

# Screening for Differentially Expressed Genes: Are Multilevel Models Helpful?

Dongmei Liu<sup>1</sup>, Giovanni Parmigiani<sup>2\*</sup> and Brian Caffo<sup>1</sup>

<sup>1</sup>Department of Biostatistics, Johns Hopkins School of Public Health, USA

<sup>2</sup>Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, USA

## Abstract

Screening for changes in gene expression across biological conditions using high throughput technologies is now common in biology. In this paper we present a broad Bayesian multilevel framework for developing computationally fast shrinkage-based screening tools for this purpose. Our scheme makes it easy to adapt the choice of statistics to the goals of the analysis and to the genomic distributions of signal and noise. We empirically investigate the extent to which these shrinkage-based statistics improve performance, and the situations in which such improvements are larger. Our evaluation uses both extensive simulations and controlled biological experiments. The experimental data include a so-called spike-in experiment, in which the target biological signal is known, and a two-sample experiment, which illustrates the typical conditions in which the methods are applied.

Our results emphasize two important practical concerns that are not receiving sufficient attention in applied work in this area. First, while shrinkage strategies based on multilevel models are able to improve selection performance, they require careful verification of the assumptions on the relationship between signal and noise. Incorrect specification of this relationship can negatively affect a selection procedure. Because this inter-gene relationship is generally identifiable in genomic experiments, we suggest a simple diagnostic plot to assist model checking. Secondly, no statistic performs optimally across two common categories of experimental goals: selecting genes with large changes, and selecting genes with reliably measured changes. Therefore, careful consideration of analysis goals is critical in the choice of the approach taken.

## Background

Many genomics investigations using expression arrays take the form of searching for genes whose expression level is different across experimental conditions or phenotypes. The list of gene transcripts produced by a microarray analysis is usually the starting point for extensive additional biological work, including independent validation, and both in-silico and laboratory work on sequences and proteins related to the transcripts selected. In this context, microarray experiments are screening, not testing, experiments. Because of the wide range of important questions that can be explored using these arrays, and the costs involved, comparisons across conditions are often made using a limited number of replications. Efficient use of data is critical in improving a laboratory's ability to correctly identify important biological hypotheses and proceed to test them by appropriate further experimentation.

Specific screening goals vary with the study. Two simple but representative situations are the selection of genes that are changed by a large amount, and the selection of genes that are changed by a reliably measured amount. In either case, the comparison of gene expression across two conditions based on replicated experiments requires a trade-off of signal, the variation of expression across the two conditions, versus noise, the variation of expression within each condition. Therefore the problem is statistical in nature [1-3]. In this paper we discuss a Bayesian multilevel framework for developing screening tools that adapt to the goals of the analysis and to the genomic distributions of signal and noise. We evaluate a representative set of these tools using both extensive simulations and controlled biological experiments in which the set of altered genes is known.

A variety of approaches for selecting differentially expressed genes have been proposed [4] for a review and Murie et al. [5]. The simplest and still the most widely used is to set a threshold on a measure of signal alone, for example an estimated fold-change. This can be motivated

by the desire to identify large changes, although often it is used by simple analogy with other gene expression essays that have much less noise. Upper and lower thresholds of two and one half are often seen in applications.

One limitation of this approach is that it does not consider how reliably gene-specific changes are measured. That is, it implicitly assumes that all genes are subject to the same level of noise. This may not be the case because even after appropriate preprocessing of the data, the within-gene variation in expression can be highly gene-dependent.

A straightforward way to account for both signal and noise is to select genes based on statistics motivated by two-sample testing, such as the T-ratio or the Wilcoxon statistic. For each gene, the T-ratio is an estimate of the signal-to-noise ratio. Because it requires estimating two or three parameters instead of one, when the number of replicates is small, the T-ratio does not necessarily perform better than fold-change, even when the goal is point-null-like. Gains in efficiency over both fold-change and T-ratios can be obtained by considering the ensemble of gene expression measures at once, rather than each gene in isolation. This occurs for at least two reasons. First, genes measured on the same array type in the same laboratory are all affected by a number

**\*Corresponding author:** Giovanni Parmigiani, Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, 450 Brookline Avenue, Boston MA 02215, USA, Tel: (617) 632-5323; Fax: (617) 632-2444; E-mail: [gp@jimmy.harvard.edu](mailto:gp@jimmy.harvard.edu)

Received March 10 2014; Accepted April 09, 2014; Published April 15, 2014

**Citation:** Liu D, Parmigiani G, Caffo B (2014) Screening for Differentially Expressed Genes: Are Multilevel Models Helpful? J Biomet Biostat 5: 192. doi:10.4172/2155-6180.1000192

**Copyright:** © 2014 Liu D, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

of common sources of noise. Secondly, many changes in expression are part of common biological mechanisms.

A widely used approach that uses genome-wide information is “significance analysis of microarrays” or SAM [6]. SAM involves transforming the signal-to-noise ratios so that they are approximately independent of noise across genes. The type of transformation used by SAM is designed to protect against false discoveries generated by very small denominators in the T-ratios. The denominators that are really small are highly likely to be so by chance, because genes share many sources of variability, and a certain amount of variation is to be expected from all of them.

More broadly, joint estimation of many related quantities is often approached by multilevel modeling, and the associated Empirical Bayesian [7,8] and Hierarchical Bayesian [9] estimation techniques. See Carlin and Louis [10] for a detailed discussion. In genomics, these may represent variation in two stages. The first stage defines summaries at the gene level, for example test statistics, or estimates of fold change and noise. These describe variability of samples within each gene.

The second stage posits a “genomic” distribution for these gene-level summaries. Such multilevel modeling provides tools for borrowing strength from other genes when making inference on each gene. Some examples of implementations in microarrays are provided by Baldi and Long [11], Newton et al. [12], Efron et al. [13], L’onnstedt and Speed et al. [14], Ibrahim et al. [15], Parmigiani et al. [16], Wright and Simon [17] among others.

In practice, a question often raised by genomics practitioners is the extent to which simple, real-time, shrinkage statistics motivated by multilevel models would outperform single-gene-at-a-time analysis or SAM. In this paper we set out to systematically address this question. To this end, we found it necessary to develop a general framework for developing and evaluating these fast shrinkage statistics. Our answer will turn out to be that shrinkage can furnish substantial improvement over single-gene-at-a-time analysis or SAM, provided that the statistics chosen will a) take into account the goals of the screening experiments and b) will be chosen based on examination of the properties of the genomic distribution of signal and noise. Even though we considered an extensive collection of statistics, the goal was not that of providing an exhaustive comparison of all approaches that have been proposed, but rather that of highlighting the critical role of the signal-to-noise trade-off and of providing tools to choose among alternative approaches based on the genome-wide behavior of signal and noise. We compared the performance of these statistics using both extensive simulations and real data sets in which fold changes were known.

## Methods

### Multilevel models for two-group comparisons

We consider a design in which two biological types are compared on a microarray that probes  $G$  genes. Each type is measured on  $n$  arrays using either technical or biological replicates. Here technical replicates refer to experiments that have multiple aliquots of the same RNA, while biological replicates refer to experiments that have multiple subjects from a population. Each situation requires a different interpretation of the array-to-array variability, but the formal structure is the same. We do not consider both levels of replication at the same time here. We denote by  $X_{1gj}$  the expression for gene  $g$  in sample  $j$  in the first group, and by  $X_{2gj}$  the expression for gene  $g$  in sample  $j$  in the second group. Expression levels are assumed to be centered around an overall experiment-wise mean.

Recall our interest lies in studying approaches for selecting genes that are differentially expressed between groups. We begin by describing an additive group effect and independent Gaussian errors. That is we assume that the observed expressions are conditionally independent draws from

$$X_{1gj} | \mu_g, \sigma_g^2, \delta_g \sim N\left(\mu_g - \frac{1}{2}\delta_g, \sigma_g^2\right)$$

$$X_{2gj} | \mu_g, \sigma_g^2, \delta_g \sim N\left(\mu_g + \frac{1}{2}\delta_g, \sigma_g^2\right).$$

Here,  $\delta_g$  is the difference in expression level for gene  $g$  across groups,  $\mu_g$  is an overall expression level for gene  $g$ , also referred to as abundance, or intensity, and  $\sigma_g^2$  is the variance of expression level for gene  $g$  in both groups. We refer to  $\delta_g$  as true signal, and to  $\sigma_g$  as true noise. In a multilevel setting, our parameterization is different from that assuming  $E\{X_{1gj}\} = \mu_g$ ,  $E\{X_{2gj}\} = \mu_g + \delta_g$ , which would lead to two different marginal variances in the two groups.

The fit of the normal distribution can often be improved by a suitable transformation of the data. Departures from normality, equal variances, and additivity may occur but are not considered in this manuscript.

Multilevel models postulate a distribution for the abundance, signal, and noise parameters across genes. A common assumption to many of the multilevel models used in microarray analysis is that of conjugate distributions for the second stage of the statistical model, in short a “conjugate model”. In the case of Gaussian data, the conjugate model implies that the gene-specific signal-to-noise ratios and abundance-to-noise ratios are independent of the corresponding gene-specific noise [18,19]. This assumption leads to convenient mathematical representations for many of the steps required by the data analysis, and is sometimes adopted solely for this reason. In practice, however, some microarray experiments follow this independence pattern closely, while others depart from it substantially. The loss of efficiency of screening based on the conjugate model in the latter case can be large.

Here we broaden the conjugate scheme and we investigate four model varieties, that result from the combination of two factors: (i) whether the gene-specific signal is independent of the gene-specific noise, (ii) whether the gene-specific abundance is independent of the gene-specific noise. Formally, for (i) the independence models assumes that  $\delta_g$  and  $\sigma_g$  are independent, while the conjugate model assumes that  $\delta_g/\sigma_g$  and  $\sigma_g$  are independent.

The remainder of our distributional assumptions are standard for normal multilevel models [9,20]. The models are summarized in Table 1.

We use the notation  $d_g$  for the mean difference of expression across two groups,  $a_g$  for the overall mean expression, and  $s_g$  for the pooled estimate of the standard deviation. Notationally:

$$a_g = \frac{1}{2}(\bar{X}_{1g} + \bar{X}_{2g})$$

$$d_g = \bar{X}_{2g} - \bar{X}_{1g}$$

$$s_g^2 = \frac{\text{RSS}}{n-1} = \frac{1}{n-1} \sum_{j=1}^n (X_{1gj} + \bar{X}_{1g})^2 + \frac{1}{n-1} \sum_{j=1}^n (X_{2gj} - \bar{X}_{2g})^2.$$

In all four models of Table 1, these statistics are independent conditional on gene-specific parameters and have distributions

$$a_g \sim N\left(\mu_g, \frac{1}{2n}\sigma_g^2\right)$$

Data Level	
$X_{1gj}   \mu_g, \sigma_g^2, \delta_g \sim N\left(\mu_g - \frac{1}{2}\delta_g, \sigma_g^2\right)$ $X_{2gj}   \mu_g, \sigma_g^2, \delta_g \sim N\left(\mu_g + \frac{1}{2}\delta_g, \sigma_g^2\right).$	
<b>II. Independence</b>  $\mu_g   \tau^2 \sim N(0, \tau^2)$ $\delta_g   \lambda^2 \sim N(0, \lambda^2)$ $\sigma_g^{-2}   \nu, \beta \sim Ga(\nu, \beta)$	<b>CI. Independence of Signal and Noise</b>  $\mu_g   \tau^2, \sigma_g^2 \sim N(0, \sigma_g^2 \tau^2)$ $\delta_g   \lambda^2 \sim N(0, \lambda^2)$ $\sigma_g^{-2}   \nu, \beta \sim Ga(\nu, \beta)$
<b>IC. Independence of Abundance and Noise</b>  $\mu_g   \tau^2 \sim N(0, \tau^2)$ $\delta_g   \lambda^2, \sigma_g^2 \sim Ga(\nu, \beta)$ $\sigma_g^{-2}   \nu, \beta \sim Ga(\nu, \beta)$	<b>CC. Complete Conjugacy</b>  $\mu_g   \tau^2, \sigma_g^2 \sim N(0, \sigma_g^2 \tau^2)$ $\delta_g   \lambda^2, \sigma_g^2 \sim N(0, \sigma_g^2 \lambda^2)$ $\sigma_g^{-2}   \nu, \beta \sim Ga(\nu, \beta)$

**Table 1:** The four classes of multilevel models investigated. The array-to-array variation is modeled in the same way in all four cases. In all cases,  $j = 1, 2, \dots, n$  and  $g = 1, 2, \dots, G$ . All quantities denoted by Greek letters are unknown. A further set of prior distributions for the hyperparameters is described in the text.

