Change point detection for clustered expression data

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 Michaela Kirschner
Institute of Translational Physiology, Charité - Universitätsmedizin Berlin
Jochen Kruppa
Institute of Medical Informatics, Charité - Universitätsmedizin Berlin

Research Article

Keywords: simultaneous confidence intervals, change point detection, multiple contrast tests, linear mixed models, expression analysis

Posted Date: November 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-605289/v3

Version of Record: A version of this preprint was published at BMC Genomics on July 6th, 2022. See the published version at https://doi.org/10.1186/s12864-022-08680-9.

RESEARCH ARTICLE

Change point detection for clustered expression data

Miriam Sieg[1*], Lina Katrin Sciesielski[2], Karin Michaela 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: To detect changes in biological processes samples are often measured at several time points. We observe expression data measured at different developmental stages, or more broadly, historical data. Hence, the main assumption of our proposed methodology is the independence between the observed samples over time. In addition, the observations are clustered at each point in time. The clustering is caused by measuring litter mates from relatively few mother mice at each development stage. The examination is lethal. Therefore, we have an independent data structure over the entire history, but a dependent data structure at a particular point in time. Over the course of the historical data, we want to identify abrupt changes in the outcome - a change point.

Results: In this paper, we demonstrate the application of generalized hypothesis testing using a linear mixed effects model as one possible method for detecting change points. The coefficients from the linear mixed model are used in multiple contrast tests. The effect estimates are then visualized with simultaneous confidence intervals. The figure of the confidence intervals can be used for the determination of the change point. Multiple contrast tests depend on the choice of the used contrast. A variety of possible usable contrasts exists. In small simulation studies, we model different courses with abrupt changes and illustrate different contrasts. We found two contrasts, both capable of answering different research questions in change point detection. Sequen contrast to detect individual points of change or McDermott contrast to illustrate overall progression. In addition, we show the application on a clinical pilot study.

Conclusion: Simultaneous confidence intervals estimated by multiple contrast tests using the model fit from a linear mixed model are usable to determine possible change points in clustered expression data. The confidence intervals deliver direct interpretable effect estimates on the scale of the outcome for the strength of the potential change point. Hence, scientists can define biologically relevant limits of change depending on the research question. We found two rarely used contrast with the best properties to detect a possible change: the Sequen and McDermott contrast. We provide R code for the direct application with examples.

Keywords: simultaneous confidence intervals; change point detection; multiple contrast tests; linear mixed models; expression analysis

Background

Independent observations over time are counter intuitive. If we observe samples at different points in time, we would assume a dependent data structure between these points in time. Each observation is then measured more than once. In our

work the observations between points in time are independent. The observations are measured at defined stages during gestation and later life. We call them development stages or in a broader sense historical data. We observe such data because the measurement of the outcome, the gene expression, is lethal at these early stages of development. Therefore, our motivated 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. At each developmental stage, litter mates and non-littermates are observed. Hence, we have a data setting with independent development stages but a dependent and independent data structure at each stage. The described setting is not common but can be observed in development studies in small mammals. Therefore, we want to present a novel methodology to find abrupt changes - so called change points - in clustered historical gene expression data.

At first glance, there are two possible approaches of how the data could be analyzed. Our motivating biological data consists of independent developmental stages of the mouse. These developmental stages naturally run over time. Therefore, two methodological approaches can be identified. On the one hand, change point detection could be applicable, and on the other hand, a dose-response analysis. However, both ignore important aspects of our research question. A change point analysis assumes that the same subject is measured repeatedly over time. The data would therefore be dependent over time. However, repeated measurement over time is not given in our data structure with lethal measurement on the mice (pups). Therefore, classical change point detection algorithms cannot be applied. Moreover, not only the position of the change point, hence the corresponding developmental stage, is of interest. We also want to report the effect of the change in expression at the change point position. The effect size is of great biological interest here. Further, we do not want to simply report the mean difference or the median difference, but also be able to adjust the effect of the change point for possible confounders. This is not possible with classical machine learning methods for change point detection. In our view, the significance is not as important as the relevance [1]. 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 [2, 3]. Our approach allows estimating the effect of the change point. In our work we used a log-normal transformation of measured expression values. Depending on the outcome, the linear mixed models also have a generalized implementation to model the full range of the exponential family [4].

