Simultaneous Statistical Inference for Epigenetic Data

Konstantin Schildknecht[1], Sven Olek[2], Thorsten Dickhaus[3]*

1 Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany, **2** Ivana Türbachova Laboratory for Epigenetics, Epiontis GmbH, Berlin, Germany, **3** Institute for Statistics, University of Bremen, Bremen, Germany

* dickhaus@uni-bremen.de

Abstract

Epigenetic research leads to complex data structures. Since parametric model assumptions for the distribution of epigenetic data are hard to verify we introduce in the present work a nonparametric statistical framework for two-group comparisons. Furthermore, epigenetic analyses are often performed at various genetic loci simultaneously. Hence, in order to be able to draw valid conclusions for specific loci, an appropriate multiple testing correction is necessary. Finally, with technologies available for the simultaneous assessment of many interrelated biological parameters (such as gene arrays), statistical approaches also need to deal with a possibly unknown dependency structure in the data. Our statistical approach to the nonparametric comparison of two samples with independent multivariate observables is based on recently developed multivariate multiple permutation tests. We adapt their theory in order to cope with families of hypotheses regarding relative effects. Our results indicate that the multivariate multiple permutation test keeps the pre-assigned type I error level for the global null hypothesis. In combination with the closure principle, the family-wise error rate for the simultaneous test of the corresponding locus/parameter-specific null hypotheses can be controlled. In applications we demonstrate that group differences in epigenetic data can be detected reliably with our methodology.

Introduction

Epigenetic mechanisms, such as deoxyribonucleic acid (DNA) methylation, constitute a central principle of gene regulation. In contrast to other forms of regulation, e. g., transcriptional or translational control, DNA methylation occurs without changing the primary DNA sequence, see [1]. It refers to the selective addition of a methyl group to the 5′-carbon of the cytosine base and occurs exclusively in the dinucleotide cytosine phosphate guanine (CpG). DNA methylation occurs non-randomly and, if the target CpGs are located in the proximity of coding regions, is often associated with inactive gene expression. Oppositely, demethylation of CpG in regulatory elements is commonly accompanied by activation of expression. Shifts in DNA methylation have been observed in cells for various diseases. These changes reflect the loss of

tight gene regulation as often observed in cancer. Aberrant methylation is a hallmark of dysregulation of gene control, see [2].

On the other hand, substantial variations in methylation signals in tissues or body fluids may—while still disease associated—be derived from different changes. They may result from the changing abundance of specialized cells. For example, bacterial infections cause an innate immune response to infection, and consequently the number of neutrophils is abruptly escalated. This is invariably accompanied by an equivalent increase of neutrophil-specific methylation marks. Similarly, during human immunodeficiency virus infection, a drop of CD4+ T-cells is observed and along with that the CD4 T-cell specific DNA methylation signals drop when measuring patient whole blood or leukocytes. Thus, and insofar as methylation controlled genes are cell-type specific, the concept of differential DNA methylation is employed for the discrimination and quantification of cell types, as initially shown in [3] and used in [4].

In this work, we are concerned with statistical methods in order to identify differentially methylated loci, which may be disease markers, between two groups (typically: diseased versus non-diseased). Irrespective of the exact application, numerous different technologies are employed to identify specific methylation markers, see [5]. As consequence, studies remain disparate with limited comparability between different experiments. This also extends to statistical analysis, although for this aspect a more unified approach appears feasible, since for the majority of the approaches an estimate of the methylation proportion (β-value) at each CpG locus for each observational unit is reported. Most current studies involve the analysis of the methylation status of multiple loci separately (e. g., Illumina methylation arrays) or on a fixed sequence/pattern of loci (e. g., Methylight). Statistically, this leads to a multiple test problem. It requires a multiple statistical hypothesis test in order to find significant differences between the groups in terms of the β-value at each locus or the methylation status of a sequence of loci, respectively. Depending on the objective of the study different types of multiplicity correction are appropriate. In confirmatory studies one typically aims at strong control of the family-wise error rate (FWER), meaning that the probability for at least one type I error among the locus-specific tests is bounded by a pre-defined significance level α. In this context, a particular challenge for statistical methodology is constituted by pronounced dependencies among the β-values between the loci.

Such dependencies result from at least two different principles: On one hand, due to linkage disequilibrium (see [6]), physical proximity of different CpG sites may cause bivariate dependency, with an increasing distance between two loci generally resulting in lower bivariate dependency, cf. [7]. With respect to this, however, functionally relevant gene regulation may limit the linkage (both in extent and distance). In the Foxp3 gene, for example, the promotor region is demethylated in all T-cell types whereas the regulatory T-cell (Treg) specific demethylated region is fully methylated in most and fully demethylated in just one cell type; cf. [8]. On the other hand, when considering cell type specific markers, there is also a functional-biological dependency, which must be taken into account. For example, the number of overall T-cells in peripheral blood also influences (or depends on) the number of all cells and the number of, e. g., regulatory T-cells. Hence, the number of demethylated CD3-intergenic regions—present only on all T-cells—somewhat correlates with the number of demethylated glyceraldehyde-3-phosphate dehydrogenase (GAPDH) copies—present in all cells—and the number of Foxp3 demethylated gene copies—present only in Tregs.

These pronounced dependencies, at least within blocks of loci with small genomic distance, lead to conservativity of traditional multiple test procedures like the Bonferroni correction, meaning that α is not exhausted. For the classes of multiple test procedures considered in this work, non-exhaustion of α is equivalent to sub-optimal power characteristics of the multiple tests; cf. Lemma 3.1 of [9].

Furthermore, several confounding factors of DNA methylation have been identified in previous work, e. g., methylation is known to be highly correlated with age. A test for case-control methylation data addressing this dependency was developed in [10] and extended in [11].

Several parametric models for the distribution of the β-values have been proposed, see [12]. Their parametric nature limits their applicability in practice. A nonparametric analysis of methylation data was suggested in [13] and [14]. However, a formal notion of multiplicity adjustment is lacking in their work. In the present paper we develop a nonparametric statistical framework for tight FWER control in the context of analyzing epigenetic methylation studies, taking the described dependencies among loci into account, leading to multivariate procedures.

Methods

Throughout the remainder, we label reported results from the literature as propositions. Our major own mathematical contribution is Theorem 1.

Basic Model

Suppose that we have two experimental groups denoted by A and B, for instance given by a disease status. We consider $N \in \mathbb{N}$ observational units with n_A observables in group A and n_B in group B, such that $N = n_A + n_B$. We assume that all N observables are stochastically independent and that the observations in group $i \in \{A,B\}$ are realizations of independent and identically distributed (iid.) d-dimensional random vectors $\mathbf{X}_{ik} = (X_{ik}^{(1)}, \ldots, X_{ik}^{(d)})^{\top}$, where the index $i \in \{A, B\}$ denotes the group and $1 \leq k \leq n_i$ indexes the k-th observational unit within group i, while the superscript denotes the coordinate. The random vectors are assumed to follow the distribution $\mathcal{L}(\mathbf{X}_{A1}) = P$ or $\mathcal{L}(\mathbf{X}_{B1}) = Q$, respectively.