MOTIVATING MODEL	ANALYSIS GOAL	
	Large change	Reliably measured change
Independence/Normality of Genes	Difference in Expression (F)	T-statistic (T)
Exchangeability of Genes	Significance Analysis of Microarray (SAM)	
Complete Conjugacy	Signal (CC.F)	Standardized Signal (CC.T)
	Tail probability (CC.TP)	Bayes factor (CC.BF)
Independence of Abundance and Noise	Signal (CI.F)	Standardized Signal (CI.T)
	Tail probability (CI.TP)	Bayes factor (CI.BF)
Independence of Signal and Noise	Signal (IC.F)	Standardized Signal (IC.T)
	Tail probability (IC.TP)	Bayes factor (IC.BF)
Independence	Signal (II.F)	Standardized Signal (II.T)
	Tail probability (II.TP)	Bayes factor (II.BF)

**Table 2:** Summary of statistics examined, by goal and motivating model structure.

$$d_g \sim N\left(\delta_g, \frac{2}{n}\sigma_g^2\right)$$

$$\frac{n-1}{\sigma^2} s_g^2 \sim \chi_{2(n-1)}^2$$

conditional on gene-specific parameters. All four combinations of Table 1 occur commonly in practice. For example, our two experimental data sets show two markedly different relationships between signal and noise.

The vector of unknown genome-wide parameters will be denoted by  $\xi=(\nu, \beta, \lambda, \tau)$ . There are several estimation approaches available for models of this kind. State-of-the art, computationally intensive approaches are usually based on MCMC [21]. Instead we focus on a faster and simpler empirical Bayesian approach based on estimating  $\xi$  by method of moments from the empirical distributions of  $s_g^2, d_g$ 's. The resulting estimators are computationally cheap and may include shrinkage of the signal, of the noise, or both. Several method of moments alternatives are available, and results can be strongly affected

by this choice. For example, in our experience, the method of moments applied to the distribution of  $s_g^2$ , which is inverse gamma, performs poorly, while the same applied to  $d_g$  performs well.

We approach the task of generating a list of candidate genes by ranking genes according to a one-dimensional statistic, and then selecting all genes whose statistic is above a certain cutoff. This is the norm in practice. While more general decision theoretic approaches evaluating the trade-off between false and missed discoveries are available [22,23], these are complex, and would have been prohibitive in our extensive simulation study. The cutoff is often determined by the ability of a laboratory to perform validity analyses, or, more inferentially, by false discovery rates [24-27]. In our presentation, to simplify the comparison of approaches, we focus on the ranking of genes implied by the statistics, and the ability of each statistic of identifying the top  $g$  genes.

Throughout, we draw a distinction between the selection of genes that are changed by a large amount, and genes that are changed by a reliably measured amount. Accordingly we consider two broad families of statistics, ones that estimate the signal,  $\delta_g$ , and ones that estimate the signal-to-noise ratio,  $\delta_g/\sigma_g$ . Because we use statistics as ranking devices and compare them based on ROC curves, we only need to define statistics up to constants that are not gene-specific. A proportionality sign will indicate omission of such constants. Table 2 summarizes the statistics we examined, organizing them by goal and motivating model structure. Table 3 summarizes the expressions of statistics motivated by multilevel models.

**Statistics**

In this subsection we enumerate and briefly comment on each of the statistics we considered. The remainder of this section is provided as a reference for future sections. Details of the derivations are given in the Appendix.

**Difference in Expression (F):** This is the observed average difference  $d_g$ . Usually expression data are analyzed in the logarithmic scale, in which case F corresponds to an estimate of the log fold change across conditions.

**T-statistic (T):** This is the common statistics  $T \propto d_g/s_g$  used for testing the null hypothesis of  $\delta_g=0$  one gene at the time.

**Significance Analysis of Microarrays (SAM):** This was proposed by Tusher et al. [6] and is based on the change of gene expression relative to an adjusted standard deviation. For the two group case considered here, the SAM statistic for gene  $g$  is

$$SAM = \frac{d_g}{s_g + s_0}$$

where  $s_0$  is the so-called "exchangeability factor". This factor is estimated using information from the entire set of genes to transform the values of SAM so that noise and SAM are approximately independent.

**Statistics for the Complete Conjugacy (CC) Model:** In the Complete Conjugacy model the conditional posterior distribution of  $\delta_g$  given hyperparameters  $\xi$  can be written as

$$\delta_g | d_g, \sigma_g^2, \xi \sim N\left(\frac{\frac{n}{2}d_g}{\frac{n}{2} + \frac{1}{\lambda^2}}, \frac{\sigma_g^2}{\frac{n}{2} + \frac{1}{\lambda^2}}\right) \tag{1}$$

A set of computationally cheap statistics is derived by considering

Model	Statistics			
	F	T	BF	TP
CC.m	$d_g$	$\frac{d_g}{\sqrt{\hat{\sigma}_g^2}}$	$\left( \frac{1 + \frac{1}{\hat{\beta}} \left( \frac{(n-1)}{2} s_g^2 + \frac{n}{4} d_g^2 + \frac{a_g^2}{2n + \tau^2} \right)}{1 + \frac{1}{\hat{\beta}} \left( \frac{(n-1)}{2} s_g^2 + \frac{d_g^2}{n + \hat{\epsilon}^2} + \frac{a_g^2}{2n + \hat{\tau}^2} \right)} \right)^{-(n+\hat{\nu})}$	$\Pr(\delta_g > D   d_g, a_g, s_g^2, \hat{\xi})$ $\delta_g   d_g, a_g, s_g^2, \hat{\xi} \sim St\left(\frac{n}{2} d_g, \left(\frac{n}{2} + \frac{1}{\hat{\lambda}^2}\right) \frac{v_n}{\beta_n}, 2(\hat{\nu} + n)\right)$
CC.c	$d_g$	$\frac{d_g}{\sqrt{\hat{\sigma}_g^2}}$	$e^{-\frac{\frac{n}{2} d_g^2 + \frac{d_g^2}{2} + \hat{\lambda}^2}{\frac{n}{2\hat{\sigma}_g^2}}}$	$\Pr(\delta_g > D   d_g, \hat{\sigma}_g^2, \hat{\xi})$ $\delta_g   d_g, \hat{\sigma}_g^2, \hat{\xi} \sim N\left(\frac{\frac{n}{2} d_g}{\frac{n}{2} + \frac{1}{\hat{\lambda}^2}}, \frac{\hat{\sigma}_g^2}{\frac{n}{2} + \frac{1}{\hat{\lambda}^2}}\right)$
IC	$d_g$	$\frac{d_g}{\sqrt{\hat{\sigma}_g^2}}$	$e^{-\frac{\frac{n}{2} d_g^2 + \frac{d_g^2}{2} + \hat{\lambda}^2}{\frac{n}{2\hat{\sigma}_g^2}}}$	$\Pr(\delta_g > D   d_g, \hat{\sigma}_g^2, \hat{\xi})$ $\delta_g   d_g, \hat{\sigma}_g^2, \hat{\xi} \sim N\left(\frac{\frac{n}{2} d_g}{\frac{n}{2} + \frac{1}{\hat{\lambda}^2}}, \frac{\hat{\sigma}_g^2}{\frac{n}{2} + \frac{1}{\hat{\lambda}^2}}\right)$
CI	$\frac{nd_g}{2\hat{\sigma}_g^2}$ $\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\hat{\epsilon}^2}$	$\frac{nd_g}{2\hat{\sigma}_g^2}$ $\sqrt{\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\hat{\epsilon}^2}}$	$\sqrt{\left(\frac{2}{n} \hat{\sigma}_g^2 + \hat{\lambda}^2\right)^{\frac{\frac{n}{2} d_g^2}{2\hat{\sigma}_g^2} + \frac{d_g^2}{2\left(\frac{2}{n} \hat{\sigma}_g^2 + \hat{\lambda}^2\right)}}} e^{-\frac{\frac{n}{2} d_g^2 + \frac{d_g^2}{2} + \hat{\lambda}^2}{\frac{n}{2\hat{\sigma}_g^2}}}$	$\Pr(\delta_g > D   d_g, \hat{\sigma}_g^2, \hat{\xi})$ $\delta_g   d_g, \hat{\sigma}_g^2, \hat{\xi} \sim N\left(\frac{\frac{nd_g}{2\hat{\sigma}_g^2}}{\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\hat{\lambda}^2}}, \frac{1}{\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\hat{\lambda}^2}}\right)$
II	$\frac{nd_g}{2\hat{\sigma}_g^2}$ $\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\hat{\epsilon}^2}$	$\frac{nd_g}{2\hat{\sigma}_g^2}$ $\sqrt{\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\hat{\epsilon}^2}}$	$\sqrt{\left(\frac{2}{n} \hat{\sigma}_g^2 + \hat{\lambda}^2\right)^{\frac{\frac{n}{2} d_g^2}{2\hat{\sigma}_g^2} + \frac{d_g^2}{2\left(\frac{2}{n} \hat{\sigma}_g^2 + \hat{\lambda}^2\right)}}} e^{-\frac{\frac{n}{2} d_g^2 + \frac{d_g^2}{2} + \hat{\lambda}^2}{\frac{n}{2\hat{\sigma}_g^2}}}$	$\Pr(\delta_g > D   d_g, \hat{\sigma}_g^2, \hat{\xi})$ $\delta_g   d_g, \hat{\sigma}_g^2, \hat{\xi} \sim N\left(\frac{\frac{nd_g}{2\hat{\sigma}_g^2}}{\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\hat{\lambda}^2}}, \frac{1}{\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\hat{\lambda}^2}}\right)$

**Table 3:** Summary of functional forms of all the statistics motivated by multilevel models. For the complete conjugacy case (CC) we consider both statistics that use analytic integration with respect to  $\sigma_g$  (labelled CC.m) and statistics that use plug-in estimates of  $\sigma_g$  (labelled CC.c).

empirical Bayes estimates of this distribution, obtained by replacing hyperparameters  $\xi$  with an estimate  $\hat{\xi}$ . A regularized estimate of signal  $\delta_g$  is

$$CC.F = \frac{\frac{n}{2} d_g}{\frac{1}{\hat{\lambda}^2} + \frac{n}{2}} \propto d_g.$$

Regularization is independent of the gene, so for any given experiment CC.F will be proportional to F. For this reason we only consider F, although we keep this correspondence in mind when interpreting the results.