The other methodological approach would be to analyze the data in the setting of a dose-response analysis. In the setting of a dose-response analysis, different increasing doses are administered. The goal of the analysis is to find the dose at which the response changes relevantly which is in our data case the gene expression. The analysis for each dose is mostly lethal but basically independent for each dose. Therefore, the setting would fit in principle. Nevertheless, dose-response data is a type of progression. Doses would correspond to developmental stages in our biological setting. However, the developmental stages do not lead to a monotonic

increase in gene expression but can be up- and down-regulated over the course. In a dose-response setting a monotonic increase would be expected or an increase with an sudden decrease. In addition, the dose has defined units and therefore the distance between each dose is nearly the same. In the dose-response setting, the multiple contrast tests are widely used [5, 6, 7]. The developmental stage intervals are not equidistant. This is caused by the gene expression data analysis based on snapshots of the transcriptome of many individuals at one development stage. 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 [8, 9, 10]. The change could be gradual over time or very abrupt. The point in development of an abrupt major change in gene expression is called a change point in our work. The multiple contrast test using for example the Changepoint contrast has been checked for statistical properties [11, 12, 4], but not discussed for the purpose of detecting outside of the dose-response setting. Hothorn (2006) [13] shows the properties and visualization of the Williams and Changepoint contrast in the setting of a randomized dose-response trials with a confidence interval oriented approaches without clustering effects. In addition, Hothorn (2006) [13] presents user specific contrasts, which might be to complicated to build for a practitioner.

Therefore, a change point algorithm is required to analyze historical data and return estimands for detected points. We define historical data in our case as data consisting of a dependent structure between points in time and a mixture dependence and independence at each time point. 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 possible confounder effects 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 the application of three different established types of contrast matrices to provide an overview on their applicability for this specific data setting. Therefore, we are able to achieve confounder adjusted effect estimates to detect change points. The effect at each possible change point can be easily interpreted by the practitioner.

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 a development data set. The respective pups are nested through their 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 historical 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 biological sets of historical gene expression data and eleven simulated data sets. The simulation settings were designed by (basic research) scientists to ensure applicability.

Biological expression data

We present a biological data set as a motivational example. 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 mouse development stages may be Theiler Stages (TS) and the day of birth (postnatal day: P) [15]. Data series in those cases consists of around 12-15 independent development stages. Additionally, at certain developmental stages and with certain data acquisition techniques, the examination is lethal and an individual can only be tested once. 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 between certain data points. As each litter introduces its own variance, this information has to be taken into account when analyzing the data.

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 illustrate 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 genes in two mouse organs. We analyzed mouse livers and kidneys from thirteen 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 normality assumptions of the linear mixed model. We provide more information on the biological data in the supplementary material 2.1. 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. 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 historical gene expression data courses, 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 historical data 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.

