protection law. The study was carried out in compliance with the ARRIVE guidelines.

Consent for publication Not applicable.

Availability of data and material Online as supplementary material and code chunks and further information are also available from https://github.com/msieg08/clustered_data_changepoint_detection

Competing interests The authors declare that they have no competing interests.

Funding This work was supported by a German Research Foundation (DFG) grant to LKS (SC132/3-1) and KK (KK1441/4-1).

Author's contributions MS wrote, coded and provided bioinformatical insights. LKS wrote and provided clinical insights. KK wrote and provided clinical insights. JK suggested the problem and wrote. All authors have read and approved the final manuscript.

Acknowledgments The technical expertise of Nicole Dinse is gratefully acknowledged.

Author details
[1]Institute of Biometry and Clinical Epidemiology, Charité - University Medicine, Charitéplatz 1, 10117 Berlin, Germany. [2]Department of Neonatology, Charité - University Medicine, Charitéplatz 1, 10117 Berlin, Germany. [3]Institute of Translational Physiology, Charité - University Medicine, Charitéplatz 1, 10117 Berlin, Germany. [4]Institute of Medical Informatics, Charité - University Medicine, Charitéplatz 1, 10117 Berlin, Germany.

References

1. Salas-Huetos, A., James, E.R., Aston, K.I., Jenkins, T.G., Carrell, D.T., Yeste, M.: The expression of mirnas in human ovaries, oocytes, extracellular vesicles, and early embryos: a systematic review. Cells **8**(12), 1564 (2019)
2. Hu, B., Zheng, L., Long, C., Song, M., Li, T., Yang, L., Zuo, Y.: Emexplorer: a database for exploring time activation of gene expression in mammalian embryos. Open biology **9**(6), 190054 (2019)
3. Frye, M., Harada, B.T., Behm, M., He, C.: Rna modifications modulate gene expression during development. Science **361**(6409), 1346–1349 (2018)
4. Barry, D., Hartigan, J.A.: A bayesian analysis for change point problems. Journal of the American Statistical Association **88**(421), 309–319 (1993)
5. Fryzlewicz, P., *et al.*: Wild binary segmentation for multiple change-point detection. Annals of Statistics **42**(6), 2243–2281 (2014)
6. Cho, H., Fryzlewicz, P.: Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. Journal of the Royal Statistical Society: Series B: Statistical Methodology, 475–507 (2015)

7. Truong, C., Oudre, L., Vayatis, N.: Selective review of offline change point detection methods. Signal Processing **167**, 107299 (2020)
8. Aminikhanghahi, S., Cook, D.J.: A survey of methods for time series change point detection. Knowledge and information systems **51**(2), 339–367 (2017)
9. Tartakovsky, A., Nikiforov, I., Basseville, M.: Sequential Analysis: Hypothesis Testing and Changepoint Detection. CRC Press, Boca Raton (2014)
10. Sturludottir, E., Gunnlaugsdottir, H., Nielsen, O.K., Stefansson, G.: Detection of a changepoint, a mean-shift accompanied with a trend change, in short time-series with autocorrelation. Communications in Statistics-Simulation and Computation **46**(7), 5808–5818 (2017)
11. Menne, M.J., Williams Jr, C.N.: Detection of undocumented changepoints using multiple test statistics and composite reference series. Journal of Climate **18**(20), 4271–4286 (2005)
12. Sowell, F.: Modeling long-run behavior with the fractional arima model. Journal of monetary economics **29**(2), 277–302 (1992)
13. Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., Ciccozzi, M.: Application of the arima model on the covid-2019 epidemic dataset. Data in brief **29**, 105340 (2020)
14. Sneddon, L.U., Halsey, L.G., Bury, N.R.: Considering aspects of the 3rs principles within experimental animal biology. Journal of Experimental Biology **220**(17), 3007–3016 (2017)
15. Theiler, K.: The House Mouse: Atlas of Embryonic Development. Springer, Zürich, Switzerland (2013)
16. Lewis, D.I.: Animal experimentation: Implementation and application of the 3rs. Emerging Topics in Life Sciences **3**(6), 675–679 (2019)
17. Ernst, J., Bar-Joseph, Z.: Stem: a tool for the analysis of short time series gene expression data. BMC bioinformatics **7**(1), 191 (2006)
18. Muggeo, V.M., Atkins, D.C., Gallop, R.J., Dimidjian, S.: Segmented mixed models with random changepoints: a maximum likelihood approach with application to treatment for depression study. Statistical Modelling **14**(4), 293–313 (2014)
19. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823 (2014)
20. Mcdermott, M.P.: Generalized orthogonal contrast tests for homogeneity of ordered means. Canadian Journal of Statistics **27**(3), 457–470 (1999)
21. Hothorn, T., Bretz, F., Westfall, P.: Simultaneous inference in general parametric models. Biometrical Journal: Journal of Mathematical Methods in Biosciences **50**(3), 346–363 (2008)
22. Kruppa, J., Hothorn, L.: A comparison study on modeling of clustered and overdispersed count data for multiple comparisons. Journal of Applied Statistics, 1–13 (2020)
23. Goldfeld, K., Wujciak-Jens, J.: simstudy: Illuminating research methods through data generation. Journal of Open Source Software **5**(54), 2763 (2020)
24. Schad, D.J., Vasishth, S., Hohenstein, S., Kliegl, R.: How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. Journal of Memory and Language **110**, 104038 (2020)
25. Bretz, F., Hothorn, T., Westfall, P.: Multiple Comparisons Using R. CRC Press, Boca Raton (2016)
26. Hasler, M., Hothorn, L.A.: Multiple contrast tests in the presence of heteroscedasticity. Biometrical Journal: Journal of Mathematical Methods in Biosciences **50**(5), 793–800 (2008)
27. Schaarschmidt, F., Vaas, L.: Analysis of trials with complex treatment structure using multiple contrast tests. HortScience **44**(1), 188–195 (2009)
28. Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., Bates, D.: Balancing type i error and power in linear mixed models. Journal of memory and language **94**, 305–315 (2017)
29. Barr, D.J., Levy, R., Scheepers, C., Tily, H.J.: Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language **68**(3), 255–278 (2013)
30. Balcombe, J., Ferdowsian, H., Briese, L.: Prolonged pain research in mice: trends in reference to the 3rs. Journal of applied animal welfare science **16**(1), 77–95 (2013)
31. Alosh, M., Bretz, F., Huque, M.: Advanced multiplicity adjustment methods in clinical trials. Statistics in medicine **33**(4), 693–713 (2014)
32. Li, G., Taljaard, M., Van den Heuvel, E.R., Levine, M.A., Cook, D.J., Wells, G.A., Devereaux, P.J., Thabane, L.: An introduction to multiplicity issues in clinical trials: the what, why, when and how. International journal of epidemiology **46**(2), 746–755 (2017)
33. Wasserstein, R.L., Schirm, A.L., Lazar, N.A.: Moving to a world beyond "p¡ 0.05". Taylor & Francis (2019)
34. Akacha, M., Bretz, F., Ohlssen, D., Rosenkranz, G., Schmidli, H.: Estimands and their role in clinical trials. Statistics in Biopharmaceutical Research **9**(3), 268–271 (2017)
35. Ratitch, B., Bell, J., Mallinckrodt, C., Bartlett, J.W., Goel, N., Molenberghs, G., O'Kelly, M., Singh, P., Lipkovich, I.: Choosing estimands in clinical trials: Putting the ich e9 (r1) into practice. Therapeutic innovation & regulatory science **54**(2), 324–341 (2020)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Siegetalchangepointsupplement.pdf