A standardized estimate of signal is derived as the ratio of the conditional posterior mean and standard deviation of  $\delta_g$  from expression (1),

$$CC.T \propto \frac{d_g}{\sqrt{\hat{\sigma}_g^2}}.$$

The denominator incorporates a linear shrinkage estimate of the gene-specific variance with gene-varying coefficients, penalizing more heavily genes whose signal or abundance are outlying. For this reason, it is critical that the conjugacy assumption be checked, or very valuable information may be lost. On the other hand, when the

assumption is met, an increase in efficiency is gained from estimating the denominator.

An empirical Bayes estimate of the Bayes factor [28] for the null hypothesis of no gene-specific differential expression,  $\delta_g = 0$ , is:

$$CC.BF \propto \frac{\left(1 + \frac{1}{\hat{\beta}} \left( \frac{(n-1)}{2} s_g^2 + \frac{n}{4} d_g^2 + \frac{\frac{a_g^2}{2}}{\frac{1}{2n} + \hat{T}^2} \right)\right)^{-(n+\nu)}}{\left(1 + \frac{1}{\hat{\beta}} \left( \frac{(n-1)}{2} s_g^2 + \frac{\frac{d_g^2}{2}}{n + \hat{\lambda}^2} + \frac{\frac{a_g^2}{2}}{\frac{1}{2n} + \hat{T}^2} \right)\right)^{-(n+\nu)}}$$

Finally, we consider the empirical Bayes approximation,

$$CC.TP = Pr(\delta_g > D | d_g, a_g, s_g^2, \hat{\xi}),$$

of the probability that the true change  $\delta_g$  exceeds D [29]. Here, D represents a target change across conditions. This tail probability reflects the observed change, its variability and the likely magnitude of biologically significant changes.

Statistics for the Independence of Abundance and Noise (IC) Model: In this model the posterior distribution of  $\delta_g$  given  $\sigma_g$  and hyperparameters  $\xi$  can be written as

$$\delta_g | d_g, \sigma_g^2, \xi \sim N \left( \frac{\frac{n}{2} d_g}{\frac{n}{2} + \frac{1}{\lambda^2}}, \frac{\sigma_g^2}{\frac{n}{2} + \frac{1}{\lambda^2}} \right) \quad (2)$$

Unlike in the complete conjugate case, a closed form marginalization with respect to  $\sigma$  is not possible. Therefore we derive results assuming  $\sigma_g$  is known. In the actual calculations, to obtain a real-time statistics,  $\sigma_g^2$  is estimated by the posterior mode of the distribution of  $\sigma_g^2 | s_g^2, \xi$ . Then our estimate of the normalized signal is

$$IC.F = \frac{\frac{n}{2} d_g}{\frac{n}{2} + \frac{1}{\lambda^2}} \propto d_g.$$

As with CC.F, IC.F is proportional to F, so we only consider F in our results section.

A standardized estimate of signal based on regularized estimates of signal is the ratio of the marginal posterior mean and standard deviation of  $\delta_g$  from expression (2), that is

$$IC.T \propto \frac{d_g}{\sqrt{\hat{\sigma}_g^2}}.$$

The Empirical Bayes estimate of the Bayes factor, conditional on gene specific variance is:

$$IC.BF \propto e^{-\frac{\frac{n}{2} d_g^2 + \frac{d_g^2}{2 + \lambda^2}}{2 \hat{\sigma}_g^2}}$$

Finally, we consider the empirical Bayes tail probability

$$IC.TP = Pr(\delta_g > D | d_g, \hat{\sigma}_g^2, \hat{\xi}).$$

Statistics for the Independence of Signal and Noise (CI) Model: In this model the posterior distribution of  $\delta_g$  given  $\sigma_g$  and hyperparameters  $\xi$  can be written as

$$\delta_g | d_g, \sigma_g^2, \xi \sim N \left( \frac{\frac{nd_g}{2\sigma_g^2}}{\frac{n}{2\sigma_g^2} + \frac{1}{\lambda^2}}, \frac{1}{\frac{n}{2\sigma_g^2} + \frac{1}{\lambda^2}} \right) \quad (3)$$

Again, to obtain a real-time statistic we develop results conditional on  $\sigma_g^2$  and estimate it with its posterior mode in actual calculation. The estimate of the signal is

$$CI.F = \frac{\frac{nd_g}{2\hat{\sigma}_g^2}}{\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\lambda^2}}$$

A standardized estimate of signal based on regularized estimates

of signal is the ratio of the marginal posterior mean and standard deviation of  $\delta_g$  from expression (3), that is

$$CI.T = \frac{\frac{nd_g}{2\hat{\sigma}_g^2}}{\sqrt{\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\lambda^2}}}$$

The Empirical Bayes estimate of the Bayes factor is:

$$CI.BF \propto \sqrt{\frac{2}{n} \hat{\sigma}_g^2 + \hat{\lambda}^2}^{-\frac{\frac{n}{2} d_g^2 + \frac{d_g^2}{2(\frac{2}{n} \hat{\sigma}_g^2 + \hat{\lambda}^2)}}{2 \hat{\sigma}_g^2}} e^{-\frac{\frac{n}{2} d_g^2 + \frac{d_g^2}{2(\frac{2}{n} \hat{\sigma}_g^2 + \hat{\lambda}^2)}}{2 \hat{\sigma}_g^2}}$$

while the tail probability approximation is

$$CI.TP = Pr(\delta_g > D | d_g, \hat{\sigma}_g^2, \hat{\xi}).$$

**Statistics for the Complete Independence (II) Model:** In this model the posterior distribution of  $\delta_g$  given  $\sigma_g$  and hyperparameters  $\xi$  can be written again as

$$\delta_g | d_g, \sigma_g^2, \xi \sim N \left( \frac{\frac{nd_g}{2\sigma_g^2}}{\frac{n}{2\sigma_g^2} + \frac{1}{\lambda^2}}, \frac{1}{\frac{n}{2\sigma_g^2} + \frac{1}{\lambda^2}} \right) \quad (4)$$

A regularized estimate of  $\delta_g$ , motivated by the independence model is obtained by replacing  $\xi$  with  $\hat{\xi}$  and  $\sigma_g^2$  with its conditional posterior mode evaluated at  $\hat{\xi}$ , and approximating the posterior mean by

$$II.F = \frac{\frac{nd_g}{2\hat{\sigma}_g^2}}{\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\lambda^2}}$$

Unlike IC.F and CC.F, both II.F and CI.F imply a linear shrinkage which depends on the genomic variability of the signal. Dividing II.F by the square root of the variance of  $\delta_g$ , and approximating as before, we obtain

$$\text{II.T} = \frac{\frac{nd_g}{2\hat{\sigma}_g^2}}{\sqrt{\frac{n}{2\hat{\sigma}_g^2} + \frac{1}{\hat{\lambda}^2}}}$$

The Empirical Bayes estimate of the Bayes factor is:

$$\text{II.BF} \propto \sqrt{\frac{\left(\frac{2}{n}\hat{\sigma}_g^2 + \hat{\lambda}^2\right)^{\frac{n}{2}} \hat{\sigma}_g^2}{\left(\frac{2}{n}\hat{\sigma}_g^2 + \hat{\lambda}^2\right)^{\frac{n}{2}} \hat{\sigma}_g^2}} e^{-\frac{\frac{n}{2}d_g^2}{2\hat{\sigma}_g^2} - \frac{d_g^2}{2\left(\frac{2}{n}\hat{\sigma}_g^2 + \hat{\lambda}^2\right)}}$$

while the empirical Bayes approximation to the tail probability is

$$\text{II.TP} = \Pr(\delta_g > D | d_g, \hat{\sigma}_g^2, \hat{\xi})$$

Notice that the definitions of statistics for the CI and II cases would be the same if  $\sigma_g^2$  was known. Hence the only difference in practice is the posterior mode for  $\sigma_g^2$ . It should not then be a surprise that the performance of these two are very close. For the same reason, this is also true for statistics in the CC and IC settings conditional on  $\sigma_g^2$ .

The empirical Bayes estimators, both standardized and not, have functional similarities to the SAM score, although shrinkage of the noise in the denominators are determined differently. In empirical Bayes analyses, the shrinkage is driven by the parameters of the genomic distributions of signal and noise, in a form that depends on whether or not conjugacy is assumed. In SAM one applies linear shrinkage to the standard deviation rather than the variance, and the shrinkage intercept  $s_0$  is chosen to approximate independence of SAM ratios from noise.

## Simulation Results

### Overview

In this section we study the performance of the real time shrinkage statistics on a large number of data sets simulated from each of the four models in Table 1. We evaluate each statistic on the basis of the implied ranking of genes, and the ability of each statistic of identifying the top  $g$  genes. In the analyses presented here we considered two alternative goals: in one the genes of interest are the top genes by absolute change  $\delta_g$ . In the other the genes of interest are the top genes by signal-to noise ratio  $\delta_g/\sigma_g$ . We considered the top 1%, 2% and 10% for each goal. For each cutoff we create a binary indicator of whether the true parameter is in the top list, and use this indicator as the true class assignment to be predicted. We evaluate performance by an ROC curve [13,30], which is the graph of the true positive fraction versus the false positive fraction for varying thresholds.

As summaries, we consider the overall area under the ROC curve [31] and the partial area under the ROC corresponding to false positive fractions smaller than 20%. We prefer these measures to others incorporating explicitly  $\delta_g$  and  $\delta_g/\sigma_g$  for two reasons: the goal of the microarray experiments we are focusing on is screening rather than estimation; interest usually lies in a relatively small fraction of important findings.

Based on these criteria, our simulations suggest three general conclusions about the alternative approaches for identifying differential genes: i) simple, real-time, shrinkage statistics motivated by multilevel models can outperform alternatives based on analyzing each gene separately, in some cases by a large margin; ii) the same statistics can perform better than the commonly used SAM [6] statistic, provided

that careful checking of the multilevel modeling assumptions is carried out, and iii) no statistics is optimal for both the identification of large  $\delta_g$  and large  $\delta_g/\sigma_g$ .

### Design of simulation study

Our goal was to generate a large and diverse number of scenarios, depending on sample size, conjugacy assumptions, and hyperparameter choices. We considered three sample sizes: 3, 10 and 100 per group. Use of samples as small as 3 is a common scenario in the gene screening experiments taking place during the routine activities of many laboratories, while 10 per group is a common scenario in comparisons across conditions for population genomic studies. Sample size as large as 100 per group are rare and considered here mostly as a check.