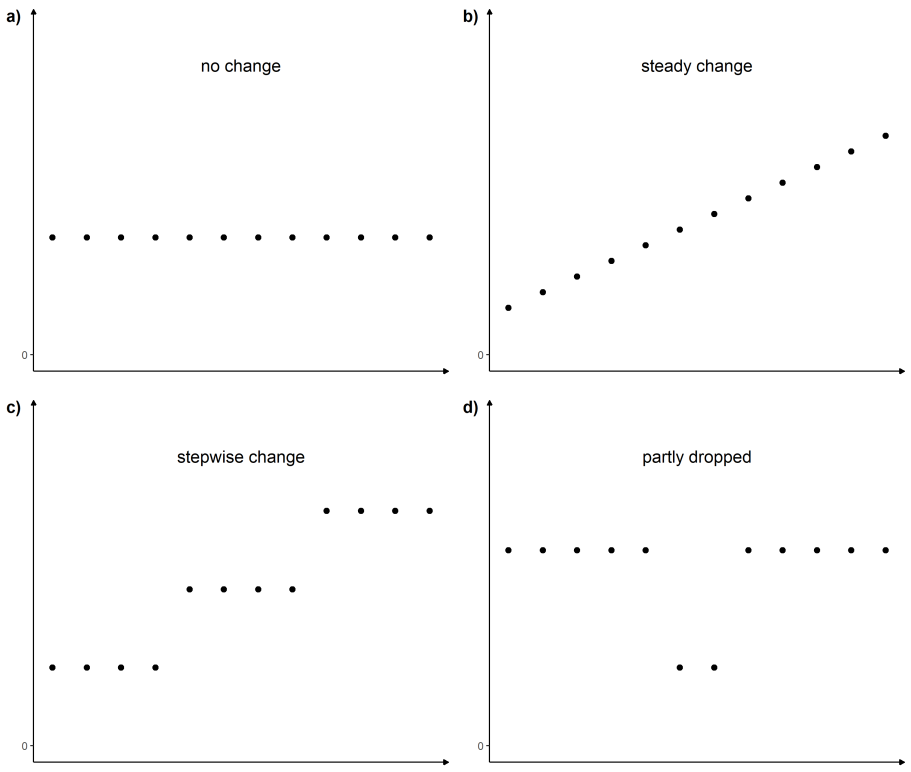

Figure 1 Possible courses of included historical data. Each subplot a) to d) represents one group of scenarios of courses within historical data. Points in time are on the x-axis, outcome on the y-axis. Scenario a) shows a course with no change. Steady changes b) and stepwise changes c) each include increase and decrease of outcome values within the historical data. 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).

We would not expect the detection of change points in the historical data 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 findings. For each of the defined historical data scenarios, 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` [17]. 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 variance was set to 2 since we expected only small differences between the expression values of the pups. We conducted a small simulation study for the variance of the mother effects with the values of 2, 6, and 10. We found no effect on the course of the confidence intervals. Therefore, the linear mixed model is able to take into account the different mother variance. The simulation results can be seen in the supplementary section 5. 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) [18]. Also Bretz et al. (2011) [19] and Hothorn et al. (2008) [20] 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 [11], complex data models [12], and even under overdispersion and small sample sizes [4]. Therefore, we consider the use of linear mixed models a valid and unbiased way to determine the estimates for the multiple contrast testing.

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 with mean parameterization can be written with fewer time points, for simplicity, as follows:

$$\overbrace{\mathbf{y}}^{150\times 1} = \overbrace{\underbrace{\mathbf{X}}_{150\times 5}\ \underbrace{\boldsymbol{\beta}}_{5\times 1}}^{150\times 1} + \overbrace{\underbrace{\mathbf{Z}}_{150\times 15}\ \underbrace{\boldsymbol{u}}_{15\times 1}}^{150\times 1} + \overbrace{\boldsymbol{\varepsilon}}^{150\times 1} \tag{1}$$

with

$$\mathbf{y} = \begin{bmatrix} 26.45 \\ 23.71 \\ \vdots \\ 12.10 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} 24.13 \\ 20.25 \\ 18.23 \\ 13.52 \\ 10.81 \end{bmatrix} \begin{matrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \end{matrix}$$

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 1 \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} 0.08 \\ 5.69 \\ \vdots \\ -4.71 \end{bmatrix}$$

where

- $\mathbf{y}$ is the 150×1 vector of normally distributed expression values,
- $\mathbf{X}$ is the 150×5 design matrix for the fixed effects considering five time points $(t_1, ..., t_5)$,
- $\boldsymbol{\beta}$ is the 5×1 vector of the fixed effects coefficients due to mean parametrization the mean of each of the five time points $(t_1, ..., t_5)$,
- $\mathbf{Z}$ is the 150×15 design matrix for the random effects of the fifteen mothers with a constant intercept,
- $\mathbf{u}$ is the 15×1 vector of the random effects coefficients i.e. the effect of the mother on the expression with $u \sim N(0, 5)$.

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 illustrate the general framework. The effects of the time points can be adjusted as in any other multiple linear regression analysis. For further clarification, we present a very short R code chunk as an example with the Changepoint contrast. The R terms can be matched to the formula 1 as follows. The `expression` indicates the $\mathbf{y}$, the variable `timepoint` the $\mathbf{X}\boldsymbol{\beta}$ as fixed effect, and the term `(1 | mother)` the $\mathbf{Zu}$ as random effects. The 1 in `(1 | mother)` indicates a constant intercept for all mothers. Mean parameterization is achieved by removing the intercept and placing 0 at the beginning of the `lmer()` formula. More complex code chunks are available in the supplementary material. In addition, we provide further R code and functions on a connected GitHub repository (`https://github.com/msieg08/clustered_data_changepoint_detection`).

