Weighing biology Vol. 38, No. 2, 127-139 (2017)

Confirmatory analysis and exploratory analysis of omics research: And multiple testing *P* Focusing on value

Con fi rmatory and Exploratory Analyses in Omics Studies with Particular Focus on Multiple Testing and *P*- value

Shigeyuki Matsui

Shigeyuki Matsui

Nagoya University Graduate School of Medicine biostatistics field

Department of Biostatistics, Nagoya University Graduate School of Medicine
Mail-E: Smatsui Atto Med.Nagoya-u.Ac.Jp

In this article, we discuss the role of *P* values in multiple testing to associate a large number of genetic or molecular features with a phenotypic variable of interest in biomedical omics studies. For multiple tests in such association analyses, we distinguish those conducted for con fi rmatory purpose, as seen in genome-wide association studies to determine disease-associated variants, from those for exploratory screening of associated features. For the latter, exploratory analysis, we discuss application of the ROC curve analysis used in diagnostic medicine, as an alternative, but more relevant framework, rather than the standard framework based on multiple testing that controls false positives only. Finally, partly based on arguments made in the fi eld of omics studies, we make some comments on future endeavors by statisticians to disseminate **discussions given in the ASA's Statement on *P*- Values (Wasserstein and Lazar, 2016, The American Statistician, 70, 129-133)** to improve statistical practice in various scienti fi c fi elds.

Key words: **con fi rmatory analysis; exploratory analysis; multiple testing; omics studies; *P*- value; ROC analysis.**

1. INTRODUCTION

Recently, it is possible to capture the tissue and cells derived from a living body at the molecular level by the development of significant biotechnology, DNA Sequence (point mutation, insertion or deletion, copy number, chromosome structure and the like), transcript (gene expression),

DNA Methylation, protein expression, such as metabolites, and can be measured for various molecular data. Furthermore, the cost of sequencing and array technology, genomic, transcript, comprehensive analysis of such proteins widespread, big data of molecular data is **advanced. Whole genome, the data representing the molecular entire such as whole genes omics data (omics data) To be collectively.**

(In the example of medicine, the development of the disease, condition, treatment reactivity, etc.) o mix data and the biological phenotype of examining the relationship between is one of the fundamental analysis in omics research. More specifically, one of one

Omics variable (gene polymorphism, a probe set, etc.) test is performed on the associated phenotype variables (multiple testing; multiple testing). Since omics variable is with several thousand to several million, the vast number of *P* So that the value is calculated.

Omics research, or, I wonder if the above-mentioned multiple testing is carried out under what purpose? Taking of medical science research as an example, as the goal of omics research, 1) Generation of the disease, progression, pathology, understanding of the molecular mechanism (mechanisms) behind such treatment response, 2) Prophylaxis of medical technology (disease, diagnostic methods, the development of treatment), is mainly cited. In particular, the "new discoveries" such as the discovery of new markers that can help in the discovery and diagnosis of unknown important molecular mechanism, or, it is possible to find a great significance and value of omics research in the "generation of new hypotheses." Inevitably, new discoveries and hypotheses from omics studies will be discussed further in subsequent studies. Therefore, multiple testing in omics studies, the position and exploratory analyzes is that appropriateness. However, research can also be seen as an exception to multiple testing for the purpose of substantially verification has been carried out. Genome-wide association study for the identification of disease-related genes (genome-wide association study; GWAS) Is its representative.

This paper, 2016 A year American Statistical Association (ASA) From the issued of statistical hypothesis testing *P* Statement value (Wasserstein and Lazar, 2016) Receiving, to introduce the current situation of omics research field, in among them *P* An object of the present invention is to organize the significance and role of value. It is discussed divided into larger confirmatory analysis and exploratory analysis.

2. Confirmatory analysis genome-wide association study

2.1 Genome-wide significance level

Genome-wide association study GWAS So attempt the identification of patients and healthy subjects (respectively several hundred to several hundreds of thousand people) two frequencies different polymorphisms between groups (single nucleotide polymorphisms). The number of genetic polymorphisms hundreds of thousands to millions, test of the difference in frequency is carried out for the one by one. Of course, it is essential adjustment of multiplicity.

GWAS In, as the level of significance, genome-wide significance level (genome-wide signi fi cance level), Called a 5.0×10^{-8} Has been established that the use of. One interpretation of this significance level, approximately over the entire genome chromosomes in the general human population (European population) 100 Assume thousands of independent fragments are present, the significance level is introduced when this is applied a Bonferroni criteria (Risch and Merikangas, 1996;. Hoggart et al, 2007). It has led almost the same level of significance also some of the simulation study (Hogaart et al, 2007;. Pe'er et al, 2008;. Dudbridge and Gusnanto, 2008). Bonferroni method is the probability that a least one of the false positives in the entire test results family-wise error rate

(FWER Multiplexing to control)

The most conservative way in preparation method (e.g., Westfall and Young, 1993), Substantially, multiple testing as a confirmatory analysis can be said that are implemented. In fact GWAS In, multiple testing *P* Value, and the presence or absence of statistical significance outcomes of R (It has become one of the most important decisions of the publication).