For each combination of conjugacy assumption (CC, IC, CI and II) and sample size, we simulated data from 2009 hyperparameter combinations, resulting in a total of 24108 datasets. The 2009 combinations of hyperparameters are based on the grid:

$$E[\sigma_g^{-2}] \in \left(\frac{1}{100}, \frac{1}{25}, \frac{1}{5}, 1, 5, 25, 100\right)$$

$$\text{var}[\sigma_g^{-2}] \in \left(\frac{1}{100}, \frac{1}{25}, \frac{1}{5}, 1, 5, 25, 100\right) E[\sigma_g^{-2}]$$

$$\lambda^2 \in \left(\frac{1}{100}, \frac{1}{25}, \frac{1}{5}, 1, 5, 25, 100\right) E[\sigma_g^{-2}]$$

$$\tau^2 \in \left(\frac{1}{100}, \frac{1}{25}, \frac{1}{5}, 1, 5, 25, 100\right) E[\sigma_g^{-2}]$$

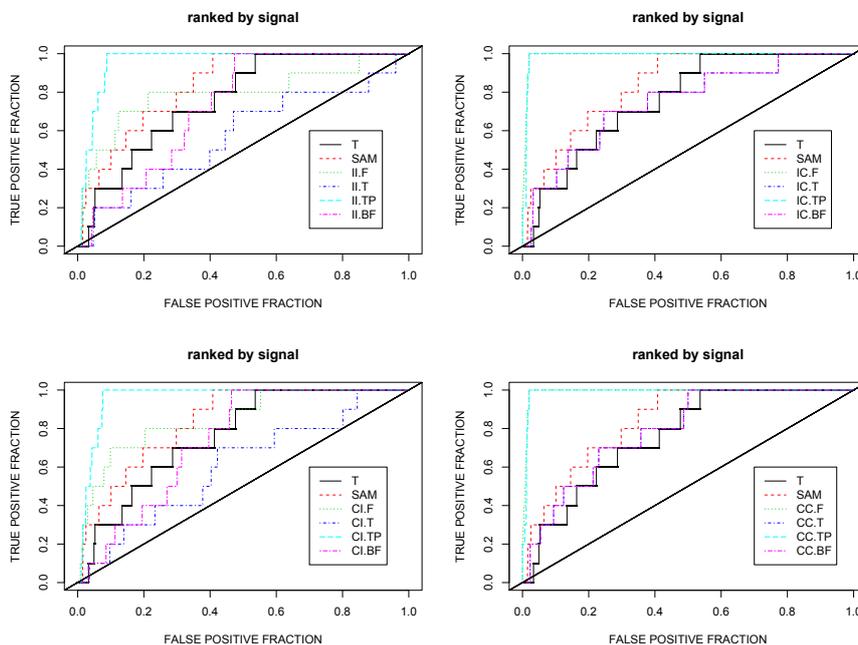
For each combination we derive  $\nu$  and  $\beta$  from  $E[\sigma_g^{-2}]$  and  $\text{var}[\sigma_g^{-2}]$ . Here the total number of combinations is 2009 rather than the full 2401 because some expectation/variance combinations lead to unrealistic settings for  $\nu$  yielding numerically unstable results.

### Simulation results

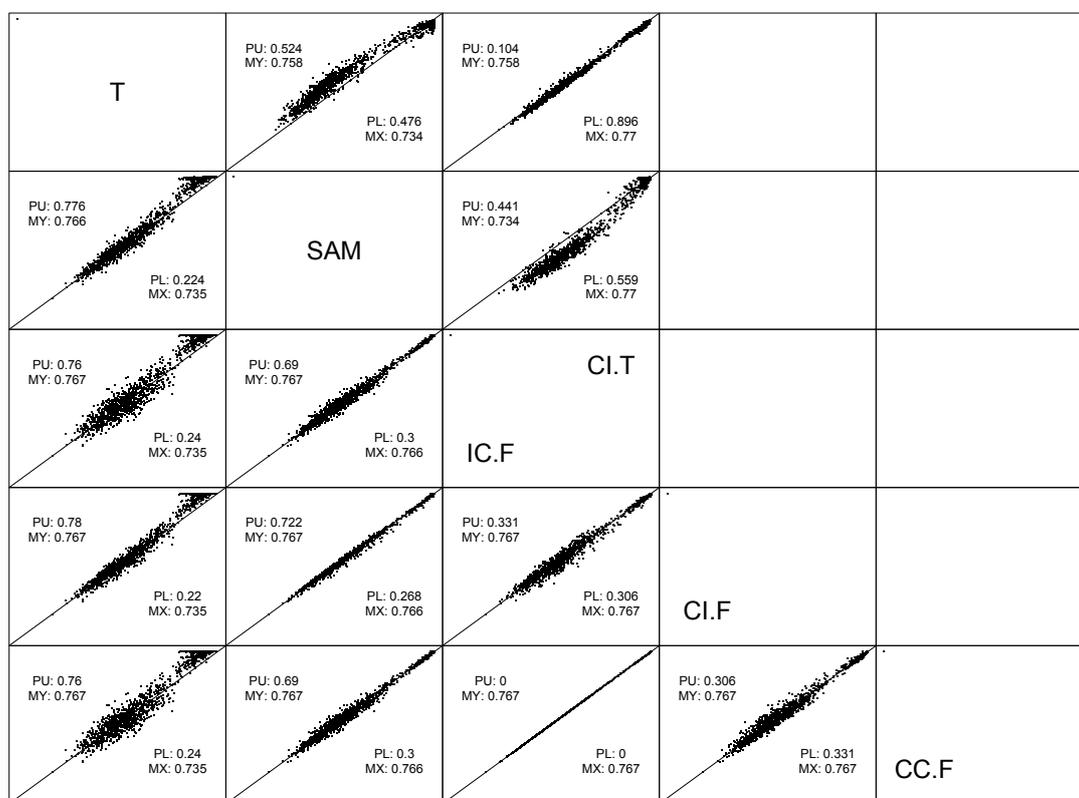
Mining the massive information generated by the thousand datasets required drastic summarization. Here, we present one dataset in detail, and then provide the following summaries: scatterplots in which each dataset/statistic combination is represented by a single point, summaries of pairwise comparisons of statistics by model/goal/sample size, and summaries of best performing statistics by model/goal/sample size.

Figure 1 shows the ROC curves for a single simulation. Data are generated from the II model and the ROC is based on identifying genes with large signal. In this data set, we see a clear separation in the performance of the statistics, both within and across conjugacy structures. The tail probability statistics perform best irrespective of the motivating model, stressing the importance of correctly specifying the analysis goal.

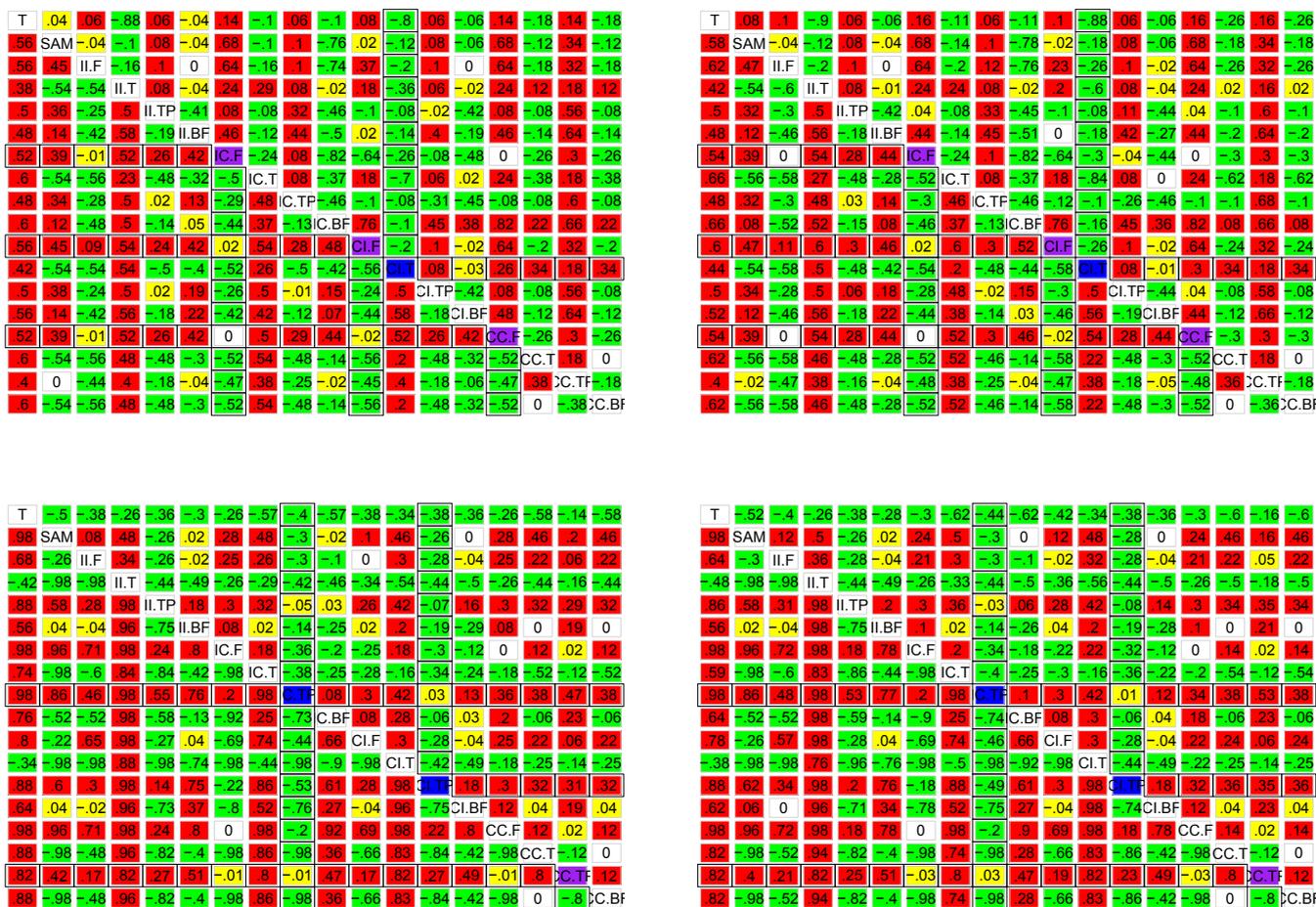
To summarize this type of comparison for all 2009 simulated data sets, we display pair scatterplots of areas under the ROC curves (AUC). Figure 2 summarizes results for II data with three replicates. We focus on T, SAM and on the three shrinkage-based statistics that perform best in II data. Whiskers at the top of the graphs for the T statistics indicate that, for a subset of simulation for which other statistics achieve a perfect separation, the T is can still miss a fraction of differentiated genes. The reciprocal situation does not occur, suggesting a substantial inefficiency in the use of the T statistic. For



**Figure 1:** ROC curves for all the statistics under study in a single simulation with 3 replicates and 1000 genes. The hyperparameter values used to simulate the data were  $v=2$ ,  $\beta=1$ ,  $\gamma=1$  and  $\tau=1$ . The ROC curves are grouped by statistical model to unclutter the displays. The curves for T and SAM are repeated in each panel.



**Figure 2:** Comparison of performance of T, SAM and the shrinkage statistics that perform best in independent (II) data. Each point represents the areas under the curve for two statistics for a particular simulated data set. In the plots above the diagonal, we report the areas when identifying genes with high signal to noise ratio, while in plots below the diagonal we report areas when identifying genes with large signal. The shrinkage statistics that perform best change across the two goals, and therefore different shrinkage statistics are shown above and below the diagonal. Because over-plotting points may lead to visually misleading results in some cases, we also print the percentages PU and PL of points lying above and below the diagonal, and the averages MX and MY of the horizontal and vertical variables.



**Figure 3:** Heat maps of the results for data simulated from the four models, with sample size 3. The color encodes the sign of the difference between percentages PU and PL of points lying above and below the diagonal in Figure 2 for all pairs of statistics. Green indicates PU<PL, red indicates PU>PL, while yellow indicates those cases where the difference was less than 0.05. The best statistics to identify genes with large signal are labeled with purple, while the best statistics to identify reliably measured differentially expressed genes are labeled with blue. These results are further summarized in Table 4.

both goals, the best performing model is based on CI, suggesting that the model chosen to represent the abundance/noise relationship is less critical than that chosen for the signal/noise relationship. A complete set of such comparisons including IC, CI and CC is available in the supplementary materials on our website.