Example 1 (Identifying differentially methylated CpG loci) *Consider an epigenetic methylation dataset comprising d CpG loci. For each locus ℓ a methylation ratio (occasionally referred to as β-value) is defined as*

$$X^{(\ell)} = \frac{M^{(\ell)}}{M^{(\ell)} + U^{(\ell)}}, \tag{1}$$

where $M^{(\ell)}$ ($U^{(\ell)}$) is an intensity value for the amount of methylated (unmethylated) cells at locus ℓ, where we assume that suitable preprocessing steps have been performed prior to the statistical analysis. In previous literature the family of beta distributions has been considered as a model for the distribution of $X^{(\ell)}$, e. g., in [15]. However, often bimodality and skewness are encountered, questioning this parametric assumption. Notice also that numerator and denominator in Eq (1) are highly dependent. As we are not aware of a model capturing the aforementioned distributional characteristics, we propose a nonparametric approach as in [14]. An application of our general methodology to a two-sample problem involving such data is presented in Section "Identification of differentially methylated CpG loci".

Example 2 (Group differences for immune relevant parameters) *As a second example, consider the comparison of human colorectal tissue for two different stages of cancer as well as healthy controls. In Section "Association of immune cell counts with cancer" we analyze data from a study in which three immune relevant parameters were measured utilizing novel epigenetic markers based on methylation signatures in tissue. Since no prior information about distributional properties of these marker data are at hand, our nonparametric approach is applied, only making use of our basic model assumptions. Three two-group comparisons are made regarding differences of the immune relevant parameters between the disease stages.*

Aim of the statistical analysis

We denote by F_i the cumulative distribution function (cdf) of $\mathbf{X}_{i1}$ with marginal cdfs $F_i^{(\ell)}$ for each coordinate $1 \leq \ell \leq d$. We are interested in testing two families of marginal hypotheses, say $\mathcal{H} = (H_\ell : 1 \leq \ell \leq d)$ and $\mathcal{H}' = (H_\ell' : 1 \leq \ell \leq d)$. The family $\mathcal{H}$ corresponds to marginal homogeneity in the sense of [16]. This means, one is interested in testing which of the coordinate-specific marginal distributions are the same in both groups A and B, i. e.,

$$H_\ell : F_A^{(\ell)} = F_B^{(\ell)} \quad \text{versus} \quad K_\ell : F_A^{(\ell)} \neq F_B^{(\ell)}.$$

The family $\mathcal{H}'$ corresponds to finding a particular type of coordinate-specific differences. Namely, one is interested in detecting coordinates in which there are group differences in the central tendencies of the marginal distributions. To this end, recall the definition of the relative effect in the sense of [17].

Definition 1 (Relative effect) *Let X_A and X_B denote two stochastically independent random variables which are defined on a common probability space with probability measure $\mathbb{P}$. Assume that X_A and X_B have non-degenerate distributions and denote the normalized version of their cdf, as considered in [18], by F_A and F_B, respectively. Then, the relative effect of F_A with respect to F_B is defined as*

$$p_{AB} = \mathbb{P}(X_A < X_B) + \frac{1}{2}\mathbb{P}(X_A = X_B) = \int F_A \, dF_B.$$

For a d-variate distribution the relative effects can be defined coordinate-wise for each $1 \leq \ell \leq d$ by

$$p_{AB}^{(\ell)} = \int F_A^{(\ell)} \, dF_B^{(\ell)}.$$

Let $\mathbf{p}_{AB} = (p_{AB}^{(1)}, \ldots, p_{AB}^{(d)})^\top$ denote the vector of marginal relative effects in the latter case.

The functional p_{AB} is capturing central tendencies, i. e., whether realizations of one of the distributions are tending to larger values than the ones from the other. Hence, we let $H_\ell' : p_{AB}^{(\ell)} = 1/2$ with two-sided alternatives $K_\ell' : p_{AB}^{(\ell)} \neq 1/2$, $1 \leq \ell \leq d$.

Let $S \subseteq \{1, \ldots, d\}$. In the remainder, we make use of the notation

$$H_S = \bigcap_{\ell \in S} H_\ell, \quad H_0 = H_{\{1, \cdots, d\}} = \bigcap_{\ell=1}^{d} H_\ell,$$

and refer to H_0 as the global hypothesis in $\mathcal{H}$. An analogous notation applies for intersection hypotheses in $\mathcal{H}'$.

Test statistics and multiple test procedures

For the univariate nonparametric two-sample problem, i. e., for testing one particular hypothesis H_ℓ, Wilcoxon's rank sum test (or, equivalently, the Mann-Whitney U test) is commonly applied. We make use of multivariate generalizations described in [19] (for testing $\mathcal{H}$), and in [20] (for testing $\mathcal{H}'$).

Wilcoxon-Mann-Whitney (WMW) statistic.

Definition 2 (Mann-Whitney U-Statistic) *For $1 \leq \ell \leq d$, we let*

$$U^{(\ell)} = \frac{1}{n_A n_B} \sum_{j=1}^{n_A} \sum_{k=1}^{n_B} \phi^{(\ell)}(\mathbf{X}_{Aj}, \mathbf{X}_{Bk})$$

with

$$\phi^{(\ell)}(\mathbf{X}_{Aj}, \mathbf{X}_{Bk}) = \mathbb{I}\{\mathbf{X}_{Aj}^{(\ell)} > \mathbf{X}_{Bk}^{(\ell)}\}.$$

Proposition 1 (cf. Theorem 2 (iii*) of [19]) *Assume that $n_i/N \to \tau_i \in (0,1)$ for $N \to \infty$, $i \in \{A,B\}$. Then, under H_0, it holds that*

$$\mathbf{U}_N = \sqrt{N}\left(U^{(1)} - \frac{1}{2}, \cdots, U^{(d)} - \frac{1}{2}\right) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, \Sigma), \quad N \to \infty, \tag{2}$$

where $\Sigma = (\sigma_{\ell r})_{1 \le \ell, r \le d}$ with entries

$$\sigma_{\ell r} = \frac{\zeta_{10}^{(\ell,r)}}{\tau_A} + \frac{\zeta_{01}^{(\ell,r)}}{\tau_B},$$

$$\zeta_{10}^{(\ell,r)} = \zeta_{01}^{(\ell,r)} = \mathrm{Cov}(\phi^{(\ell)}(\mathbf{X}_{Aj}, \mathbf{X}_{Bk}), \phi^{(r)}(\mathbf{X}_{Aj}, \mathbf{X}_{Bk'})),$$

where $1 \le j \le n_A$ and $1 \le k \ne k' \le n_B$. In Eq (2) and throughout, $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, and $\mathcal{N}_d(\mu, \Sigma)$ stands for the d-variate normal distribution with mean μ and covariance matrix Σ.