why, GWAS In that genome-wide significance level, emergence wonder extremely stringent significance criteria are established? Generally, the occurrence of the disease, not genetics only, the influence of the environment is also conceivable, heredity since regarded as temporally precedes the environmental exposures, disease-causing mutations by searching the related gene polymorphism

(causal variant) There is a possibility that lead to the discovery of. Disease-causing mutation or gene discovery itself, as advances in science (in public) recognition easy, also easy to understand as further new therapeutic social significance because some can lead to the development of. In addition, many of the GWAS It is implemented as a major research project of the nation and the consortium. GWAS The results of the Nature And such as the series magazine, but it is often reported in the Science and top journals, more than that is considered to be the main background. These journals have a great social impact, even greater risk that where published results to walk alone in an unexpected direction. Therefore, easy to imagine the children are always holding a big fear with respect to false positives. In particular, it examines hundreds of thousands to millions of genetic polymorphisms at a time GWAS In Note Is the will allo La Defense. As a strategy journal, genome-wide significance level (verification of additional also requested) and went, can be understood that to adopt a very conservative standards (On the other hand, because the researchers want to anyway research published in top jar Anal, what trying to somehow clear even in the strict standards and do my best!). However, again, GWAS The second 1 When viewed as a omics research described in the section, the use of genome-wide significance level, must be said that detracts greatly the possibility of which is the original value of omics research "new discovery".

2.2 Other indicators, criteria

Of course, it is well recognized in academic and research community this problem is related. Just a problem that only not been elucidated part of the genetic disease factors that are estimated from pedigree analysis, so-called, missing heritability (Lost heritability) (e.g., Maher, 2008 It is also clear that a cause of). there, P Value, or, standard alternative to the genome-wide significance level has been proposed several (Note: In addition the 3 Section ROC It will be able to be added to the proposed curve analysis).

As one of the suggestions from the Bayesian approach, false-positive report probability (FPRP) There (Wacholder et al., 2004). For any of the genetic polymorphisms, of the test P The value P , The level of significance α Expressed in, FPRP It is,

$$FPRP = \Pr(H_0 \mid P \le \alpha) = \frac{\Pr(P \le \alpha \mid H_0)\Pr(H_0)}{\Pr(P \le \alpha \mid H_0)\Pr(H_0) + \Pr(P \le \alpha \mid H_1)\Pr(H_1)} = \frac{\alpha\pi}{\alpha\pi + (1-\beta)(1-\pi)}$$

Denoted. The formula of the first line are using the Bayes' theorem. In the second line $1-\beta$ Is the detection force (the original in the effect size that is assumed). π It is, the null hypothesis H_0 There is the prior probability is true. FPRP It is, of Bay's flow false discovery rate (Efron and Tibshirani, 2002) And it can be interpreted (Note: 3.3 Section FDR (c) Corresponding to).

Bayes factor (Bayes factor; BF Use of) has also been proposed (Wake fi eld, 2009):

$$BF(y) = \frac{\Pr(Y = y \mid H_0)}{\Pr(Y = y \mid H_1)}$$

here, Y It represents the test statistic. Effect size (e.g., corresponding to the true value of the log odds ratio for the presence of gene polymorphism and disease) θ Expressed in the denominator of $\Pr(Y = y \mid H_1)$ Is, any of the effect size θ Probability density of the original $f(Y = y \mid \theta)$ The prior distribution of effect size $g(\theta)$ For obtained by taking the integral. $BF(y)$ It is

H_0 For the post odds ratio and pre-odds on,

$$\frac{\Pr(H_0 \mid Y = y)}{1 - \Pr(H_0 \mid Y = y)} = BF(y) \times \frac{\pi}{1-\pi}$$

Some of the relationship. From the viewpoint of Bayesian decision is determined based on Bayes factor (data Y At a given of the original) H_0 Corresponding to the determination based on the posterior probability, the losses associated with type II error against loss due to the type I error can be viewed as minimizing (e.g., Robert, 2007, pp. 224-228). This point differs significantly from the test to consider only the control of Type I error.

Challenges in the Bayesian approach, the prior information (a priori probability π, Prior distribution of effect size $g(\theta)$) Is a set of. Given the improvement of the security and accuracy of objectivity, approach to estimate directly the prior information from the research data is attractive (experience Bayes). Fortunately, GWAS The including vast omics variables (polymorphism, probe sets, etc.) is omics research data have been measured. In other words, the data of variables with the distribution and structure something similar exists in abundance. Approach of experience Bayesian attempting to share information between the variables is considered to be particularly useful in the analysis of omics data (Efron, 2008; 2010).

As loose significant criteria than the genome-wide significance level, suggestive level Called the $1.0 \times$ Ten $^{-6}$ There is sometimes needed use. Also, I do not want to allow even one false positive FWER But not a reference, suppressed to a certain extent below the rate of false-positive (in the set of significant test) false discovery rate (FDR Control of) (Benjamini and Hochberg, 1995; Storey, 2002) Are also used in some of the research (for example, Nelson et al., 2017).