Rather than reproducing these plots for each of the simulation scenarios, we present, in Figure 3, “heat maps” synthesizing pairwise comparisons of estimators for the four simulation settings. For each pair of statistics, maps encode how often one statistics’ AUC is better than the other’s, allowing for rapid comparison of any two statistics on data generated under each of the four model families. To further summarize, the best performing statistics from the heat maps for sample sizes 3, 10 and 100 replicates are reported in Tables 4 and 5 (for partial AUC).

Overall, these results emphasize the importance of matching the statistic to the analysis goal: fold changes and tail probabilities appear to be the best statistics for estimating large signals changes. In contrast the signal-to-noise ratios and Bayes factors appear to be optimal for estimating reliably measured differential expression. Also, these result confirm the importance of shrinkage: in only one case did a T statistic out-perform other statistics to identify high signal changes while in no cases did fold change out-perform other statistics to identify high

SNRs. IN shrinkage correctly identifying the appropriate modelling assumptions becomes increasingly important as the number of replicates increases. The SAM statistic performs well across models, though often worse than the best shrinkage statistics within a model.

To provide a bound on the improvement in performance that can be achieved by shrinkage, in Tables 4 and 5 we provide results obtained by plugging in the true values of the parameters of the genomic distributions instead of their estimates in the calculation of the statistics.

When there is no close form, we have used the conditional estimation by plugging in the posterior mode of  $\sigma_g^2$ . To verify how reasonable these real-time statistics are, we considered the Complete Conjugacy model where closed forms are available for all the statistics. We performed another set of simulation with exactly the same hyperparameters comparing the results based on CC.TP and CC.BF in two cases: one based on the conditional posterior of  $\delta_g$  with posterior mode of  $\sigma_g^2$  plugged in, the other based on the marginal posterior of  $\delta_g$  integrating out  $\sigma_g^2$ . Results are shown in the supplementary materials. Real-time statistics generally perform as well as the exact statistics, although there is a minority of cases in which the exact statistics does far better than real-time approximation, especially for tail probabilities.

Model		II			CI			IC			CC		
top %		1	2	10	1	2	10	1	2	10	1	2	10
MM	S/N	CI.T	CI.T	CI.T	CI.T	CI.T	CI.T	IC.TP CI.TP	IC.TP CI.TP	IC.TP CI.TP	IC.TP CI.TP	IC.TP CI.TP	IC.TP CI.TP
	Signal	CI.F *C.F	CI.F *C.F	CI.TP IC.TP	CI.F *C.F	CI.F *C.F	IC.TP	*C.TP	*C.TP	IC.TP	*C.TP	*C.TP	*C.TP
TT	S/N	CI.T	CI.T T	CI.T T	CI.T T	CI.T T	CI.T T	IC.BF *I.TP	IC.BF	IC.BF CI.TP	IC . B F CI . B F II.TP	IC . B F CI . B F CI.TP	IC . B F CI . B F CI.TP
	Signal	CI.F	CI.F	*I.TP IC.TP	CI.F	*I.F	*I.TP IC.TP	CC.TP	CC.TP	CC.TP	CC.TP	CC.TP	CC.TP

Model		II			CI			IC			CC		
top %		1	2	10	1	2	10	1	2	10	1	2	10
MM	S/N	*I.T	*I.T	II.T	CI.T	CI.T	CI.T	IC.BF *I.TP	IC.BF *I.TP	IC.BF CI.TP	*I.TP	IC.BF *I.TP	IC.BF *I.TP
	Signal	*I.TP CI.F	*I.TP CI.F	CI.F CI.TP IC.TP	II.TP IC.TP	CI.F II.TP IC.TP	*I.TP *F IC.TP	*C.TP	*C.TP	*C.TP	*C.TP	CC.TP	*C.TP
TT	S/N	*I.T	*I.T	II.T	CI.T	CI.T	CI.T	IC.BF *I.TP	IC.BF *I.TP	IC.BF CI.TP	IC.BF *I.TP	IC.BF *I.TP CI.BF	IC.BF CI.TP
	Signal	*I.TP CI.F	*I.TP CI.F	CI.F CI.TP	II.TP IC.TP	*I.TP	*I.TP *F IC.TP	CC.TP	CC.TP	CC.TP	CC.TP	CC.TP	CC.TP

Model		II			CI			IC			CC		
top %		1	2	10	1	2	10	1	2	10	1	2	10
MM	S/N	II.T	II.T IC.BF	II.T	CI.T IC.BF	CI.T IC.BF	CI.T	IC.BF	IC.BF	IC.BF	IC.BF *I.BF	IC.BF	IC.BF
	Signal	*I.F *I.TP	*I.F *I.TP	*I.F *I.TP	*I.TP	*I.TP *F	*I.TP *F	IC.TP	IC.TP	IC.TP	IC.TP	IC.TP	IC.TP
TT	S/N	II.T	II.T IC.BF	II.T	CI.T IC.BF	CI.T IC.BF	CI.T	IC.BF	IC.BF	IC.BF	IC.BF *I.BF	IC.BF *I.BF	IC.BF
	Signal	*I.F *I.TP	*I.F *I.TP	*I.F *I.TP	*I.TP	*I.TP *F	*I.TP *F	IC.TP	IC.TP	IC.TP	IC.TP	IC.TP	IC.TP

**Table 4:** Summary of best performing statistics by AUC. Here “model” is the true model used for simulation. Statistics were evaluated by their ability to identify the top 1%, 2% and 10% of genes with large signals (labeled Signal) and large signal-to-noise ratios (S/N). The rows labeled MM correspond to parameter estimation using the method of moments while those labeled TT correspond to using the true hyperparameter. Instances where the best performing statistic did not match the appropriate true model and goal are high-lighted in red. Instances where the best performing statistic did not match the model but was consistent with goal are highlighted in blue. A wildcard \* indicates the statistic from either the conjugate or independence model was the best performer. For example, “\*C.F” indicates that the CC.F and IC.F statistics were roughly equivalent best performers.

Finally, we further investigated the seemingly counterintuitive result where the best performing statistic for the data simulated from the Complete Conjugacy model are the IC.TP, IC.BF for genes ranked by both signal alone and signal-to-noise ratio. This behavior persists at larger sample sizes. The reason for the counterintuitive behavior is that some of the hyperparameter combinations lead to simulated dataset that have diagnostic plots consistent with an IC model, in which case IC statistic performs well while the abundance-based shrinkage applied by the CC statistics leads to loss of some of the signal. Additional details are provided in the supplementary materials.

## Experimental Results

### Datasets

We analyzed two data sets. The first was reported by Tusher et al. [6] in the context of comparing irradiated and unirradiated cell lines. A subset of the genes’ changes, identified based on the SAM statistic, were

subsequently validated by independent essays. While the experiment includes some blocking, we analyze it here as though it were a two-class comparison with 4 replicates.

The second data set is from an experiment reported by Dudley et al. [32]. They performed a so called “spike-in” experiment in which they selected 9 genes with very low natural expression and “spiked-in” Cy3-labeled gene-specific oligonucleotides in increments from 0.5 fold to 200 fold. Their experiment used cDNA microarrays including a total of 6307 genes, and had two replicates. We work from ratios of Cy3-to-Cy5 channels, after normalization [33]. While spike in experiments are useful in that true fold changes are known, both the magnitudes of the changes, and the sparsity of changes in the genome are unlikely to be realistic.

### Graphical diagnostics for conjugacy structure

We begin by investigating the relationship between signal,

nrep=3													
Model		II			CI			IC			CC		
top %		1	2	10	1	2	10	1	2	10	1	2	10
MM	S/N	CI.T	CI.T	CI.T	CI.T	CI.T	CI.T	SAM	SAM	SAM	SAM	SAM	SAM
	Signal	CI.F	CI.F	CI.F	CI.F	CI.F	CI.F	*C.F	*C.F	*C.F	*C.F	*C.F	*C.F
Truth	S/N	CI.T	CI.T	CI.T T	CI.T	CI.T	CI.T T	CI.T	CI.T	CC.BF CC.T	CI.T	CC.BF CC.T CI.T	CC.BF CC.T
	Signal	*I.F	*I.F	*I.F	CI.F	CI.F	*I.F	*C.F	*C.F	*C.F	*C.F	*C.F	*C.F

nrep=10													
Model		II			CI			IC			CC		
top %		1	2	10	1	2	10	1	2	10	1	2	10
MM	S/N	*I.T	*I.T	II.T	CI.T	CI.T	CI.T	SAM	SAM	SAM	SAM CC.BF CC.T	SAM CC.BF CC.T	SAM CC.BF CC.T
	Signal	CI.F	CI.F	*I.F	CI.F	CI.F	CI.F	*C.F	*C.F CC.TP	*C.F	*C.F	*C.F	*C.F
Truth	S/N	*I.T	*I.T	II.T	CI.T	CI.T	CI.T	CC.T CC.BF CI.T	CC.BF CC.T	CC.BF *C.T	CC.BF CC.T CI.T	CC.BF CC.T CI.T	CC.BF CC.T
	Signal	CI.F	CI.F	*I.F	CI.F	CI.F	CI.F	*C.F	*C.F	*C.F	*C.F	*C.F	*C.F

nrep=100													
Model		II			CI			IC			CC		
top %		1	2	10	1	2	10	1	2	10	1	2	10
MM	S/N	*I.T	II.T	II.T	CI.T	CI.T	CI.T	*C.T CC.BF CI.T	*C.T CC.BF CI.T	*C.T CC.BF	*C.T CC.BF CI.T	*C.T CC.BF	*CC.T CC.BF
	Signal	*I.F	*I.F	*I.F	*I.F	*I.F	*I.F	*C.F	*C.F	*C.F	*C.F	*C.F	*C.F
Truth	S/N	*I.T	II.T	II.T	CI.T	CI.T	CI.T	*C.T	*C.T	*C.T	*C.T	*C.T	CC.T
								CC.BF	CC.BF	CC.BF	CC.BF	CC.BF	CC.BF
	CI.T	CI.T		CI.T									
Signal	*I.F	*I.F	*I.F	*I.F	*I.F	*I.F	*I.F	*C.F	*C.F	*C.F	*C.F	*C.F	*C.F

**Table 5:** Best statistics for each simulation scenario with three replicates. Here “model” corresponds to the true model used for simulation. Statistics were differentiated in their ability to identify the top 1%, 2% and 10% of genes with large signals (labeled Signal) and large signal-to-noise ratios (S/N). The rows labeled “MM” correspond to parameter estimation using the method of moments. Results using the actual true parameter values (labeled “Truth”) are also given. Instances where the best statistic did not matching the appropriate true model and goal are highlighted in red. Instances where the best statistic did not match the model but was consistent with goal are highlighted in blue. The results highlight the importance of matching the statistic to the goal; in only one case did a T statistic out-perform other statistics to identify high signal changes while in no cases did a fold change out-perform other statistics to identify high SNRs.