Corollary 1 (Theorem 9.1 in [19]) *Let $\hat{\Sigma}$ be a consistent estimator of Σ. Assuming that $\det(\Sigma) > 0$ it holds that*

$$W_N^U = N\left(\mathbf{U}_N - \frac{1}{2}\mathbf{1}_d\right)^{\top} \hat{\Sigma}^{-1}\left(\mathbf{U}_N - \frac{1}{2}\mathbf{1}_d\right)$$

is under H_0 asymptotically χ^2-distributed with d degrees of freedom as $N \to \infty$.

Empirical relative effects. The empirical counterpart of the vector $\mathbf{p}_{AB}$ of relative effects is denoted by $\hat{\mathbf{p}}_{AB} = (\hat{p}_{AB}^{(1)}, \ldots, \hat{p}_{AB}^{(d)})^{\top}$ with $\hat{p}_{AB}^{(\ell)} = \int \hat{F}_A^{(\ell)} d\hat{F}_B^{(\ell)}$, $1 \le \ell \le d$, where $\hat{F}_i^{(\ell)}$, given by $\hat{F}_i^{(\ell)}(x) = n_i^{-1}\sum_{k=1}^{n_i} \frac{1}{2}\left(\mathbb{I}_{(-\infty,x]}(X_{ik}^{(\ell)}) + \mathbb{I}_{(-\infty,x)}(X_{ik}^{(\ell)})\right)$, denotes the normalized version of the empirical cdf in group $i \in \{A,B\}$ pertaining to coordinate ℓ. Notice that $\hat{p}_{AB}^{(\ell)} = 1 - U^{(\ell)}$ almost surely for all $1 \le \ell \le d$, where $U^{(\ell)}$ is as in Definition 2, under the assumption of absolutely continuous distributions (that there is zero probability for ties).

Proposition 2 (Theorem 3.3 in [20]) *Let $V_N \in \mathbb{R}^{d \times d}$ denote the matrix with entries*

$$v_N^{(\ell,r)} = \frac{N}{n_A}c_A^{(\ell,r)} + \frac{N}{n_B}c_B^{(\ell,r)},$$

$$c_i^{(\ell,r)} = \mathrm{Cov}(Y_{i1}^{(\ell)}, Y_{i1}^{(r)}), i \in \{A, B\},$$

with the transformed random variables $Y_{Ak}^{(\ell)} = F_B^{(\ell)}(X_{Ak}^{(\ell)})$ and $Y_{Bk}^{(\ell)} = F_A^{(\ell)}(X_{Bk}^{(\ell)})$. Assuming that V_N converges to a positive definite covariance matrix V as $N \to \infty$, it holds that

$$\mathbf{T}_N = \sqrt{N}(\hat{\mathbf{p}}_{AB} - \mathbf{p}_{AB}) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, V), \quad N \to \infty.$$

Furthermore, in (4.6) of [20] a consistent estimator $\hat{V}_N$ defined via the ranks of the observations has been provided.

Corollary 2 *Making use of Proposition 2 and a Studentization by $\hat{V}_N$, it follows by Slutsky's lemma in analogy to Corollary 1 that, under $H_0' : \mathbf{p}_{AB} = \mathbf{1}_d/2$, the statistic*

$$W_N = N\left(\hat{\mathbf{p}}_{AB} - \frac{1}{2}\mathbf{1}_d\right)^{\top} \hat{V}_N^{-1}\left(\hat{\mathbf{p}}_{AB} - \frac{1}{2}\mathbf{1}_d\right)$$

is asymptotically χ_d^2-distributed as $N \to \infty$.

Closure principle. A (non-randomized) multiple testing procedure $\varphi = (\varphi_1,\ldots,\varphi_d)^{\top}$ for testing $\mathcal{H}$ or $\mathcal{H}'$, respectively, is a vector of measurable mappings (individual tests) from the sample space into $\{0,1\}^d$. In this, the event $\{\phi_\ell = 1\}$ means rejection of the ℓ-th null hypothesis H_ℓ or H_ℓ', respectively. For given distributions P and Q, the FWER of φ is defined as the probability under (P,Q) of at least one type I error, i. e.,

$$\mathrm{FWER}_{(P,Q)}(\varphi) = \mathbb{P}_{(P,Q)}\left(\bigcup_{\ell \in I_0(P,Q)} \{\varphi_\ell = 1\}\right),$$

where $I_0(P,Q) \subseteq \{1,\ldots,d\}$ denotes the index set of true null hypotheses in $\mathcal{H}$ or $\mathcal{H}'$, respectively. The multiple test φ is said to control the FWER strongly at a given level $\alpha \in (0,1)$, if $\mathrm{FWER}_{(P,Q)}(\varphi) \leq \alpha$ for all possible pairs (P,Q).

One construction principle for FWER-controlling multiple tests is the closed test principle according to [21]. The general idea behind this method is to add to the system of hypotheses of interest all their intersections H_S or H_S', respectively, where $S \in 2^{\{1,\ldots,d\}}$. Even if these intersection hypotheses are not of scientific interest, they are tested auxiliarly in order to provide a multiplicity correction. Namely, a closed test procedure tests every such intersection hypothesis at full level α by an arbitrarily chosen level α test φ_S or φ_S', respectively. The adjustment for multiplicity is then performed via the decision rule that only those coordinate-specific hypotheses H_ℓ or H_ℓ', respectively, are rejected for which all intersection hypotheses H_S (H_S') with $\ell \in S$ have been rejected by φ_S (φ_S'). Thus, the price to pay for the multiplicity of the problem is that one has to perform 2^d tests. A concise description of this principle can for instance be found in Section 3.3 of [9].

Remark 1 *Application of the closed test principle is particularly convenient in our context by noticing that the assertions of Propositions 1 and 2 and their corollaries remain valid if the respective full d-dimensional vector of test statistics is replaced by a subvector which only contains the indices in the subset S to which φ_S or φ_S', respectively, refer. In the corollaries, only the degrees of freedom of the asymptotic χ^2-distributions have to be changed from d to $|S|$.*

Resampling-based approach. The results from the previous sections can also be used to construct asymptotically pivotal statistics for usage in a resampling approach. This strategy is assumed to keep α more accurately for finite N than the asymptotic methods resulting from Corollaries 1 and 2. In [22], multivariate multiple permutation tests have been developed for more restrictive families of hypotheses than $\mathcal{H}$ or $\mathcal{H}'$, namely, for families where differences of coordinate-specific functionals of P and Q, respectively, are of interest. In contrast, the relative effect depends both on P and on Q. In Theorem 1 we adapt the theory derived in [22] to the case that multivariate relative effects are of interest. Thereby, we obtain an asymptotically FWER-controlling resampling procedure based on the statistic W_N or W_N^U, respectively.