Therefore, we used the `lme4` package [21] 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. If needed and the sample size is high enough, one could add additional fixed or random effect variables like id of the Polymerase Chain Reaction run or gender of the pups.

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 possible 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 [22] from the R `multcomp` package [20]. Constructions of the contrast matrices to represent each of these types can be found in the supplementary material section 4.

Table 1 Changepoint contrast for five points in time 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 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 points in time 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 points in time 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_12 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 course 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 in the historical data, 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 historical gene expression data. 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 steps

The presented approach has a theoretical limitation in the number of significant detectable differences. If many comparisons are included, we will correct each comparison for the type I error. Therefore, at a given number of comparisons depending on the maximal observed effect size δ_{max} and the corresponding standard deviation s, no *significant* change point will be detected as significant. However, the point estimator of the confidence interval will not be influenced. In addition, the approximation also depends on the chosen contrast matrix. In the following, we will examine an approximation of how many comparisons 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 significant 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. We illustrate here only an approximation, see the discussion section for further considerations. A small numeric example is given from Figure 2.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 run approximately 37 comparisons in our analysis with at least one significant confidence interval. We recommend not to concentrate on the significance but to consider the course of the point estimators. Since the confidence intervals directly represent the effect estimator, the user must decide whether the change point is relevant for the biological question. The confidence intervals provide a measure for the uncertainty, but it must be considered that the number of comparisons is included in the width of the confidence intervals.

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

We present as a motivation example biological data of the *Glut1* gene expression in the liver in figure 2 and kidney in the supplementary material, 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. Supplementary figure 3 presents the *Glut1* expression data from kidney. Table 4 shows the numerical effect estimates for the *Glut1* data from kidney. All 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 biological relevance limits. The limits are user-specific and depend on the research question. We decided to choose ±1 for our example.

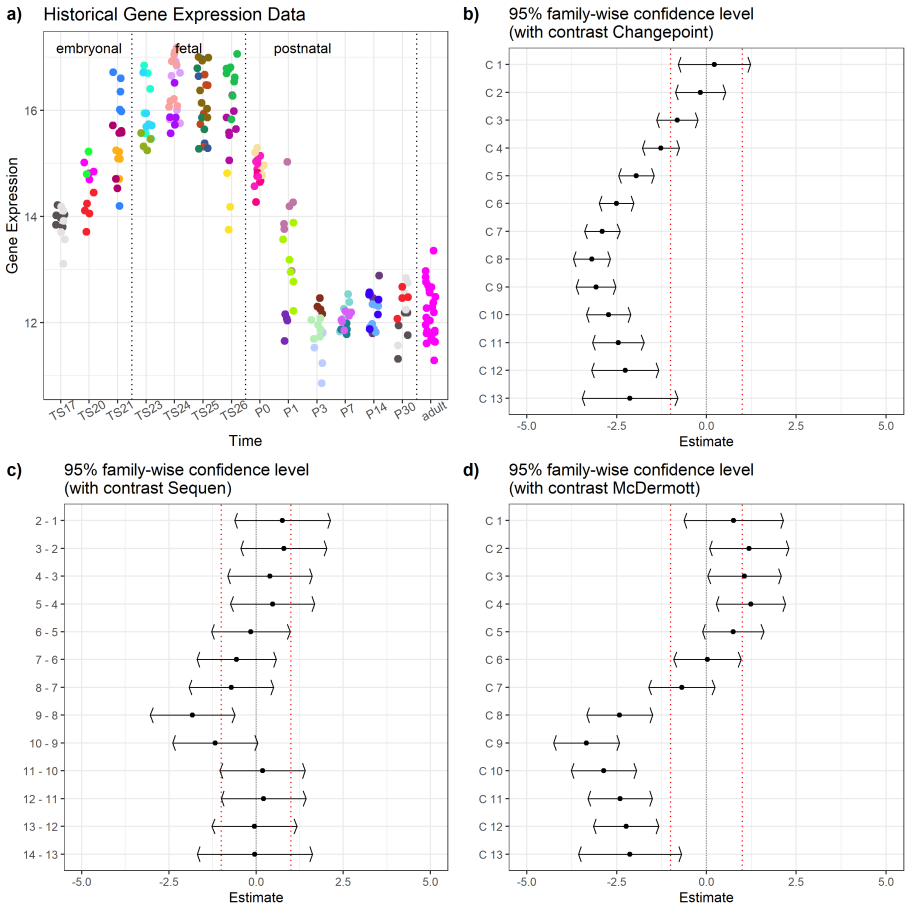

Figure 2 Biological example data for Glut1 expression in the developing liver. Subplot a) shows the biological example data set (log-transformed). Each point in time on the x-axis represents an independent developmental stage. Each data point represents a pup and each color a mother animal. The pups 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

Figure 2 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 4. The Changepoint contrast visualizes the overall course of the time 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.