2.3 Recommendations

However, 2.1 As described in the section GWAS Given the personality, practice of "using a very conservative criteria, including genome-wide significance level" would be unlikely be overturned in the future. With this recognition, "ask the pros and cons of genome-wide significance level", "genome-wide significance level, FDR One is better of? Have remained on the idea, such as "is a big step forward seems to be no. One realistic approach, in addition to confirmatory analysis using genome-wide significance level, above Bayesian index, suggestive level, FDR It is conceivable to implement the exploratory analysis using such simultaneously. For example, in the commentary deserves clinical trials therapy usually analysis of analysis and other by-the following specific items of the primary endpoint (primary endpoint) (secondary endpoint) are carried out simultaneously. Of course, it must be by strictly ward. GWAS In the paper, for example, " Exploratory Analysis for Hypothesis Generation What about, "such as the billed was to try be provided with a section or supplements? On the other hand, the search criteria (suggestive level, FDR When the like), such as combined with confirmatory studies using another independent population (corresponding to the evaluation of external validity), devised for sufficiently reduce the likelihood of false positives obtained (For example, Nelson et al., 2017). Such direction was hardly seen the study was still strains for, is considered in the future become important as an approach to replace the genome-wide significance level.

3. Exploratory analysis-related screening

3.1 Or framework of multiple testing appropriate?

The 1 As I mentioned in the section, association analysis of omics data and phenotypic variables in omics studies in general, positioned as exploratory analyzes for the associated screening. In other swing so far of the discussion of the relevant analysis, most of which have been carried out within the framework of multiple testing. Omics analysis (or

Before the analysis of high-dimensional data) the advent of, targeted the relatively limited number of test "of Me other verification analysis FWER Discussion of control "was the mainstream. There, in recent years, represents out analysis of high-dimensional data, how to adapt the methodology, method of multiple testing of up to it (adaptation) Or should it do? There came to be discussed, in which, FWER Alternative to the standard FDR New development occurs, such as a standard, and has been established a framework of multiple testing in omics research of today's authors understand. That is, in drew the flow of original confirmatory analysis, or it can be said that an extension of the confirmatory analysis is established. However, confirmatory analysis (Related verification) and the exploratory analysis (related screening) is, naturally, is not different from the root this manner? It will not occur question.

Again, given the significance and value of omics research, the 1 As discussed in the section, it is in the "new discovery". In fact, to researchers who carried out the omics research has much to look forward to "new discovery", Tokimei to have (the author's experience). For omics research, how it has how much ability to new discoveries, that it is considered very important. When referred to in the framework of multiple testing, it can detect how much the true-positive, aspect that is important. However, in the framework of multiple testing only false positive are considered, subject to controls. Now it does not say that unfair is?

3.2 ROC Curve analysis

Diagnostic Medicine (diagnostic medicine In a field of), there is a disease screening intended for healthy individuals of the population. For the detection of certain diseases, as a first step, is relatively noninvasive by a simple diagnostic method is performed, the subject suspected of diseases are sieving (screening of the disease). there " Positivity " The By now, the Target Company, as the second stage, more involving invasion, an attempt is made more definitive diagnosis. As used diseases screening in the first stage of the (simple) diagnostic methods performance, sensitivity (sensitivity) And specificity (speci fi city) By Ri being evaluated is the basic. Sensitivity is the probability of positive in the population with disease, specificity is the probability of negative in a population without disease. Often, the determination of the positive, for a particular marker, it threshold (cutoff, cut-o ff) It is made on whether or not exceeded. Moving the cutoff, the sensitivity, even specificity varies. This relationship, the vertical axis sensitivity, the horizontal axis 1 - It is a representation taking specificity (the probability of false positives) ROC curve(receiver operating characteristic curve) It is. Selection of the cut-off is performed in consideration of the balance between sensitivity and false positive rate (or specificity). For the author, this ROC Framework of analysis, natural framework and considered Erareru to smoothly fit to related screening in omics research.

So, in omics research, how ROC I wonder if it is sufficient to constitute a curve? Corresponds to the diagnostic marker used in disease screening is appropriate statistics capture relevant individual OMICS variable phenotype variables (such as the test statistic). Corresponds to the cutoff of the positive determination is cut off on the statistic. Test statistic (P When using a value), critical values (significance level) corresponds to the cutoff.