To this end, let π denote an arbitrary but fixed permutation of the set $\{1,\ldots,N\}$ and let $\mathbf{X}^{\pi} = (\mathbf{X}_{A1}^{\pi}, \ldots, \mathbf{X}_{An_A}^{\pi}, \mathbf{X}_{B1}^{\pi}, \ldots, \mathbf{X}_{Bn_B}^{\pi})$ be the matrix containing the permuted observation vectors from $\mathbf{X} = (\mathbf{X}_{A1},\ldots,\mathbf{X}_{An_A}, \mathbf{X}_{B1},\ldots,\mathbf{X}_{Bn_B})$. We make the convention that the first n_A columns of $\mathbf{X}$ and $\mathbf{X}^{\pi}$ correspond to group A and the remaining n_B columns to group B. Denote by $\tau = \tau(\pi, n_A, n_B)$ the fraction of observations from group B within the first n_A columns of $\mathbf{X}^{\pi}$, and let

$\mathbf{p}_{AB}^{\pi} = \mathbf{p}_{A'B'}$, where

$$\mathcal{L}(\mathbf{X}_{A'1}) = \tau Q + (1 - \tau)P = P' \text{ and}$$
$$\mathcal{L}(\mathbf{X}_{B'1}) = (n_A/n_B)\tau P + (1 - (n_A/n_B)\tau)Q = Q'.$$

Analogously, let $\hat{\mathbf{p}}_{AB}^{\pi} = \hat{\mathbf{p}}_{A'B'}(\mathbf{X}^{\pi})$ denote the estimator of the vector of relative effects based on the permuted data set $\mathbf{X}^{\pi}$. A simple calculation yields that
$\mathbf{p}_{AB}^{\pi} = \tau(1 + n_A/n_B)\mathbf{1}_d/2 + [(1 - \tau(1 + n_A/n_B)]\mathbf{p}_{AB}$. Finally, let
$\mathfrak{p}_{AB}^{\pi} = \tau(1 + n_A/n_B)\mathbf{1}_d/2 + [(1 - \tau(1 + n_A/n_B)]\hat{\mathbf{p}}_{AB}$.

Theorem 1 *Under the general setup from above, assume that the sample sizes n_A and n_B fulfill the regularity assumptions given in Lemma 5.3 of [23] as $N \to \infty$. Define the statistic*

$$W_N^{\pi} = N(\hat{\mathbf{p}}_{AB}^{\pi} - \mathfrak{p}_{AB}^{\pi})^{\top}(\hat{V}_N^{\pi})^{-1}(\hat{\mathbf{p}}_{AB}^{\pi} - \mathfrak{p}_{AB}^{\pi}), \qquad (3)$$

where $\hat{V}_N^{\pi}$ denotes the estimator from (4.6) in [20] applied to $\mathbf{X}^{\pi}$. Then, the permutation distribution of W_N^{π} (i. e., its discrete distribution induced by letting π be uniformly distributed on all $N!$ possible permutations of the set $\{1,\dots,N\}$, while keeping the data $\mathbf{X}$ fixed), the cdf of which we denote by $\hat{R}_N^W$, satisfies

$$\forall t \in [0, \infty) : |\hat{R}_N^W(t) - F_{\chi_d^2}(t)| \xrightarrow{P} 0, \quad N \to \infty.$$

A result analogous to Theorem 1 can be obtained for the statistic W_N^U. Based on them, an asymptotic null distribution for W_N or W_N^U, respectively, is given by its permutation distribution. This permutation distribution, in connection with Remark 1, can be used instead of $\chi_{|S|}^2$ in order to calibrate each test ϕ_S or ϕ_S', respectively, for type I error control at level α.

Proof of Theorem 1. We approximate the conditional distribution (given $\mathbf{X}$) of $\mathbf{X}^{\pi}$ by an asymptotically equivalent unconditional two-groups model. To this end, denote by $\mathbf{Z} = (\mathbf{Z}_1,\dots, \mathbf{Z}_N)$ a random matrix, the columns of which are stochastically independent such that the first n_A columns are distributed as P' and the remaining n_B columns are distributed as Q'. Following the argument of Theorem 3.5 in [20] the statistic $\mathbf{T}_N(\mathbf{Z})$ has asymptotically a centered d-variate normal distribution with some covariance matrix which is non-degenerate for eventually all large N under our general assumptions. Also, we note that both $\hat{\mathbf{p}}_{AB}^{\pi}$ and $\mathfrak{p}_{AB}^{\pi}$ consistently estimate $\mathbf{p}_{AB}^{\pi}$. Applying the reasoning of Lemma 5.3 in [23], together with the continuous mapping theorem and Slutsky's lemma, completes the proof.

Remark 2 *In dimension $d = 1$, a similar Studentized permutation approach has been discussed in [24]; see also [25].*

Results

Computer simulations

In this section we consider the performance of the proposed tests in terms of type I error control and power. To this end we present results of computer simulations under the following model.

Model 1 *For each coordinate $\ell \in \{1,\dots,d\}$ the marginal cdf $F_i^{(\ell)}$, $i \in \{A,B\}$, is the cdf of the beta distribution with shape parameters equal to $a_i^{(\ell)}$ and $b_i^{(\ell)}$. In all simulations, the value of the second shape parameter $b_i^{(\ell)}$ was fixed as $b_i^{(\ell)} = 4$ for all $1 \le \ell \le d$ and $i \in \{A,B\}$. We consider $d \in \{2,5,10\}$ and set the values of the first shape parameter equal to $a_i^{(\ell)} = 3$ in both groups for coordinates $1,\dots,d_0$, where d_0 denotes the number of true null hypotheses. For the remaining $d_1 = d - d_0$ coordinates the values of the first shape parameter in group A are taken as $a_A^{(\ell)} = 3$, while in*

group B the corresponding values are given as $a_B^{(\ell)} = 3 + \delta$, where δ takes values in
$\{0.5, 1, 1.5, 2, 2.5, 3\}$.

The dependency between the marginals is modeled by the correlation matrix R of a Gaussian
copula C_Σ, where Σ is the covariance matrix originating from R and the marginal variances in-
duced by the shape parameters. The correlation matrix $R = (R_{\ell,r})$ is of AR(1) structure, i. e., $R_{\ell,r}$
$= \rho^{|\ell-r|}$, $1 \leq \ell, r \leq d$, where ρ takes values in $\{0, 0.2, 0.4, 0.6, 0.8\}$. This model is motivated by inter-
preting coordinates as epigenetic loci and considering a decreasing strength of dependency with
increasing epigenetic distance.