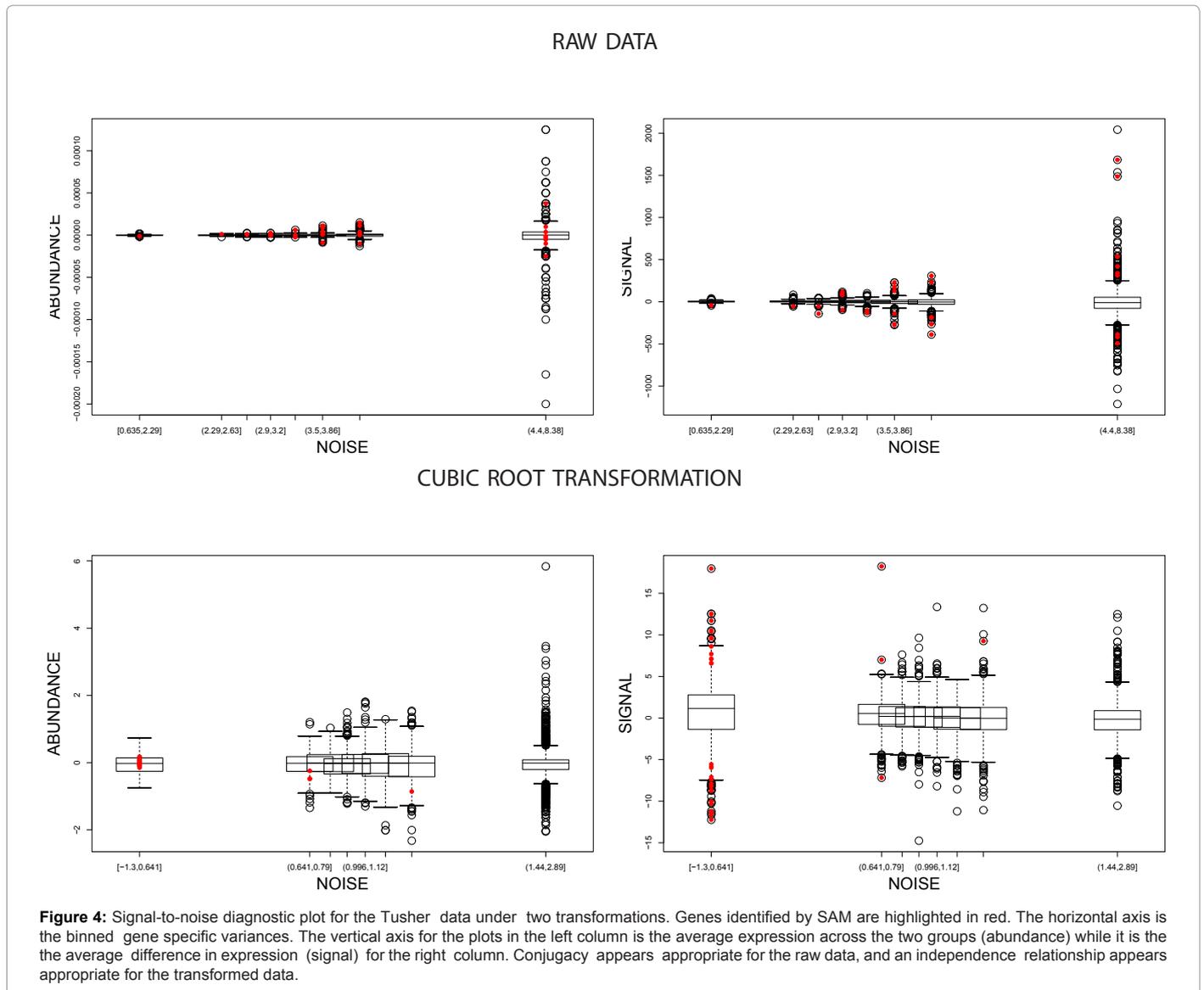
abundance and noise by displaying boxplots of signal and abundance by noise level. These elaborate on ideas of Dudoit et al. [34] and Tusher et al. [6]. Figure 4 considers the Tusher data. An SN plot in which the location and dispersion of signal are stable across noise levels suggests the use of an independence model, while one in which the location and dispersion of signal increase with noise level suggests the use of a conjugate model. Thus, a constant box size indicates independence while an increasing box size indicates conjugacy. In simulated data (see Supplementary materials), the diagnostic plot clearly distinguishes conjugacy with respect to signal and noise as well as conjugacy with respect to abundance and noise.

Because results are sensitive to the type of transformation applied to the expression measurements, we display both the original scale and the cube root. Untransformed data show a pattern consistent with the conjugate model, while data transformed using the cube root appears consistent with the independent model. An alternative visualization to the SN plot is a simple scatterplot of signal versus noise. A limitation

of this approach is that it can be difficult to establish whether increased variation in signal at different level of noise is due to a true relationship or simply to a higher number of genes at that noise level.

Figure 5 shows the spike-in data using three transformations. The original scale shows a marked positive relationship between estimated signal and noise, the cube root scale a mild positive relationship, and the logarithm an almost stable relationship, with some indication of larger variation in the signal at lower noise level. These figures do not inform us about absolute intensity, so the larger variation of signal at the low end after the log transformation is not the same as the well-known “fishtail” effect observed in MVA plots.

In this data set the cube root transformation is the most effective in helping identify the truly differentially expressed genes, performing better than the commonly used log. In practical application one does not have the advantage of knowing the true changes when choosing a transformation. The important lesson here is, however, that choosing



transformations based on convenient statistical properties such as variance stabilization does not necessarily improve, and could prejudice, our ability to detect signal.

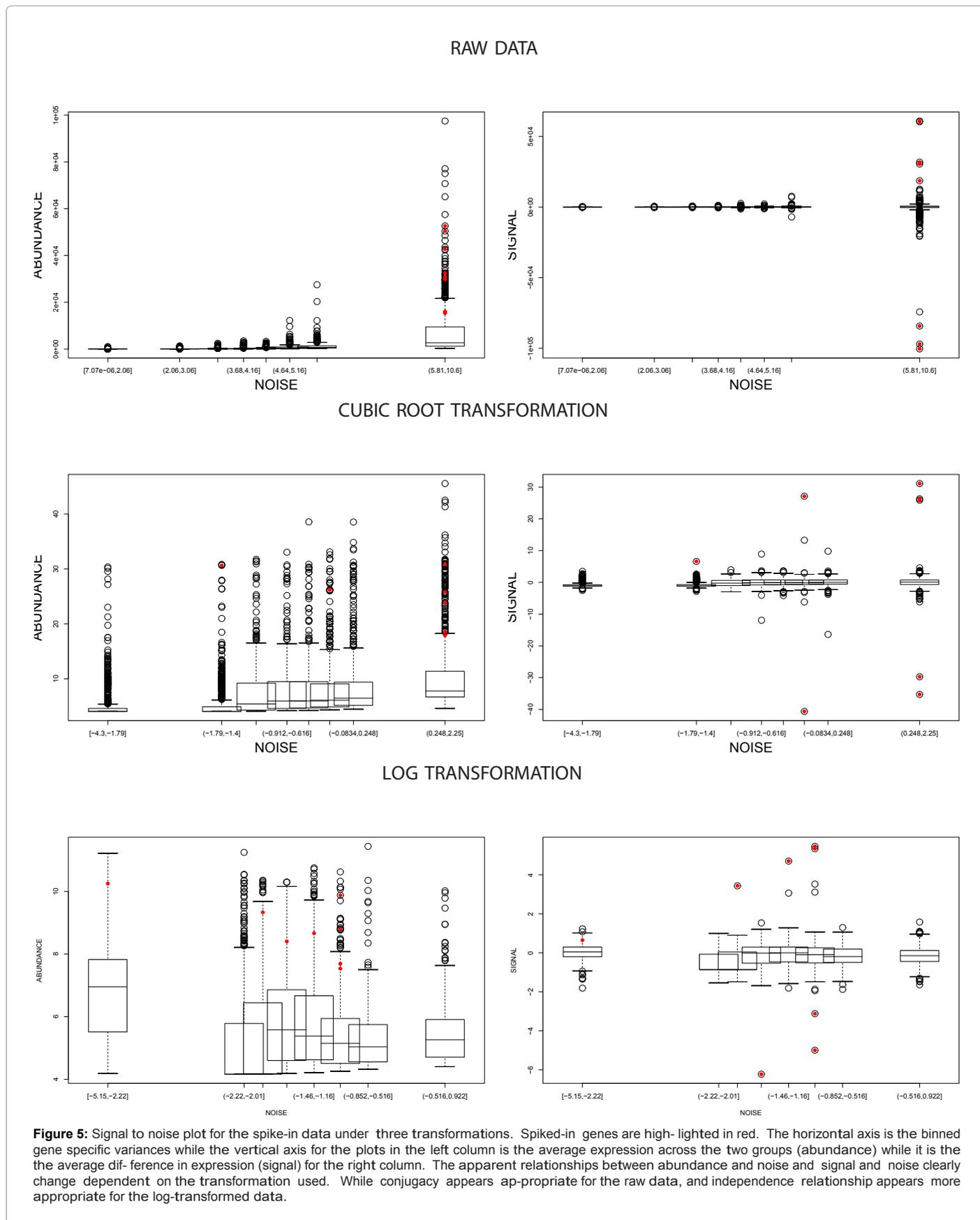
These two datasets stress that both the independence of signal and noise and independence of signal-to-noise ratio and noise may need to be tackled in real applications. While transformation of the measured intensities may allow one to achieve independence, it is not clear that such transformations would be optimal in terms of gene screening.

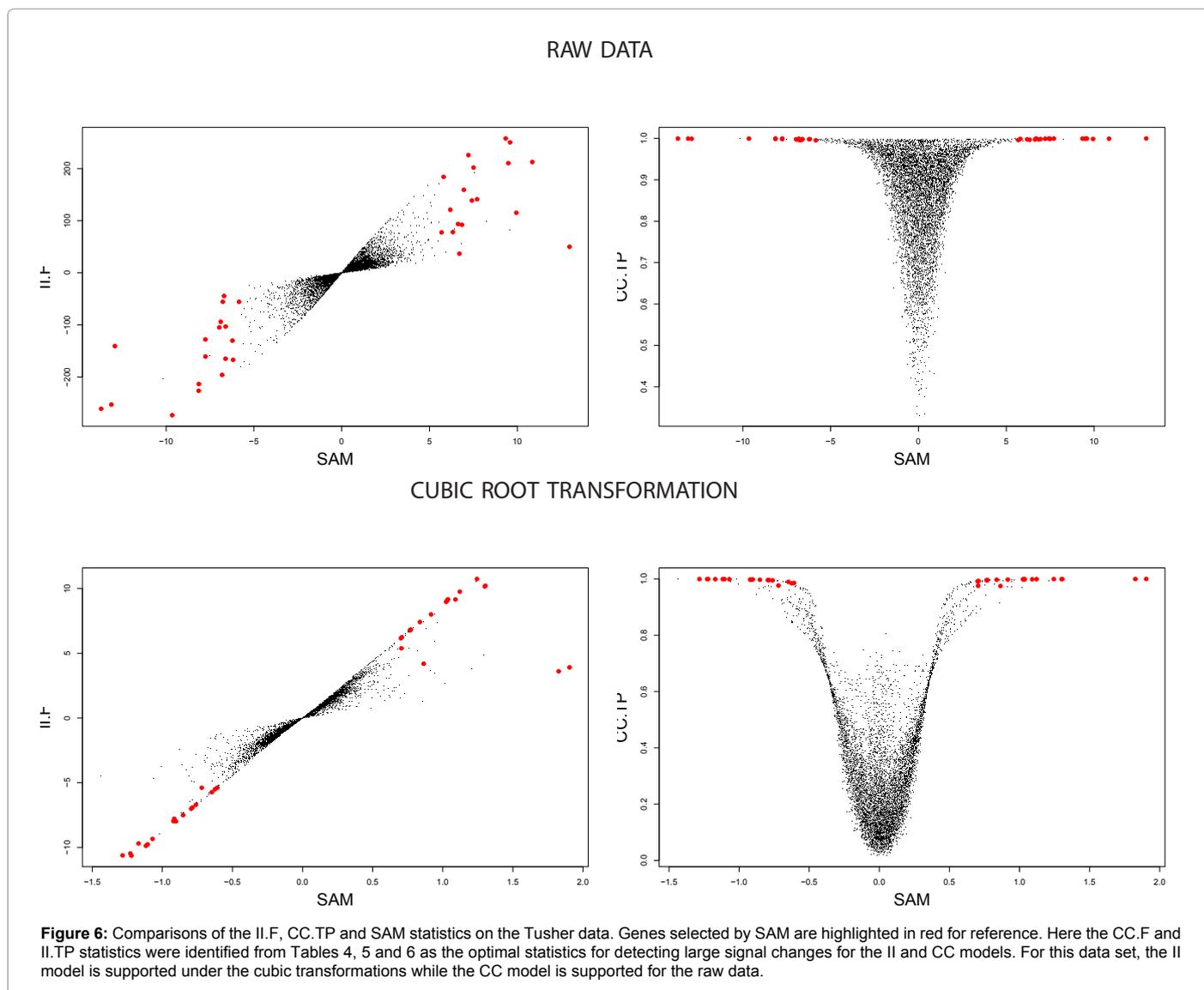
### Comparison of real-time statistics

Figure 6 shows pairwise scatterplots of the statistics CC.TP, II.F, and SAM for the two transformations in Figure 4. CC.TP and II.F are the two best performing real-time shrinkage statistics for identifying genes with large signal. The two best statistics for selecting reliably measured genes, CC.BF and II.T, are given in the supplementary materials. All the genes originally identified by Tusher et al. [6] receive high tail probability using both transformed and untransformed data, though additional genes also receive tail probability close to one. On