Table 4 Contrasts and estimates of figure 2. 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 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	0.22	-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.

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 [23] and assure consistency of the model with the experimental design [24]. 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 the supplementary material there is no obvious expression change in the two biological examples, *Car9* expression in the developing liver and *Glut* expression in the developing kidney. All three contrasts stay within the relevance limits. The examples illustrates that both, biological visualisation and confidence intervals, are required.

Simulation data

We simulated eleven simulation settings according to figure 1 and therefore motivated by the biological examples. 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 simulation results. Figures 3 and 4 show the course in figure 1 c) and d). Table 5 and 6 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.

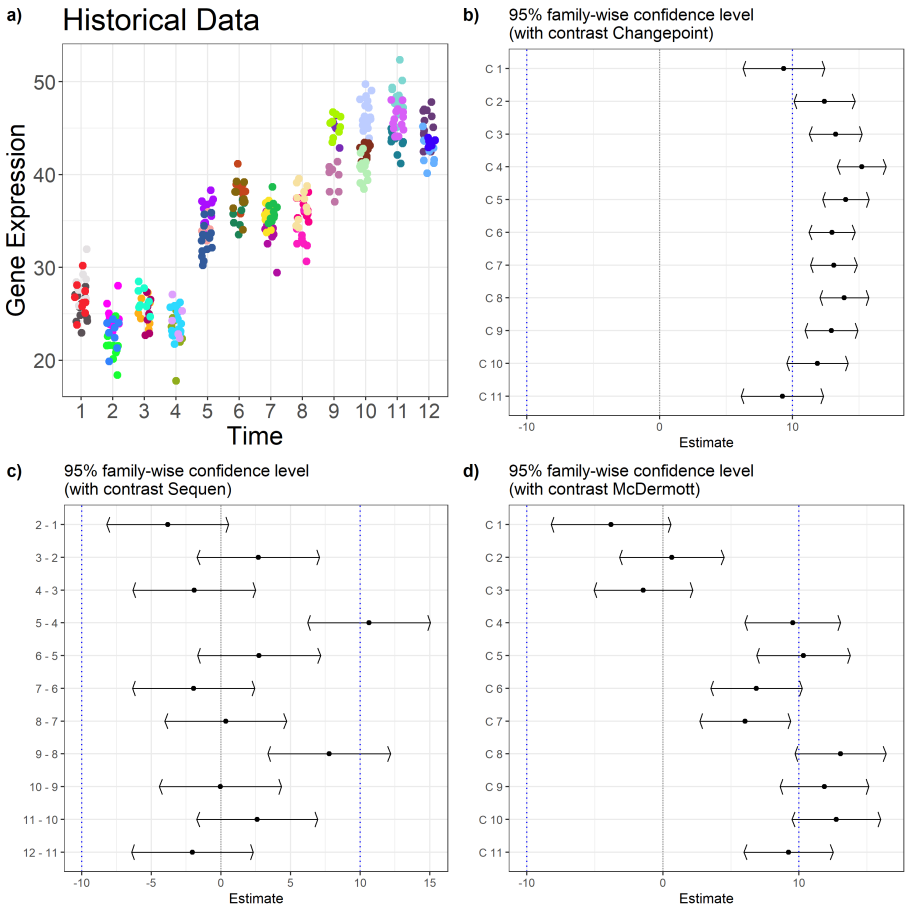

Figure 3 Confidence intervals of estimates from linear mixed model coupled with contrast matrix for historical data with two change points. Figure a) shows the points in time (x-axis) of the sampled historical data in association with gene expression (y-axis) with two expected change points. Each color 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.

Figures 3 shows a stepwise increase of expression, table 5 the corresponding numeric values. We observe two distinct change points. For illustration 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 the position and direction (Sequen) or the overall course (McDermott) of predefined change points.