First, we assessed the accuracy of the χ^2 approximation (Proposition 2 in connection with
Corollary 2) and the permutation-based approximation (Theorem 1) of the null distribution of
W_N, respectively, under the global null hypothesis. The empirical type I error rate was calculat-
ed as the relative frequency of occurrences of type I errors when testing the global null hypothe-
sis ($d_1 = 0$), i. e.,

$$\frac{1}{K} \sum_{k=1}^{K} \mathbb{I}\{\varphi^{(k)}(\mathbf{x}^{(k)}) = 1\},$$

where $\varphi^{(k)}$ denotes the test of the global hypothesis H_0 in the k-th of K simulation runs and $\mathbf{x}^{(k)}$
the pseudo-sample in simulation run k. The empirical power of the test of the global null hy-
pothesis was calculated as the same frequency for the cases with $d_1 > 0$.

Second, type I error control of the multiple tests employing the closure principle was as-
sessed by the FWER. Empirical values of the FWER were calculated as the relative frequency of
the occurrence of at least one type I error, i. e.,

$$\widehat{\text{FWER}} = \frac{1}{K} \sum_{k=1}^{K} \mathbb{I}\{\exists 1 \leq j \leq d_0 : \varphi_{j(k)}(\mathbf{x}^{(k)}) = 1\},$$

where $\varphi_{(k)} = (\varphi_{1(k)}, \ldots, \varphi_{d(k)})^\top$ stands for the multiple test in the k-th simulation run.

For the sample size N, we considered two different regimes, namely moderate ($n_A = 20$, $n_B =$
30) and large ($n_A = 100$, $n_B = 150$). Tables 1 and 2 display empirical type I error rates for testing
the global hypothesis in the moderate and large sample regimes, respectively. The empirical
power for testing the global hypothesis is presented in Tables 3 and 4. Finally, Table 5 displays
empirical values of the FWER, both in the moderate and in the large sample regime. The nomi-
nal significance or FWER level, respectively, was set to 5% in all simulations. The permutation
test was carried out as a Monte Carlo permutation test employing 9,999 randomly chosen per-
mutations of $\{1, \ldots, N\}$, together with the identity permutation.

In both sample size regimes, the empirical type I error rate of the permutation test is below
the desired level of 0.05, indicating its applicability even for moderate sample sizes. In contrast,
the test depending on critical values from the limiting χ^2 distribution performs liberally in all
simulation settings displayed in Tables 1 and 2. With increasing dimension this test even be-
comes more and more liberal. For example, its empirical type I error rate rises up to 20% for
$d = 10$. On the other hand the stronger the dependency between the coordinates, the less liberal
the χ^2-based test.

Of course, the more stringent type I error control of the permutation test, compared with
the asymptotic χ^2-based test, leads to lower power, see Tables 3 and 4. However, the differences
in power become smaller for increasing δ.

Regarding the empirical FWER (Table 5), we again observe that the permutation test keeps
the level better than the χ^2-based multiple test, where level exceedances of the latter occur for
large δ and small $\rho > 0$ in the moderate sample size regime.

Table 1. Type I error for the global hypothesis, moderate sample sizes.

	$d = 2$		$d = 5$		$d = 10$	
ρ	χ^2	Perm	χ^2	Perm	χ^2	Perm
0	0.0654	0.0428	0.1034	0.0432	0.2154	0.0480
0.2	0.0668	0.0432	0.1092	0.0478	0.2064	0.0408
0.4	0.0730	0.0488	0.1092	0.0482	0.2092	0.0476
0.6	0.0654	0.0426	0.1012	0.0494	0.1898	0.0468
0.8	0.0628	0.0460	0.0848	0.0410	0.1662	0.0448

Monte Carlo simulation results, based on $K = 10,000$ repetitions, regarding the type I error rate for testing the global hypothesis in the moderate sample size regime ($n_A = 20, n_B = 30$) for the asymptotic χ^2-based test (χ^2) and the permutation test (Perm). The data have been generated according to Model 1 with correlation parameter ρ. The nominal significance level was set to $\alpha = 5\%$ in all simulations. The permutation test was carried out as a Monte Carlo permutation test employing 9,999 randomly chosen permutations of $\{1,\ldots,N\}$, together with the identity permutation.

doi:10.1371/journal.pone.0125587.t001

Table 2. Type I error for the global hypothesis, large sample sizes.

	$d = 2$		$d = 5$		$d = 10$	
ρ	χ^2	Perm	χ^2	Perm	χ^2	Perm
0	0.0527	0.0464	0.0604	0.0448	0.0734	0.0460
0.2	0.0551	0.0456	0.0554	0.0396	0.0772	0.0500
0.4	0.0543	0.0453	0.0590	0.0440	0.0792	0.0476
0.6	0.0520	0.0440	0.0526	0.0396	0.0708	0.0458
0.8	0.0547	0.0486	0.0585	0.0460	0.0640	0.0468

Monte Carlo simulation results, based on $K = 10,000$ repetitions, regarding the type I error rate for testing the global hypothesis in the large sample size regime ($n_A = 100, n_B = 150$) for the asymptotic χ^2-based test (χ^2) and the permutation test (Perm). The data have been generated according to Model 1 with correlation parameter ρ. The nominal significance level was set to $\alpha = 5\%$ in all simulations. The permutation test was carried out as a Monte Carlo permutation test employing 9,999 randomly chosen permutations of $\{1,\ldots,N\}$, together with the identity permutation.

doi:10.1371/journal.pone.0125587.t002

Table 3. Power for rejecting the global hypothesis, moderate sample sizes.

	δ	0.5	1	1.5	2	2.5	3
$d_1 = 1$	χ^2	0.1456	0.2696	0.4540	0.6512	0.7948	0.8984
	Perm	0.0682	0.1524	0.2948	0.4964	0.6674	0.8176
$d_1 = 2$	χ^2	0.1702	0.3384	0.5834	0.7986	0.9152	0.9700
	Perm	0.0890	0.2016	0.4148	0.6556	0.8270	0.9314
$d_1 = 3$	χ^2	0.1976	0.4108	0.6882	0.8780	0.9700	0.9926
	Perm	0.1008	0.2722	0.5354	0.7824	0.9178	0.9736
$d_1 = 4$	χ^2	0.2082	0.4744	0.7768	0.9296	0.9882	0.9982
	Perm	0.1098	0.3170	0.6402	0.8592	0.9642	0.9932
$d_1 = 5$	χ^2	0.2236	0.5182	0.8168	0.9580	0.9946	0.9992
	Perm	0.1188	0.3560	0.6894	0.9056	0.9806	0.9962

Monte Carlo simulation results, based on $K = 10,000$ repetitions, regarding the power for testing the global hypothesis in the moderate sample size regime ($n_A = 20, n_B = 30$) for the asymptotic χ^2-based test (χ^2) and the permutation test (Perm). The data have been generated according to Model 1 with correlation parameter ρ and $d = 5$. The nominal significance level was set to $\alpha = 5\%$ in all simulations. The permutation test was carried out as a Monte Carlo permutation test employing 9,999 randomly chosen permutations of $\{1,\ldots,N\}$, together with the identity permutation.

doi:10.1371/journal.pone.0125587.t003

Table 4. Power for rejecting the global hypothesis, large sample sizes.