The corresponds to the sensitivity of the vertical axis, H_0 But not satisfied, ie, non-null Detection output when the correct, or it is conceivable to truly using the detection power in phenotype assay of the associated OMICS variables. More specifically, all of non-null An average detection strength in the OMICS variables. This has been used in the detection force evaluation of Omiku be studied (Tsai et al, 2005;. Shao and Tseng, 2007; Tong and Zhao,

2008 Such). (Note: Detection force evaluation, these only multiple testing, or has a sample size design purposes, Ya effect size distribution from the data as described below ROC The curve to estimate, there is no idea to select a cut-off). The average detection force of more than, non-null Omics variables " The entire " Since the average detection strength on the entire detection power (overall power) And will be referred to as (Matsui and Noma, 2011a). However, non-null Some OMICS variables, since what effect size is small is present quite a few of the calculation of the overall detectability is generally enough to disappointed small value (for example, < 20% Often become). First of all, including those that have a very small effect size, non-null The Kokoromiruko the omics variable overall detection of will hardly unreasonable thing. So, it has the effect size of certain level or more non-null Limited omics variables, or is the average detection strength at Moto targeting " portion " Detection force (partial power) (Matsui and Noma, 2011a Use of) can be considered (below for specific settings Method 3.3 See the section). Omics variables (genetic polymorphisms, gene, etc.) having a large effect size, if it is possible to obtain considered the more important in the biological, the use of partial detection force would be appropriate.

On the other hand, as an index corresponding to the false positive of the horizontal axis FDR It would be good to use. As its reason, that is easy to interpret as a percentage of false positives in the screening are omics variable set was, also, its use is such that you are already well established in many of omics research and the like.

ROC By drawing a curve, average detection force (true positive degree), FDR (False positivity) relationships visually captured can Rukoto, it is possible to obtain useful information to help in the selection of the cut-off in consideration of both. Omi box research screened omics variables (such as genes), function analysis to examine the biological mechanisms, pathway analysis, or a more convenient platform (e.g., prevalent in clinical practice RT-PCR Subject to validation studies of, for example). Research resources in these subsequent studies (funding, between the time, manpower, etc.) Given the constraints, would be is not practical to screen for too many omics variable. In that sense, FDR Is, also it includes some severe level, 1 20% Degree would be a range of commonly accepted put. On the other hand, the average detection strength on the vertical axis, again, (all non-null Omi for the box variable) the entire power to detect a lot can be quite low level, can not be too much recommended. The use of partial detection force at a limited effect sizes greater than a certain would be more realistic. It should be noted that, since it is variously considered to put the way of the effect size limit, can define some part detection force. below 3.3 An example is shown in the section. It should be noted that, 3.3 Section, the statistical model, Dale contains a little technical content on Estimation Method is, ROC Image Figure curve (overall detectability, section detecting force) 2 It would seize by looking at the. Eventually, FDR Also based, in consultation and co-workers (such as biological and medical researchers), it is preferable to select a cut-off.

3.3 ROC The estimation of the curve

above ROC Analysis, from data ROC For the first time possible by estimating the curve. In omics research ROC Estimation problem of the curve has hardly been discussed so far. The reason for this is that, it is believed that the flow of multiple testing overemphasis of Until now also, of course, ROC Estimation framework is fully aware of the modeling for the entire omics data to help in the curve, wax that may be mentioned that has not been organized.

In the following, first, ROC Touch on omics data whole of modeling to estimate the curve,

Moreover ROC Given the estimate of the curve. It should be noted, Runode seems that is easy to understand more specific person who was using the case, here, Setlur et al. (2008) Consider the gene expression microarray study of prostate cancer by. Cancer fusion gene to be associated with malignancy and prognosis of prostate cancer has been known (TMPRSS2-ERG) have $n_1 = 103$ Not the name of the prostate cancer patients have this $n_0 = 352$ For the name of prostate cancer patients, the cancer cells cDNA Array analysis (6,144 Number of probe sets) is performed. Here, two of the expression levels of different probe sets between patient groups (hereinafter, for convenience referred to as a gene) consider the screening of.

First of all, gene j ($j = 1, \cdots, 6, 144$) Respect, the difference between the average expression level between groups effect size parameter $\delta_j = \mu_j(1)$

$_j - \mu_j(0)$ $_j / \sigma_j$ It is defined as. here, $\mu_j(1)$ $_j$, $\mu_j(0)$ $_j$ Cancer fusion gene holdings group, The average expression level of non-carrier group, in the group (common) standard deviation σ_j In dividing and you are standardized. As statistics for this, $Y_j = (\hat{}$ $\mu_j(1) - \mu_j(0)$ σ_j It is used. In addition, two-sample t Test statistic, $t_j = Y_j / \tau_n$, $\tau_n = p\,1/n_1 + 1/n_0$ Since denoted, substantially, t Is the same as the use of the test.