Table 5 Contrasts and estimates to figure 3. 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.

Figure 4 presents a "partly dropped" change point. The corresponding numeric values are shown in table 6. 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 4 c), the Changepoint contrast confidence intervals are shown. In contrast to figure

3, 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 significant 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.

Table 6 Contrasts and estimates to figure 4. The table shows the numeric values from 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.

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 inter-

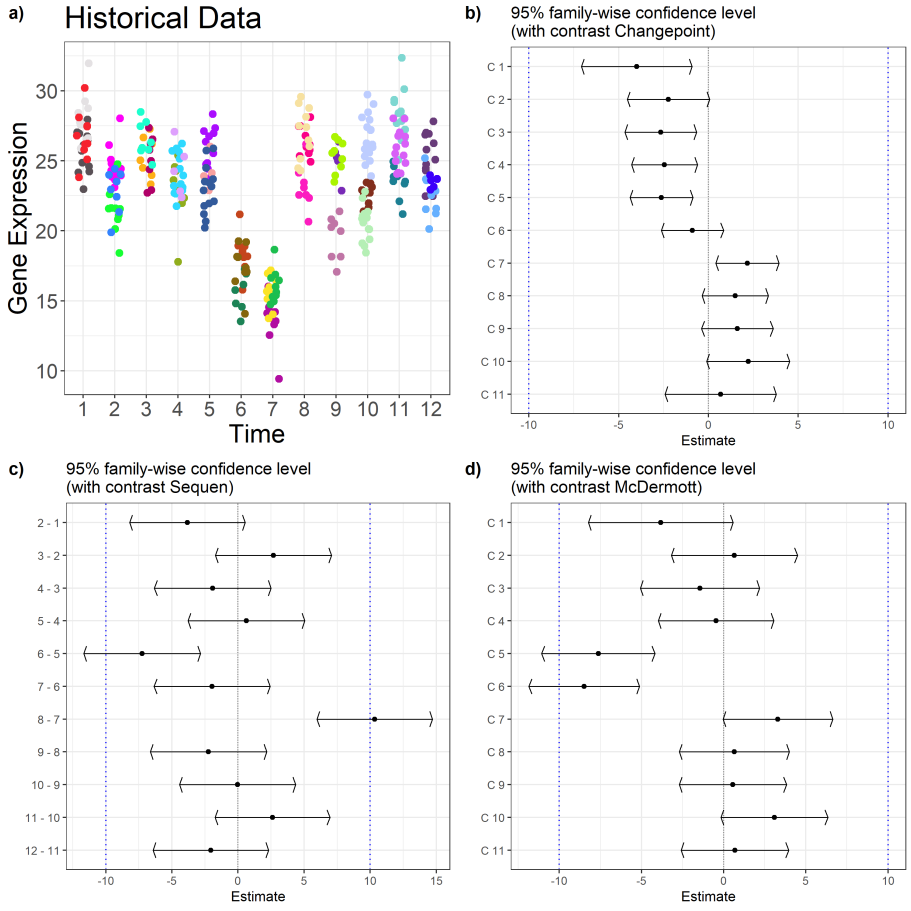

Figure 4 Confidence intervals of estimates from linear mixed model coupled with contrast matrix for historical data with a "partly dropped" change points. Figure a) shows the increasing points in time (x-axis) of the sampled historical data in association with gene expression (y-axis) with two expected 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.

vals. 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 mimics 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 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. 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 historical data of gene expression levels with independent and dependent data points.

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 historical data 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 [25, 26] will be examined in more detail.

We have discussed the possible length of historical data in terms of significance. Thus, if a confidence interval is significant, we would assume a change point. However, in the biological example data, we could also define a relevance threshold ranging from (just barely) significant to biologically relevant in our decision making. The proper choice of estimands, i.e., effect estimators, is embedded in a more general discussion of reproducibility. To date, the discussion of estimands has focused on drug development and clinical trials. Akacha et al. (2017) [27] notes that certain choices in statistical analysis can partially or completely blur the scientific question. The interested reader might read Mallinckrodt et al. (2019) [28] for a detailed discussion of estimands, estimators, and sensitivity analyses in clinical trials.

Many multiple contrast tests are well described in the literature as well as the application in statistical inference [19]. 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.