	δ	0.5	1	1.5	2	2.5	3
$d_1 = 1$	χ^2	0.2574	0.7732	0.9850	0.9998	1	1
	Perm	0.2228	0.7326	0.9804	0.9996	1	1
$d_1 = 2$	χ^2	0.3624	0.9136	0.9990	1	1	1
	Perm	0.3202	0.8938	0.9974	1	1	1
$d_1 = 3$	χ^2	0.4494	0.9676	1	1	1	1
	Perm	0.4020	0.9524	0.9998	1	1	1
$d_1 = 4$	χ^2	0.5250	0.9848	1	1	1	1
	Perm	0.4760	0.9804	1	1	1	1
$d_1 = 5$	χ^2	0.5760	0.9924	1	1	1	1
	Perm	0.5258	0.9900	1	1	1	1

Monte Carlo simulation results, based on $K = 10,000$ repetitions, regarding the power for testing the global hypothesis in the large sample size regime ($n_A = 100, n_B = 150$) for the asymptotic χ^2-based test (χ^2) and the permutation test (Perm). The data have been generated according to Model 1 with correlation parameter ρ and $d = 5$. The nominal significance level was set to $\alpha = 5\%$ in all simulations. The permutation test was carried out as a Monte Carlo permutation test employing 9,999 randomly chosen permutations of $\{1,\ldots,N\}$, together with the identity permutation.

Empirical illustration

In this section, we present applications of the proposed methods to two epigenetic studies. We applied the multiple tests based on the statistics defined in Section "Test statistics and multiple test procedures" in combination with the closure principle and Remark 1.

On one hand, we re-analyzed a representative study utilizing a whole genome approach, which aimed at the discovery of novel epigenetic markers to distinguish healthy (or good prognosis) donors from those with disease (or bad prognosis). The primary statistical challenge of such studies is the high number of locus-specific tests based on a sample with a moderate number of observations. On the other hand, we re-analyzed a data set regarding three immune

Table 5. Empirical family-wise error rates.

	d_1	0	1	2	3	4
$n_A = 20, \rho = 0.1$	χ^2	0.050	0.056	0.060	0.061	0.065
$n_B = 30, \delta = 3$	Perm	0.021	0.024	0.032	0.036	0.049
$n_A = 20, \rho = 0.5$	χ^2	0.046	0.045	0.045	0.035	0.026
$n_B = 30, \delta = 1$	Perm	0.021	0.016	0.018	0.017	0.011
$n_A = 100, \rho = 0.5$	χ^2	0.028	0.030	0.033	0.029	0.024
$n_B = 150, \delta = 0.5$	Perm	0.020	0.020	0.024	0.022	0.018

Monte Carlo simulation results, based on $K = 5,000$ repetitions, regarding the FWER for the asymptotic χ^2-based multiple test (χ^2) and the multiple permutation test (Perm). The data have been generated according to Model 1 with correlation parameter ρ and $d = 5$. The nominal FWER level was set to $\alpha = 5\%$ in all simulations. The permutation test was carried out as a Monte Carlo permutation test employing 9,999 randomly chosen permutations of $\{1,\ldots,N\}$, together with the identity permutation.

Table 6. Results for the first real data example.

Locus	cg00645579	cg00974864	cg02679745	cg08044694	cg09134726
χ^2	0.0046	0.0002	0.0002	0.0002	0.0002
Perm	0.0126	0.0029	0.0029	0.0029	0.0029
Locus	cg09303642	cg09305224	cg20070090	cg24427660	cg24777950
χ^2	0.0002	0.0047	0.0001	0.0002	0.0002
Perm	0.0029	0.0146	0.0029	0.0076	0.0029

Multiplicity-adjusted p-values of the tests for relative effects for the loci selected at the screening stage based on the asymptotic χ^2 multiple test (χ^2) and the multiple permutation test (Perm) in combination with the closure principle. The multiplicity-adjusted p-value for locus ℓ denotes the smallest significance level such that H'_ℓ is rejected for the actually observed data. The permutation test was carried out as a Monte Carlo permutation test employing 9,999 randomly chosen permutations of $\{1,\ldots,N\}$, together with the identity permutation.

doi:10.1371/journal.pone.0125587.t006

relevant parameters which were derived from cell type specific real-time PCR markers in previous work (see, e. g., [3]).

Identification of differentially methylated CpG loci. The UK Ovarian Cancer Population Study (see [26]) aimed at detecting differentially methylated loci between ovarian cancer cases and healthy controls (GEO accession number GSE19711). To this end, 274 healthy controls were compared with 131 untreated, confirmed ovarian cancer cases. Upon rigid quality control, 264 controls and 124 cases remained in the study. When applying our method, we randomly assigned 176 and 84 controls and cases, respectively, to the screening sub-sample of a two-stage selection approach (cf. [27] and references therein). We applied the univariate two-sample Wilcoxon test at each locus on the screening sample and ranked the resulting p-values in ascending order. The remaining 88 and 40 control and case subjects, respectively, were used for the confirmatory analysis (second step). The ten top-ranked loci from the screening stage were tested for a relative effect unequal 1/2 based on asymptotic critical values from the limit distribution (χ^2) and permutation-based critical values (Perm) on the confirmatory group. In Table 6, the results are presented as multiplicity-adjusted p-values. For locus $1 \leq \ell \leq 10$, the multiplicity-adjusted p-value denotes the smallest FWER level such that H'_ℓ is rejected by the respective multiple test procedure. With both methods, all ten candidate CpG sites have a multiplicity-adjusted p-value below 5%, an FWER level which is often chosen in practice.

Among the ten loci displayed in Table 6, there are two which are associated with the FUT7 gene. In turn, the FUT7 gene encodes the Alpha-(1,3)-fucosyltransferase, see [28]. This enzyme plays a role in connection with the surfaces of granulocytes, monocytes and natural killer cells.

Association of immune cell counts with cancer. As mentioned in Section "Basic Model" the discussed rank-based methods can be applied under almost no assumptions due to their nonparametric nature. Furthermore, our approach implicitly adapts to the dependency structure in the data via the permutation approach. Hence, it is especially well-suited for situations with highly dependent coordinates, for example resulting from the consideration of derived parameters.

Such a situation was present in [29]. In their study, a set of three pre-identified gene regions was considered. These regions have been shown to be associated with particular cell types. Namely, demethylated Foxp3 is associated with regulatory T-cells (Tregs), CD3 with all T-cells, and GAPDH with all leukocytes. From this, three immune relevant parameters were derived: the number of Tregs, the total number of T-cells (tTL) and the cellular ratio of immune tolerance (immunoCRIT). As the Tregs constitute a subclass of the tTL and the immunoCRIT

Table 7. Results for the second real data example.