Suppose a mixed model of following. That is, Y Probability density function of $f(y)$ Against,

$$f(y) = \pi f_0(y) + (1 - \pi) f_1(y)$$

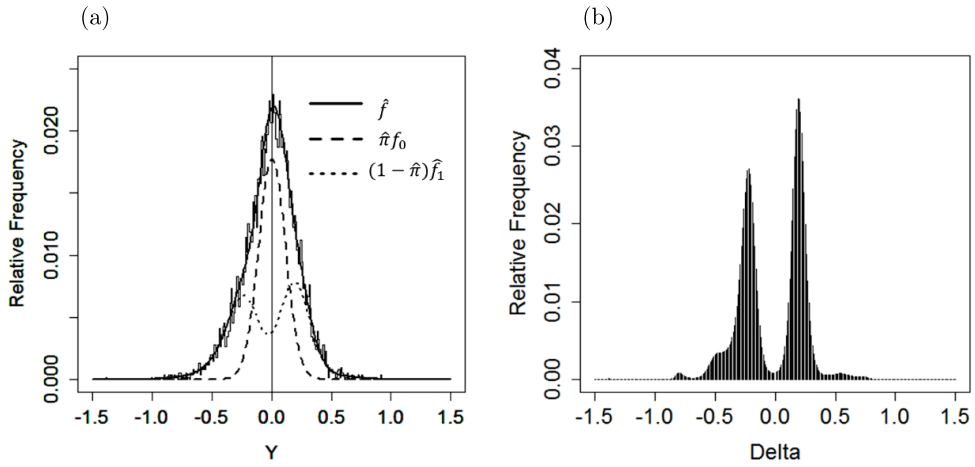
The assumed. here, π It is null It is the prior probability of the (no relation). f_0 It is null At the time of the Y Of the distribution (probability Density function), f_1 It is non-null At the time of the (relevant Yes) Y Which is a prior distribution. f_0 For, theoretically guide Wither null Distribution, for example, as the asymptotic distribution $N(0, \tau_2$ $_n)$, But to specify is generally the (theoretical null), $N(\gamma_0, \varphi_{20})$ Put and, γ_0, φ_{20} It is estimated from the data (empirical null) May be considered (Efron 2004a). Posterior probability π Pre-distribution and f_1 May be specified in the appropriate analysis's, 2.2 As discussed in the section, approach of experience Bayesian be estimated from the data is valid. In the maximum likelihood estimation, in most cases, although independence between convenience omics variables are assumed constant robustness is shown for the correlation (McLachlan, Bean, and Jones, 2006). If, as long as it can specify a correlation from a biological point of view, estimating is also possible to incorporating the same. The mixture model described above, FDR Although evaluation been widely used in the discussion of multiple testing for the purpose of the entire detection power given below (Ψ I would like to note also use it in the evaluation of).

Now, null / non-null In addition to the mixing arrangement, non-null distribution f_1 Put the hierarchy, to introduce the effect size distribution.

$$Y_j \mid \delta_j \sim N(\delta_j, \tau_2$$ $_n),$ $$\delta_j \sim g$$

here, g Corresponding to the effect size distribution. g May assume an appropriate parametric distribution is relative, it is (since the number of genes is large enough) Nonparametric estimation also realistic. In general, the advance is little information about the effect size distribution, the latter approach that it is not necessary to assume the distribution is attractive. Peripheral distribution f It estimated in the appropriate spline function, a method of obtaining by inverse transformation (Efron, 2004b) , g Assuming a discrete distribution to EM It is the algorithm smoothing-by-roughening Law (Shen and Louis, 1999 How to Apply) (Matsui and Noma, 2011a, b) It has been proposed.

Effect size distribution g Assuming a non-parametric distribution to, smoothing-by-roughening Figure the result was estimated boss by law 1 To show. Drawing 1 (a) It is, all 6,144 Statistics of gene Y The histogram of the relative frequency



Drawing 1. Setlur et al. (2008) Comparison of fusion genes possess a non-carrier using microarray gene expression data in prostate cancer research. panel(a), The total 6,144 Statistics of gene Y Of the histogram, peripheral distribution f, null component πf_0 , non-null Component (1 - π) f_1 Of the estimated (Y Of the grid: 0.01). panel(b) It is estimated nonparametric effect size distribution g .

In contrast, peripheral distribution f, null Components of πf_0 , non-null Of component (1 - π) f_1 Shows the estimates. Drawing 1 (b) , The effect size distribution g It is the result of estimation. It should be noted that, null Prior probability of π Estimates $\hat{}$

π = 0.6 Met.

Now, as genetic screening, statistics Y Against | Y | ≥ c (> 0) Consider the rules that determined to be positive at the time of the (corresponding to a two-tailed test). At this time, FDR , The entire detection force Ψ It is cut off c Represented as follows as a function:

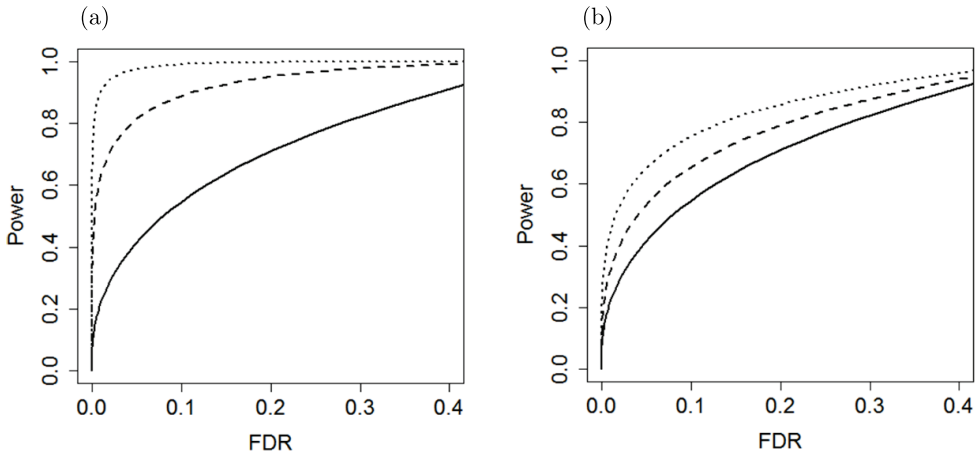
$$FDR(c) = \frac{\pi_0\{F_0(-c) + 1 - F_0(c)\}}{F(-c) + 1 - F(c)}$$