the other hand, the correspondence between SAM and II.F is good after cube root transformation but not in the raw scale. In evaluating these results, one must keep in mind that only genes that exceeded a certain SAM threshold were validated independently in the study. Therefore, direct performance comparisons with SAM are not reliable here.

Figure 7 compare statistics in the spike-in data. For the log transformed data, we would expect a better performance from II.F than CC.TP based on the SN plot. In fact, the II.F statistics shrinks the effects excessively and gives a less efficient ranking. For the cube root transformed spike-in data we would expect and, in fact, see a better performance from CC.TP than II.F in Figure 7. For the untransformed spike-in data, results, shown in the supplementary materials, confirm the intuition from the exploratory plots that the conjugacy model should outperform the independence model. Figure 7 does show this result. In a close view of comparison between CC.TP and SAM on raw data (see Supplementary materials), CC.TP also clearly picks up all the spiked genes, while SAM does not. Spiked genes are genes have large





signals, so that poor performance of II.T and CC.BF is no surprise (see Supplementary materials).

This analysis warns of a potential danger in the use of multilevel modeling. Assuming that the signals follow a Gaussian law assumes that all genes are differentially expressed to some extent, and the goal is to either detect the largest signals or the largest reliably measured signals. This is realistic in case-control comparisons and in experiments in which the experimental intervention changes a large portion of the expression, as during cell division. In contrast, in this spike-in experiment, the distribution of signals is in fact degenerate at

0 for all but the 8 spiked in genes. Therefore, the statistics motivated by the Gaussian law on the signal are not validated from the data. While extreme, the spike-in situation may be relevant in practice when experimental intervention modifies a small set of genes involved in a very specialized pathway.

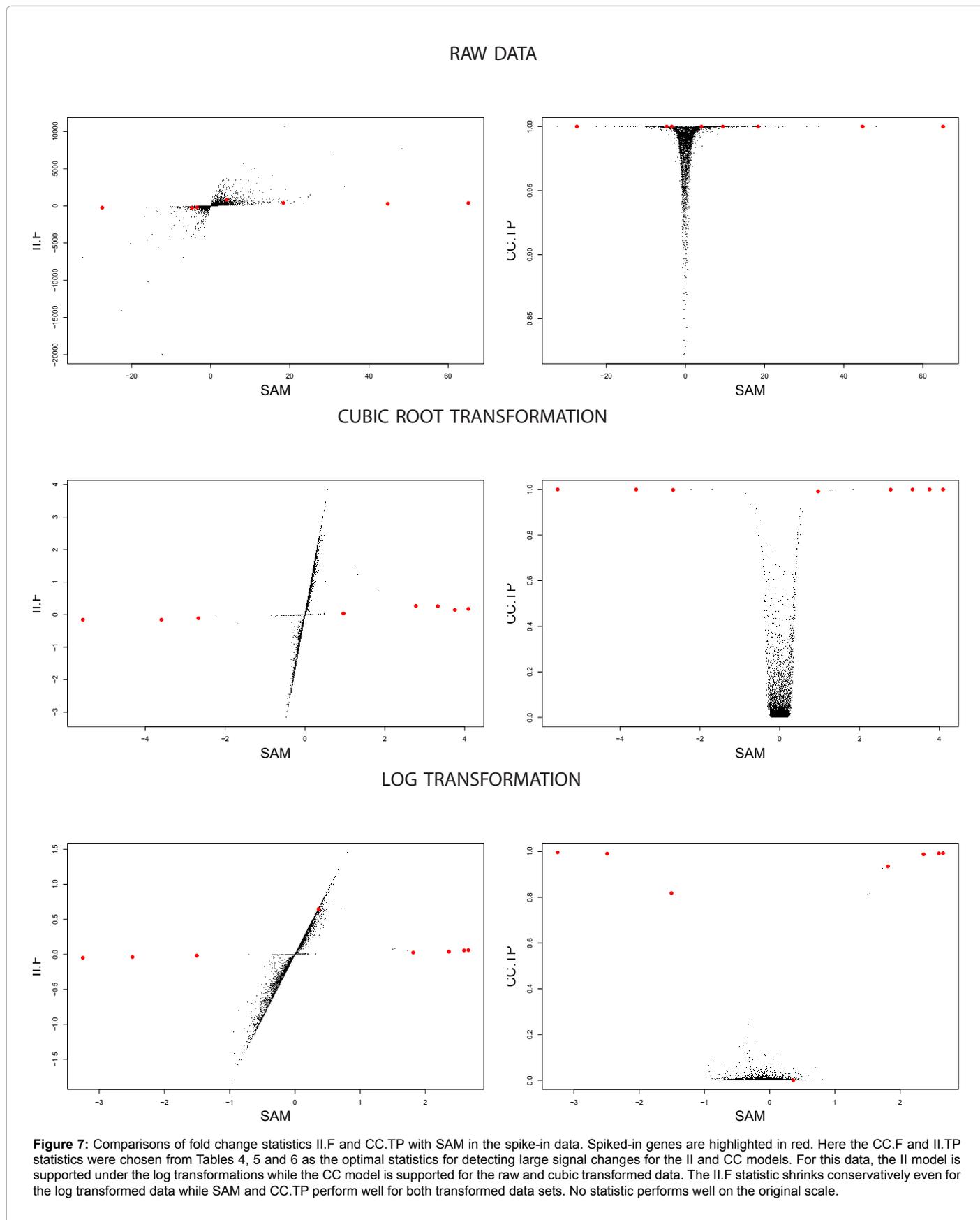
Diagnosing empirically whether the signal distribution is a mixture is difficult. Appropriate weight should be given to the biological circumstances of the experiment. For example, here an independence

relationship is suggested for the signal and noise for the cubic and log transformed data. However, as the majority of the genes are biologically known to have no signal, these plots do not inform us on the question of interest. Furthermore, Figure 7 shows that the best complete independence statistics for identifying large signal and reliably measured signal changes, II.F and II.T, perform poorly for detecting the spiked-in genes. In summary, aggressively modelling the distribution of signals when the overwhelming majority of genes have no signal can produce poor results.

## Conclusions

In this article we present a framework for interpreting, selecting, and estimating shrinkage based screening statistics used in the identification of differentially expressed genes. We also evaluated a representative set of these tools using both extensive simulations and controlled biological experiments in which the set of altered genes is known or partially known.

Our results emphasize two important practical concerns that are not receiving sufficient attention in applied work in this area. First,



while shrinkage strategies based on multilevel models are able to improve selection performance, they require careful verification of the assumptions on the relationship between signal and noise. Incorrect specification of this relationship can negatively affect a selection procedure. Because this inter-gene relationship is generally identified in genomic experiments, we suggest a simple diagnostic plot to assist model checking. Secondly, no statistic performs optimally across two common categories of experimental goals: selecting genes with large changes, and selecting genes with reliably measured changes. Therefore, careful consideration of analysis goals is critical in the choice of the approach taken.

The commonly used SAM statistics emerges as a reasonable compromise between the two goals above and is, to some extent, automatically adaptive to different relationships between signal and noise. Improving on SAM is possible but requires careful validation of the assumptions about the upper level distribution. The assumption of conjugacy in the abundance dimension requires careful attention as it is not robust. In particular, estimators based on the CC assumption can be outperformed even on data generated under CC.

Our simulation analysis relies on the assumed normality of data. In practice, two aspects of it are critical. At the lower stage, in small samples, the functional form of the error distribution across samples is hard to assess. On the other hand, at the upper stage, normality can be checked, and transformations may help, although the caveats discussed in Section 4.3 should be considered. Alternative multilevel models have been studied, for example by Newton and Kendzierski [35] who consider gamma models, and Müller et al. [22], who extend those to mixtures of gamma models. While these alternatives are worth serious consideration, here we focus on Gaussian models and statistics motivated by the Gaussian setting, primarily because in this way we can practically investigate a variety of relevant statistics on a massive number of simulation scenarios.

Other interesting multilevel approaches have been proposed to analyze designs that are more complex than the two-group comparisons considered here. We refer the reader to Kerr et al. [36], Wolfinger et al. [37], Kooperberg et al. [38], Tai and Speed [9] and Wang et al. [40] for further details. Meta-analysis of multiple microarray studies is another area where multilevel models have proven helpful, as illustrated by Conlon [41] and Scharpf et al. [42].

In recent years gene expression is increasingly measured using technologies based on sequencing short read (RNA-seq) which coexist with the hybridization-based approaches that motivated this work. These technologies generate count data. Though the normal model is sometimes used and can perform reasonably after variance stabilization, differential expression is better analyzed statistically using poisson and negative binomial models as described in [43] or Rapaport et al. [44]. An effective effective approaches using multilevel models is included in the Bioconductor package described by Love et al. [45].

In our analysis we assumed that all genes on the array are potentially changed. This is realistic in case control designs across populations or comparison of cells at different stages of the cell cycle, regulation can be expected in a large number of genes, although differences will vary randomly and many genes will be changed by amounts that are smaller than noise. In tightly controlled experiments, such as a comparison of wildtype versus mutant species, or treated versus untreated cell lines, differential expression may only involve a small number of pathways and genes. While this situation can be reasonably handled in the framework considered here, it would be more accurately modeled by

assuming that only a fraction of the genes are differentially expressed across groups A multilevel model could assume assume that a fraction of the  $\delta_g$  are identically zero, while the rest are normally distributed [14].

Our focus here has been on simple and easy-to-compute statistics for gene selection. Multilevel models were used only to provide motivation and a conceptual framework for the derivation of the shrinkage statistics. More generally, multilevel models give raise to potentially more efficient strategies than those considered here, at the price of increased computational expense, for example by MCMC of MCEM. Systematic exploration of those in the thousands of data sets considered here would have been impractical. However, our results suggest that that appropriate shrinkage is a critical part of gene selection and this will hopefully encourage practitioners to consider these more computing intensive approaches as well.