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 historical data 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 historical data. Our application is special in the sense that the individual points in time are independent of each other. Nevertheless, there is a dependent data structure within the individual development stages. We showed that generalized hypothesis testing with linear mixed-effect models can be used to detect change points in clustered expression data. We deliver an approximation of the maximal usable points in time in the historical data. 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. Confidence intervals deliver 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 changes throughout the historical data course. Both methods can also be used in sequence to verify results from historical data: First McDermott for a general overview and then Sequen for a selective examination of the course or an interval of the course.

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 KMK (KI1441/4-1).

Author's contributions MS wrote, coded and provided bioinformatical insights. LKS wrote and provided clinical insights. KMK 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. Wasserstein, R.L., Schirm, A.L., Lazar, N.A.: Moving to a world beyond "p¡ 0.05". Taylor & Francis (2019)
2. 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)
3. 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)
4. 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)
5. Rao, C.: Multiple comparison procedures-a note and a bibliography. Journal of Statistics **16**(1) (2009)
6. Bretz, F., Hothorn, L.A.: Statistical analysis of monotone or non-monotone dose–response data from in vitro toxicological assays. Alternatives to Laboratory Animals **31**(1_suppl), 81–96 (2003)
7. Bretz, F., Hsu, J., Pinheiro, J., Liu, Y.: Dose finding–a challenge in statistics. Biometrical Journal: Journal of Mathematical Methods in Biosciences **50**(4), 480–504 (2008)
8. 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)
9. 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)
10. Frye, M., Harada, B.T., Behm, M., He, C.: Rna modifications modulate gene expression during development. Science **361**(6409), 1346–1349 (2018)
11. 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)
12. Schaarschmidt, F., Vaas, L.: Analysis of trials with complex treatment structure using multiple contrast tests. HortScience **44**(1), 188–195 (2009)
13. Hothorn, L.A.: Multiple comparisons and multiple contrasts in randomized dose-response trials—confidence interval oriented approaches. Journal of Biopharmaceutical Statistics **16**(5), 711–731 (2006)
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. Goldfeld, K., Wujciak-Jens, J.: simstudy: Illuminating research methods through data generation. Journal of Open Source Software **5**(54), 2763 (2020)
18. 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)
19. Bretz, F., Hothorn, T., Westfall, P.: Multiple Comparisons Using R. CRC Press, Boca Raton (2016)
20. 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)
21. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823 (2014)
22. Mcdermott, M.P.: Generalized orthogonal contrast tests for homogeneity of ordered means. Canadian Journal of Statistics **27**(3), 457–470 (1999)
23. 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)
24. 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)
25. Alosh, M., Bretz, F., Huque, M.: Advanced multiplicity adjustment methods in clinical trials. Statistics in medicine **33**(4), 693–713 (2014)
26. 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)
27. Akacha, M., Bretz, F., Ruberg, S.: Estimands in clinical trials–broadening the perspective. Statistics in medicine **36**(1), 5–19 (2017)
28. Mallinckrodt, C., Molenberghs, G., Lipkovich, I., Ratitch, B.: Estimands, Estimators and Sensitivity Analysis in Clinical Trials. CRC Press, Boca Raton (2019)

Figure Legends

Figure 1 Possible courses of included historical data. Each subplot a) to d) represents one group of scenarios of courses within historical data. Points in time are on the x-axis, outcome on the y-axis. Scenario a) shows a course with no change. Steady changes b) and stepwise changes c) each include increase and decrease of outcome values within the historical data. 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).

Figure 2 Biological example data for Glut1 expression in the developing liver. Subplot a) shows the biological example data set (log-transformed). Each point in time 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.

Figure 3 Confidence intervals of estimates from linear mixed model coupled with contrast matrix for historical data with three change points. Figure a) shows the points in time (x-axis) of the sampled historical data 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.

Figure 4 Confidence intervals of estimates from linear mixed model coupled with contrast matrix for historical data with a "partly dropped" change points. Figure a) shows the increasing points in time (x-axis) of the sampled historical data 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.

Additional files

Additional file 1 — sieg_changepoint_supplement.pdf
Supplementary material and figures. Additional information on the biological data and additional simulation figures. Example R code is provided.

Supplementary Files

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

- Siegetalchangepointlatex.zip
- siegchangepointsupplement.pdf