$$\Psi(c) = F_1(-c) + 1 - F_1(c)$$

here, F, F_0, F_1 It is f, f_0, f_1 The corresponding distribution function.

Then, only above a certain effect sizes, or let seek partial detection power is the average power to detect that target. Obviously, the evaluation of partial detection force, non-null distribution f_1 Consider the hierarchical structure, the effect size distribution g It made possible by the introduction of. In the case of prostate cancer, effect size δ With respect to, in the negative direction $\delta \le \eta_1 (< 0)$, In the positive direction $\delta \ge \eta_2 (> 0)$ Effect size meaningful ones satisfying, or is defined as the effect size to be detected (i.e., the condition is satisfied effect size δ_j Omics variable with the j Gathering the target to be detected is). constant η_1 , η_1 For the designation, For example, the effect size distribution g Of the estimated value Ten% point, 90% It may be used as such a point. In the data of prostate cancer,

Ten% Point η_1 = - 0.34 , 90% Point η_2 = 0.26 It is. Alternatively, look at the shape of the nonparametric estimated effect size distribution, will be designated to detect the peak of interest. For example, as shown in FIG. 1 (b) Looking at the estimation result of the effect size distribution of, the negative, the peak of the large effect size in plus each direction can be seen a few. To detect these, for example, η_1 = - 0.60, η_2 = 0.42



Drawing 2. Prostate cancer research data. panel(a) It was estimated ROC curve. The solid line overall detectability, dashed-dot line portion detects force (broken line: $\eta_1 = 0.34$, $\eta_2 = 0.26$,dotted line: $\eta_1 = -0.40$, $\eta_2 = 0.40$). panel(b), The estimate of the overall detection force at the time of changing the sample size (solid line: $n = 455$, Dashed line: $n = 600$,dotted line: $n = 800$).

It would be designated as such. Part detection force,

$$\Psi_{P(C;\,\eta_1,\,\eta_2)} = \frac{\int_{|u|\geq c}\int_{\delta\leq\eta_1,\,\delta\geq\eta_2} g(\delta)\,\varphi_{\delta,\tau^2}(u)\,d\delta du}{\int_{\delta\leq\eta_1,\,\delta\geq\eta_2} g(\delta)\,d\delta}$$

And it can be expressed. here, $\varphi_{\delta,\tau^2}(\cdot)$ The average δ ,dispersion τ^2 It is a density function of a normal distribution with. Drawing 2 (A) When the entire detecting force, or, using a portion detected force ROC It is an estimated result of the curve. For example, FDR = 10% If you specify, the overall detection force Fifty percent The extent to reach (Note: Sample size is true there is a rare high level relatively large). On the other hand, partial detection force is much higher level than the overall detectability. If you want to capture the effect size is particularly large upper genes, FDR = 5% Even if strictly to the extent 80% Would the partial detection force can be achieved in excess.

Effect size distribution g To estimate the sample size design is possible. Drawing 2 (b) It is, π When g The at Moto fixed to their estimates, for any sample size, is the sample size section τ^2 Update the one in which estimated the average detection power. In particular, Setlur et al. (2008) In, TMPRSS2-ERG Fusion gene holdings, the proportion of non-carrier patients (= $n_1 / n_0 = 103/352$) In Moto was a constant, the overall sample size n The 455 (= $n_1 + n_0$) From 600 , 800 It shows the estimated value of the overall power to detect when changing the example. Such analysis is, for example, Setlur et al. (2008) A new plan similar omics research to reference, is useful in performing the sample size design. Alternatively, Setlur et al. (2008) As research is being continued, the study of this type implemented as "intermediate analysis", will be performed "Sample Size redesign" to achieve the desired detection power.