Parameter		Treg	tTL	immunoCRIT
Cancer indicator:				
Healthy colon versus colorectal cancer	χ^2	$< 10^{-16}$	4.926×10^{-13}	$< 10^{-16}$
	Perm	0.0001	0.0001	0.0001
Cancerogenesis:				
Healthy colon versus early stage cancer	χ^2	5.292×10^{-12}	0.0024	$< 10^{-16}$
	Perm	0.0001	0.0044	0.0001
Cancer progression:				
Early stage cancer versus late stage cancer	χ^2	0.9043	9.710×10^{-5}	0.0002
	Perm	0.9044	0.0005	0.0011

Multiplicity-adjusted p-values of the tests for relative effects with respect to disease groups for three different immune-relevant parameters based on the asymptotic χ^2 multiple test (χ^2) and the multiple permutation test (Perm) in combination with the closure principle. The multiplicity-adjusted p-value for parameter ℓ denotes the smallest significance level such that H'_ℓ is rejected for the actually observed data. The permutation test was carried out as a Monte Carlo permutation test employing 9,999 randomly chosen permutations of $\{1,\ldots,N\}$, together with the identity permutation. Treg: number of regulatory T-cells, tTL: total number of T-cells, immunoCRIT: cellular ratio of immune tolerance

doi:10.1371/journal.pone.0125587.t007

is the ratio of the two other values, these three parameters are highly dependent. Nonetheless each parameter is immune relevant in its own right.

We assessed the association of the three parameters with a disease indicator for cancer, with cancerogenesis, and with cancer progression. In this context, the evaluation of the individual roles of the parameters had to be investigated. This is because cancer tolerance may be either driven by the immunoCRIT or by its individual components, i. e., the shear amount of Tregs or all T-cells. In addition, it is important to understand, even if the most important part is the immunoCRIT, which of the components drives the change during cancerogenesis. The results are presented in Table 7.

Our data indicate a statistically significant role of all three parameters with respect to all three endpoints, with the exception that the Treg parameter is not significantly associated with cancer progression. Thus, our multiple permutation test confirms the notion that manifestation of cancer is strongly associated with a shift in immune tolerance as monitored by Tregs, overall T-cells and the immunoCRIT. Notably, the change of the overall immunological tolerance from healthy towards cancer tissue is driven by both the number of Tregs and the overall number of T-cells. However, once a tumor is established the continuing increase of immunoCRIT-mediated tolerance along with higher tumor stages is mainly caused by a diminished overall T-cell number and not by Treg increase. Hence, while there is an undoubted dependency among these parameters, the biological mechanisms of cancer development allow for a detachment of these parameters such that individual changes of one of the parameters can be observed and statistically evaluated.

Discussion

Epigenetic data pose their individual set of issues for their statistical interpretation, since in contrast to DNA and protein studies, they exhibit both linkage disequilibrium-type dependencies and cell type specificity issues. Hence, dependencies have to be taken into account that go

beyond the linear and parametric linkage of genetic loci, and the cell-specific linkage of expression patterns.

Here, we assessed a new method to cope with these statistical issues in a general manner. We demonstrated how group differences in epigenetic data can reliably be detected. To this end, a statistical approach based on hypotheses regarding central tendencies in combination with nonparametric Studentized multivariate multiple permutation tests has been proposed. We adapted the theory of [22] such that it can be applied to the analysis of relative effects. In particular, our methodology addresses the so-called "null dilemma" in the sense of [16], because Studentization leads to asymptotically pivotal test statistics, even if the dependency structure differs between the groups. Our approach features four important characteristics for analyzing epigenetic methylation data: (i) The use of the relative effect as a functional for the definition of differential methylation allows to declare a shift in central tendencies in case of a significant finding. This is particularly important as other studies, see [2], have found that variation in DNA methylation may play an important role in the development of complex diseases like cancer. The restriction to shift alternatives, however, is convenient for the development of certain epigenetic markers; (ii) the permutation-based approach keeps the desired type I error level even for moderate sample sizes; (iii) carrying out the permutation test as a multivariate procedure implicitly adapts to the dependency structure in the data; (iv) as we mentioned in Section "Basic Model" the discussed rank-based methods can be applied under almost no assumptions on the distribution of the data.

Computer simulations revealed that the permutation-based approach keeps the type I error level more accurately than asymptotic χ^2 approximations of the distribution of Wald-type statistics, especially in cases with moderate sample sizes. The latter finding is in line with the observations from [30]. The convergence of Wald-type statistics towards their limiting χ^2 distribution is known to be slow and this problem becomes more severe for increasing dimensionality.

As indicated in the real data examples above, epigenetic studies usually involve several loci simultaneously based on a single sample. In many medical applications, the number of observations is very limited. Each of the given examples represents one extreme—but very common —experimental set-up: Microarray analyses with thousands of mildly dependent CpGs as in Example 1 bear a substantial risk of false positives, even when relatively high sample sizes are at hand. On the other end an unknown or complicated dependency structure in the data poses a statistical challenge. This issue is true for both directly adjacent CpGs, which are usually comethylated as well as when technically independent markers functionally overlap. The latter case was considered in Example 2 with the Foxp3 gene as marker for Tregs, and CD3g/d intergenic region as marker for the overall T-cells.

As usual for resampling procedures, our approach based on permutations in combination with the closure principle is computationally much more demanding than asymptotic approximations based on tabulated χ^2-quantiles. However, computations can be parallelized with respect to the subsets S in the closed test procedure such that the computing time can be distributed among nodes in a cluster computing system. Furthermore, efficient shortcut versions (step-down variants) of the closed test procedure can be employed; see [31] for details.

Possible extensions of our methodology comprise multi-sample problems with more than two groups, as well as the consideration of other types of limit laws (e. g., coming from extreme value theory). Finally, Edgeworth expansions as in [32] for the Wald-type statistic can prevent the costly resampling steps, at least if some concrete distributional assumptions for the observational units can be justified.

Supporting Information

S1 Table. Data for the second real data example. The table contains measurements of three immune-relevant parameters for patients in different stages of colon cancer.
(ZIP)

Acknowledgments

We are grateful to the Academic Editor and an anonymous referee for their constructive comments which have led to an improvement of the presentation.

Author Contributions

Conceived and designed the experiments: SO. Performed the experiments: SO. Analyzed the data: KS TD. Contributed reagents/materials/analysis tools: SO. Wrote the paper: KS SO TD. Developed statistical methodology: KS TD.