3.4 Hierarchical mixed model and experience Bayesian estimation: further in the relevant screening analysis possibilities

Original paper Setlur et al. (2008) In, with a selective set of genes TMPRSS2-ERG Determination of the holdings of the fusion gene has been attempted. Discrimination and prediction system using a set of genes that were selected after omics research (genomic signature When developing the), the ability to discriminate how much relative selection gene set

If you can assess whether it can be expected in advance, subsequent discrimination Research and Development, go / no-go Help to the study of the judgment and research design. The experience Bayesian analysis using the hierarchical mixture models, can estimate the effect size for each gene were selected (shrinkage estimation), determination accuracy estimation considering the correlation between the gene for any subset of genes It can be performed. For more information Efron (2009), Matsui and Noma (2011b)

See. Although a FAQ List misunderstanding, discrimination and prediction capability and statistical significance is completely different. Obtaining a predictive ability meaningful statistically significantly higher hurdle than to obtain a significance (eg example, Pepe 2005). Therefore, genomic signature If priority is given to the development, it is quite simply the analysis of statistical significance insufficient discrimination accuracy evaluation of the model-based, or it is necessary to perform accuracy evaluation by standard identification and prediction analysis .

Also of interest is applied to a more complex situation. One is a case where there are a plurality of subgroups for the patient sample. For example, in randomized clinical trials to compare the survival time of cancer patients in the control treatment group and the new treatment group, the microarray gene expression analysis is performed on cancer tissue prior to treatment (baseline) to (Matsui et al., 2012). Assuming that investigate the effect of gene expression on survival time by treatment group (univariate which was one of the covariate the gene expression level Cox Analysis etc.), gene effects the expression level is common in the treatment groups prognostic markers (prognostic marker), Those that do not (i.e., interaction of the therapeutic gene), the therapeutic effect predictive marker (predictive marker It is as possible de regarded as a candidate of). The above hierarchical mixture model is extended to two dimensions, the new treatment, performed simultaneously modeling the expression of the effect obtained in two groups of control treatment, and estimates the effect size distribution nonparametric. Then, in a word, to say that the interaction of the therapeutic gene will be appreciated that there are those with a variety of effects profile. Therein they will also include not a little out of nothing that can not be adequately detected by test standard treatment interaction. It was estimated based on the (non-parametric) effect size distribution, corresponding to the likelihood ratio test optimal discovery procedure –

(Storey, 2007; Noma and Matsui, 2012) Constitute the

It is, genes having various effects profile (prognostic marker candidate therapeutic effect predictive markers) can efficiently detect (Matsui et al., 2018). Above example shows that to be estimated truth signal component contained in the omics data using a flexible hierarchical mixture model is valid in the context of screening.

3.5 *P* Meaning and role of value

Again, the test to be performed on individual omics variable *P* Let 's consider the roles and meaning of value. In the framework of multiple testing, FWER The indicators of false positives in the entire test to be representative in order to keep below a certain level *P* Value (or significance level) is adjusted. In the first place *P* Value is indicative for a statistical significance of a single test (test any one of Omi box variables). There false positives wide standards test (total omics variable) is reflected in the adjustment for determining the significance of the individual tests (test every single OMICS variable)

P Value is created. That is, If performing relevant screened side only false positive (reflecting the test constant overall false positivity) Adjustment *P* To determine the screening cutoff through value is natural.

on the other hand, ROC In the framework of the curve analysis, in the sense of determining the screening of the cut-off *P* The role of value is reduced by half. This is because, true also positive is because the framework to be considered at the same time. However,

(*P* Role as a ranking of the hidden in the role) omics variable values are leaving (However, ROC

Although not always necessarily lead to ranking based on the statistical significance in the framework of curve analysis).

Four. in conclusion

This paper focuses on the relevant screening analysis in omics research, was organized with the statistical methods and their present situation. GWAS The has been discussed confirmatory analysis and other exploratory analysis to the minute Ke and of a part of the study to be representative, related analysis in omics research is exploratory analysis and position towards the basically subsequent research Dzu kick of it is reasonable. The 1 As discussed in the section, the motive of this writing is ASA by *P* Statement value (ASA Statement) and to some relationship, but the first 3.5 As discussed in the section, in the relevant screening analysis in omics research (exploratory analysis) *P* The role and significance of the value is considered to be limited. Alternatively, the relevant screening *P* It is considered to be related to the value, or even want cormorant danger captures rather the wrong-related screening. The author, related screening, rather than a "hypothesis testing of the problem" is, omics whole of the relevant structure using a statistical model, or, ROC "Estimate of the problem" of the curve, and, based on the estimated result believes that it should be seen as "a matter of choice."

Finally, (not limited to omics research) *P* About the value ASA statement(Wasserstein and Lazar, 2016) Described a little personal opinion about even One meaning is. The main purpose of this statement is that seen in many research fields of real science *P* Solve the misunderstanding about the value, *P* More appropriate use of the value, or, it is clear that is to encourage the use of another suitable approach, methods. In many areas of research, the content of this statement is, integration Keiie is, of course, is well recognized to researchers and practitioners who are not specialized in statistics, and a goal that practice is improved statistical methods it can be said. The aim there, to separate each field of study, what kind of *P* Whether the misunderstanding and misuse of value can occur, along with the specific case, stand tried this thing to Nari author in this paper in the field of the possible solutions to be also presented together it is considered the first step (omics research is Mori is). Only show at the level of the general theory of solutions effectiveness would be poor. In each of the research field, it is essential and deeply opposed this from the front to the specific issues statisticians, together with researchers in the field, it is more sought its efforts in the future statistician ASA Seems to be such a big meaning has been put in the statement.