References

1. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nature Reviews Genetics 9: 465–476. doi: 10.1038/nrg2341 PMID: 18463664

2. Jaffe AE, Feinberg AP, Irizarry RA, Leek JT (2012) Significance analysis and statistical dissection of variably methylated regions. Biostatistics 13: 166–178. doi: 10.1093/biostatistics/kxr013 PMID: 21685414

3. Wieczorek G, Asemissen A, Model F, Türbachova I, Floess S, Liebenberg V, et al. (2009) Quantitative DNA methylation analysis of FOXP3 as a new method for counting regulatory T cells in peripheral blood and solid tissue. Cancer Res 69: 599–608. doi: 10.1158/0008-5472.CAN-08-2361 PMID: 19147574

4. Sehouli J, Loddenkemper C, Cornu T, Schwachula T, Hoffmüller U, Grützkau A, et al. (2011) Epigenetic quantification of tumor-infiltrating T-lymphocytes. Epigenetics 6: 236–246. doi: 10.4161/epi.6.2.13755 PMID: 20962591

5. Laird PW (2010) Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics 11: 191–203. doi: 10.1038/nrg2732 PMID: 20125086

6. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, et al. (2001) Linkage disequilibrium in the human genome. Nature 411: 199–204. doi: 10.1038/35075590 PMID: 11346797

7. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nature Genetics 38: 1378–1385. doi: 10.1038/ng1909 PMID: 17072317

8. Baron U, Türbachova I, Hellwag A, Eckhardt F, Berlin K, Hoffmüller U, et al. (2006) DNA methylation analysis as a tool for cell typing. Epigenetics 1: 55–60. doi: 10.4161/epi.1.1.2643 PMID: 17998806

9. Dickhaus T (2014) Simultaneous statistical inference. With applications in the life sciences. Berlin: Springer.

10. Chen Z, Liu Q, Nadarajah S (2012) A new statistical approach to detecting differentially methylated loci for case control Illumina array methylation data. Bioinformatics 28: 1109–1113. doi: 10.1093/bioinformatics/bts093 PMID: 22368244

11. Chen Z, Huang H, Liu J, Ng HKT, Nadarajah S, Huang X, et al. (2013) Detecting differentially methylated loci for Illumina Array methylation data based on human ovarian cancer data. BMC Med Genomics 6 Suppl 1: S9. doi: 10.1186/1755-8794-6-S1-S9 PMID: 23369576

12. Siegmund KD, Laird PW (2002) Analysis of complex methylation data. Methods 27: 170–178. doi: 10.1016/S1046-2023(02)00071-3 PMID: 12095277

13. Huang H, Chen Z, Huang X (2013) Age-adjusted nonparametric detection of differential DNA methylation with case-control designs. BMC Bioinformatics 14: 86. doi: 10.1186/1471-2105-14-86 PMID: 23497201

14. Chen Z, Huang H, Liu Q (2014) Detecting differentially methylated loci for multiple treatments based on high-throughput methylation data. BMC Bioinformatics 15: 142. doi: 10.1186/1471-2105-15-142 PMID: 24884464

15. Houseman EA, Christensen BC, Yeh RF, Marsit CJ, Karagas MR, Wrensch M, et al. (2008) Model-based clustering of DNA methylation array data: a recursive-partitioning algorithm for high-dimensional data arising as a mixture of beta distributions. BMC Bioinformatics 9: 365. doi: 10.1186/1471-2105-9-365 PMID: 18782434

16. Jelizarow M, Cieza A, Mansmann U (2014) Global permutation tests for multivariate ordinal data: alternatives, test statistics and the null dilemma. Journal of the Royal Statistical Society C forthcoming.

17. Brunner E, Munzel U (2013) Nichtparametrische Datenanalyse. Unverbundene Stichproben. Heidelberg: Springer Spektrum, 2nd updated and corrected edition.

18. Ruymgaart F (1980). A unified approach to the asymptotic distribution theory of certain midrank statistics. Statistique non parametrique asymptotique, Actes Journ. statist., Rouen/France 1979, Lect. Notes Math. 821, 1–18 (1980).

19. Sugiura N (1965) Multisample and multivariate nonparametric tests based on U statistics and their asymptotic efficiencies. Osaka Journal of Mathematics 2: 385–426.

20. Brunner E, Munzel U, Puri ML (2002) The multivariate nonparametric Behrens-Fisher problem. Journal of Statistical Planning and Inference 108: 37–53. doi: 10.1016/S0378-3758(02)00269-0

21. Marcus R, Peritz E, Gabriel K (1976) On closed testing procedures with special reference to ordered analysis of variance. Biometrika 63: 655–660. doi: 10.1093/biomet/63.3.655

22. Chung EY, Romano JP (2013) Multivariate and multiple permutation tests. Technical report, Technical Report 2013-05, Dept. Statistics, Stanford University.

23. Chung EY, Romano JP (2011) Asymptotically valid and exact permutation tests based on two-sample U-statistics. Technical report, Technical Report 2011-09, Dept. Statistics, Stanford University.

24. Neubert K, Brunner E (2007) A Studentized permutation test for the non-parametric Behrens-Fisher problem. Comput Stat Data Anal 51: 5192–5204. doi: 10.1016/j.csda.2006.05.024

25. Pauly M, Asendorf T, Konietschke F (2014) Permutation tests and confidence intervals for the area under the ROC-Curve. Technical report, Universität Ulm. URL http://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.105/Publikationen/FrankThomas_8.pdf.

26. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, et al. (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res 20: 440–446. doi: 10.1101/gr.103606.109 PMID: 20219944

27. Dickhaus T, Strassburger K, Schunk D, Morcillo-Suarez C, Illig T, Navarro A (2012) How to analyze many contingency tables simultaneously in genetic association studies. Stat Appl Genet Mol Biol 11: Article 12 PMID: 22850061

28. Natsuka S, Gersten KM, Zenita K, Kannagi R, Lowe JB (1994) Molecular cloning of a cDNA encoding a novel human leukocyte alpha-1,3-fucosyltransferase capable of synthesizing the sialyl Lewis x determinant. J Biol Chem 269: 16789–16794. PMID: 8207002

29. Türbachova I, Schwachula T, Vasconcelos I, Mustea A, Baldinger T, Jones KA, et al. (2013) The cellular ratio of immune tolerance (immunoCRIT) is a definite marker for aggressiveness of solid tumors and may explain tumor dissemination patterns. Epigenetics 8: 1226–1235. doi: 10.4161/epi.26334 PMID: 24071829

30. Pauly M, Brunner E, Konietschke F (2014) Asymptotic permutation tests in general factorial designs. Journal of the Royal Statistical Society: Series B (Statistical Methodology) online first,

31. Romano JP, Wolf M (2005) Exact and approximate stepdown methods for multiple hypothesis testing. J Am Stat Assoc 100: 94–108. doi: 10.1198/016214504000000539

32. Finner H, Dickhaus T (2010) Edgeworth expansions and rates of convergence for normalized sums: Chung's 1946 method revisited. Stat Probab Lett 80: 1875–1880. doi: 10.1016/j.spl.2010.08.014