——

Bibliography

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate-a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 57, 289-300.

Dudbridge, F. and Gusnanto, A. (2008). Estimation of signi fi cance thresholds for genomewide association scans. *Genetic Epidemiology* 32, 227-234.

Efron, B. (2004a) Large-scale simultaneous hypothesis testing:. The choice of a null hypothesis. *Journal of the American Statistical Association* 99, 96-104.

Efron, B. (2004b). Selection and estimation for large-scale simultaneous inference. *Technical Report* No. 2005-18B / 232, Division of Biostatistics, Stanford University.

Efron, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statistical Science* twenty three, 1-22.

Efron, B. (2009). Empirical Bayes estimates for large-scale prediction problems. *Journal of the American Statistical Association* 104, 1015-1028.

Efron, B. (2010) Large-Scale Inference:. Empirical Bayes Methods for Estimation, Testing, and Prediction New York:. Cambridge University Press.

Efron, B. and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genetic Epidemiology* twenty three, 70-86.

Hoggart, CJ, Clark, TG, De Iorio, M., Whittaker, JC, and Balding, DJ (2008). Genome-wide signi fi cance for dense SNP and resequencing data. *Genetic Epidemiology* 32, 179-185.

Maher B. (2008) Personal genomes:. The case of the missing heritability. *Nature* 456, 18-21.

Matsui, S. and Noma, H. (2011a). Estimating e ff ect sizes of di ff erentially expressed genes for power and sample-size assessments in microarray experiments. *Biometrics* 67, 1225-1235.

Matsui, S. and Noma, H. (2011b). Estimation and selection in high-dimensional genomic studies for developing molecular diagnostics. *Biostatistics* 12, 223-233.

Matsui, S., Noma, H., Qu, P., Sakai, Y., Matsui, K., Heuck, C., and Crowley, J. (2018). Multi-subgroup gene screening using semi-parametric hierarchical mixture models and the optimal discovery procedure: Application to a randomized clinical trial in multiple myeloma. *Biometrics (* In Press).

Matsui, S., Simon, R., Qu, P., Shaughnessy, JD Jr, Barlogie, B., and Crowley, J. (2012). Developing and validating continuous genomic signatures in randomized clinical trials for predictive medicine. *Clinical Cancer Research* 18, 6065-6073.

McLachlan, GJ, Bean, RW, and Jones, LBT (2006). A simple implementation of a normal mixture approach to di ff erential gene expression in multiclass microarrays. *Bioinformatics* twenty two, 1608-1615.

Nelson, CP, Goel, A., Butterworth, AS, Kanoni, S., Webb, TR, Marouli, E. et al. (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature Genetics* 49, 1385-1391.

Noma, H. and Matsui, S. (2012). The opitimal discovery procedure in multiple signi fi cance test-ing: an empirical Bayes approach. *Statistics in Medicine* 31, 165-176.

Pe'er, I., Yelensky, R., Altshuler, D., and Daly, MJ (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genetic Epidemiology* 32, 381-385.

Pepe MS. (2005). Evaluating technologies for classi fi cation and prediction in medicine. *Statistics in Medicine* twenty four, 3687-3696.

Risch, N. and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* 273, 1516-1517.

. Robert, CP (2007) The Bayesian Choice, Second Edition New York:. Springer.

Setlur, SR, Mertz, KD, Hoshida, Y., Demichelis, F., Lupien, M., Perner, S. et al. (2008). Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *Journal of the National Cancer Institute* 100, 815-825.

Shao, Y. and Tseng, C. -H. (2007). Sample size calculation with dependence adjustment for FDR-control in microarray studies. *Statistics in Medicine* 26, 4219-4237.

Shen, W. and Louis, TA (1999). Empirical Bayes estimation via the smoothing by roughening approach. *Journal of Computational and Graphical Statistics* 8, 800-823.

Storey, JD (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* 64, 479-498.

Storey, JD (2007) The optimal discovery procedure:. A new approach to simultaneous signi fi- cance testing. *Journal of the Royal Statistical Society, Series B* 69, 347-368.

Tong, T. and Zhao, H. (2008). Practical guidelines for assessing power and false discovery rate for a fi xed sample size in microarray experiments. *Statistics in Medicine* 27, 1960-1972.

Tsai, C.-A., Wang, SJ, Chen, DT, and Chan, JJ (2005). Sample size for gene expression microarray experiments. *Bioinformatics* twenty one, 1502-1508.

Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., and Rothman, N. (2004). Assess- ing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute* 96, 434-442.

Wake fi eld, J. (2009) Bayes factors for genome-wide association studies:. Comparison with P- values. *Genetic epidemiology* 33, 79-86.

Wasserstein RL and Lazar NA (2016) The ASA's statement on p-values:. Context, process, and purpose. *The American Statistician* 70, 129-133

Westfall, PH and Young, SS (1993) Resampling-Based Multiple Testing New York:.. Wiley.