#### Acknowledgment

Data from Dudley, Aach, Steffen and Church (2002) was downloaded from <http://arep.med.harvard.edu/masliner/supplement.htm>. Data from Tusher, Tibshirani and Chu (2001) was downloaded from <http://www-stat-class.stanford.edu/SAM/SAMServlet>. Work of Giovanni Parmigiani supported by grant NCI 5P30 CA006516-49.

#### References

1. Kohane IS, Kho A, Butte AJ (2002) *Microarrays for an Integrative Genomics*, MIT Press, Cambridge, MA, UK.
2. Speed TP (2003) *Statistical Analysis of Gene Expression Microarray Data*, Chapman and Hall, London.
3. Parmigiani G, Garrett ES, Irizarry RA, Zeger SL (2003) *The analysis of gene expression data: an overview of methods and software*, Springer, New York.
4. Pan W (2002) A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* 18: 546-554.
5. Murie C, Woody O, Lee AY, Nadon R (2009) Comparison of small n statistical tests of differential expression applied to microarrays. *BMC Bioinformatics* 10: 45.
6. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98: 5116-5121.
7. Robbins H (1956) An empirical Bayes approach to statistics, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1: 157-163.
8. Efron B, Morris C (1973) Combining possibly related estimation problems (withdiscussion). *Journal of the Royal Statistical Society, Series B, Methodological* 35: 379-421.
9. Lindley DV, Smith AFM (1972) Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society, Series B* 34: 1-41.
10. Bhattacharya S1, Ringo Ho MH, Purkayastha S (2006) A Bayesian approach to modeling dynamic effective connectivity with fMRI data. *Neuroimage* 30: 794-812.
11. Baldi P, Long AD (2001) A Bayesian framework for the analysis of microarrayexpression data: Regularized t-test and statistical inferences of gene changes, *Bioinformatics* 17: 509-519.
12. Newton MA, Kendzierski CM, Richmond CS, Blattner FR, Tsui KW (2001) On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data. *Journal of Computational Biology* 8: 37-52.
13. Efron B, Tibshirani R, Storey JD, Tusher V (2001) Empirical Bayes analysis of a microarray experiment, *Journal of the American Statistical Association* 96: 1151-1160.
14. Lonnstedt I, Speed T (2002) Replicated microarray data, *StatisticaSinica* 12: 3-46.
15. Ibrahim JG, Chen MH, Gray RJ (2002) Bayesian models for gene expression

- with DNA microarray data. *Journal of the American Statistical Association* 97: 88-99.
16. Parmigiani G, Garrett ES, Anbazhagan R, Gabrielson E (2002) A statistical framework for expression-based molecular classification in cancer. *Journal of the Royal Statistical Society Series B* 64: 717-736.
  17. Wright GW1, Simon RM (2003) A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* 19: 2448-2455.
  18. Raiffa H, Schlaifer R (1961) *Applied Statistical Decision Theory*, Harvard University Press, Boston.
  19. Ando A, Kaufman GM (1965) Bayesian analysis of the independent multinormal process—Neither mean nor precision known. *Journal of the American Statistical Association* 60: 347-358.
  20. Gelman A, Carlin J, Stern H, Rubin D (1995) *Bayesian Data Analysis*, Chapman and Hall, London.
  21. Gilks WR, Richardson S, Spiegelhalter DJ (1996) *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.
  22. Muller P, Parmigiani G, Robert C, Rousseau J (2005) Optimal sample size for multiple testing: the case of gene expression microarrays, *Journal of the American Statistical Association* 99: 990-1001.
  23. Lin R, Louis TA, Paddock SM, Ridgeway G (2003) Loss function based ranking in two-stage, hierarchical models; Technical report 2003-6, Department of Biostatistics, Johns Hopkins University, USA.
  24. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing, *J R Statist Soc B* 57: 289-300.
  25. Genovese C, Wasserman L (2001) Bayesian and frequentist multiple testing, Discussion paper, Department of Statistics, Carnegie Mellon University, USA.
  26. Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* 64: 479-498.
  27. Storey JS, Tibshirani R (2003) SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In: Parmigiani G, Garrett EG, Irizarry R and Zeger SL (eds). *The analysis of gene expression data: methods and software*, Springer, New York.
  28. Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association* 90: 773-795.
  29. Normand SLT, Glickman ME, Gatsonis CA (1997) Statistical methods for profiling providers of medical care: Issues and applications. *Journal of the American Statistical Association* 92: 803-814.
  30. National Research Council; Panel on Discriminant Analysis Classification and Clustering (1988). *Discriminant Analysis and Clustering*, National Academy Press, Washington, DC.
  31. Pepe MS, Longton G, Anderson GL, Schummer M (2003) Selecting differentially expressed genes from microarray experiments. *Biometrics* 59: 133-142.
  32. Dudley AM, Aach J, Steffen MA, Church GM (2002) Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *Proc Natl Acad Sci USA* 99: 7554-7559.
  33. Dudoit S, Yang J (2003) Bioconductor R packages for exploratory analysis and normalization of cDNA microarray data. In: Parmigiani G, Garrett E, Irizarry R, Zeger S (eds). *The Analysis of Gene Expression Data: Methods and Software*, Springer Verlag, New York.
  34. Dudoit S, Yang YH, Callow MJ, Speed TP (2002) Statistical methods for identifying genes with differential expression in replicated cDNA microarray experiments. *Statistica Sinica* 12: 111-139.
  35. Newton MA and Kendziorski CM (2003) *Parametric empirical bayes methods for microarrays*, The analysis of gene expression data: methods and software, Springer, New York.
  36. Kerr MK, Martin M, Churchill GA (2000) Analysis of variance for gene expression microarray data. *J Comput Biol* 7: 819-837.
  37. Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, et al. (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol* 8: 625-637.
  38. Kooperberg C, Sipione S, LeBlanc M, Strand AD, Cattaneo E, et al. (2002) Evaluating test statistics to select interesting genes in microarray experiments. *Hum Mol Genet* 11: 2223-2232.
  39. Tai YC, Speed TP (2009) On gene ranking using replicated microarray time course data. *Biometrics* 65: 40-51.
  40. Wang XV, Verhaak RG, Purdom E, Spellman PT, Speed TP (2011) Unifying gene expression measures from multiple platforms using factor analysis. *PLoS One* 6: e17691.
  41. Conlon EM (2008) A Bayesian mixture model for metaanalysis of microarray studies. *Funct Integr Genomics* 8: 43-53.
  42. Scharpf R, Tjelmeland H, Parmigiani G, Nobel AB (2009) A Bayesian model for cross-study differential gene expression. *J Am Stat Assoc* 104: 1295-1310.
  43. Kvam VM, Liu P, Si Y (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am J Bot* 99: 248-256.
  44. Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, et al. (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* 14: R95.
  45. Love M, Anders S, Huber W (2014) Differential analysis of count data—the DESeq2 package 1-41.

# 1 List of figures

## 1.1 Summary of best statistics

### 1. Number of replicates equal to 3

#### (a) Independence model

##### i. Top 1% genes are differential genes

###### A. Hyperparameters estimated by method of moment

- Figure 1 Heat map, the summary of pairwise comparison among all statistics
- Figure 2 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model

###### B. True Hyperparameters

- Figure 3 Heat map, the summary of pairwise comparison among all statistics
- Figure 4 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model

##### ii. Top 2% genes are differential genes

###### A. Hyperparameters estimated by method of moment

- Figure 5 Heat map, the summary of pairwise comparison among all statistics
- Figure 6 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model

###### B. True Hyperparameters

- Figure 7 Heat map, the summary of pairwise comparison among all statistics
- Figure 8 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model

##### iii. Top 10% genes are differential genes

###### A. Hyperparameters estimated by method of moment

- Figure 9 Heat map, the summary of pairwise comparison among all statistics
- Figure 10 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model

###### B. True Hyperparameters

- Figure 11 Heat map, the summary of pairwise comparison among all statistics
- Figure 12 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model

#### (b) Independence of Signal and Noise model

##### i. Top 1% genes are differential genes

###### A. Hyperparameters estimated by method of moment

- Figure 13 Heat map, the summary of pairwise comparison among all statistics
- Figure 14 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model

###### B. True Hyperparameters

- Figure 15 Heat map, the summary of pairwise comparison among all statistics
- Figure 16 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model

##### ii. Top 2% genes are differential genes

###### A. Hyperparameters estimated by method of moment

- Figure 17 Heat map, the summary of pairwise comparison among all statistics
- Figure 18 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model

- B. True Hyperparameters
  - Figure 19 Heat map, the summary of pairwise comparison among all statistics
  - Figure 20 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- iii. Top 10% genes are differential genes
  - A. Hyperparameters estimated by method of moment
    - Figure 21 Heat map, the summary of pairwise comparison among all statistics
    - Figure 22 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - B. True Hyperparameters
    - Figure 23 Heat map, the summary of pairwise comparison among all statistics
    - Figure 24 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- (c) Independence of Abundance and Noise model
  - i. Top 1% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 25 Heat map, the summary of pairwise comparison among all statistics
      - Figure 26 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 27 Heat map, the summary of pairwise comparison among all statistics
      - Figure 28 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - ii. Top 2% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 29 Heat map, the summary of pairwise comparison among all statistics
      - Figure 30 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 31 Heat map, the summary of pairwise comparison among all statistics
      - Figure 32 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - iii. Top 10% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 33 Heat map, the summary of pairwise comparison among all statistics
      - Figure 34 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 35 Heat map, the summary of pairwise comparison among all statistics
      - Figure 36 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- (d) Complete Conjugacy model
  - i. Top 1% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 37 Heat map, the summary of pairwise comparison among all statistics
      - Figure 38 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters

- Figure 39 Heat map, the summary of pairwise comparison among all statistics
  - Figure 40 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - ii. Top 2% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 41 Heat map, the summary of pairwise comparison among all statistics
      - Figure 42 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 43 Heat map, the summary of pairwise comparison among all statistics
      - Figure 44 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - iii. Top 10% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 45 Heat map, the summary of pairwise comparison among all statistics
      - Figure 46 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 47 Heat map, the summary of pairwise comparison among all statistics
      - Figure 48 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
2. Number of replicates equal to 10
- (a) Independence model
- i. Top 1% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 49 Heat map, the summary of pairwise comparison among all statistics
      - Figure 50 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 51 Heat map, the summary of pairwise comparison among all statistics
      - Figure 52 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - ii. Top 2% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 53 Heat map, the summary of pairwise comparison among all statistics
      - Figure 54 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 55 Heat map, the summary of pairwise comparison among all statistics
      - Figure 56 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - iii. Top 10% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 57 Heat map, the summary of pairwise comparison among all statistics
      - Figure 58 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters

- Figure 59 Heat map, the summary of pairwise comparison among all statistics
  - Figure 60 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- (b) Independence of Signal and Noise model
- i. Top 1% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 61 Heat map, the summary of pairwise comparison among all statistics
      - Figure 62 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 63 Heat map, the summary of pairwise comparison among all statistics
      - Figure 64 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - ii. Top 2% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 65 Heat map, the summary of pairwise comparison among all statistics
      - Figure 66 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 67 Heat map, the summary of pairwise comparison among all statistics
      - Figure 68 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - iii. Top 10% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 69 Heat map, the summary of pairwise comparison among all statistics
      - Figure 70 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 71 Heat map, the summary of pairwise comparison among all statistics
      - Figure 72 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- (c) Independence of Abundance and Noise model
- i. Top 1% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 73 Heat map, the summary of pairwise comparison among all statistics
      - Figure 74 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 75 Heat map, the summary of pairwise comparison among all statistics
      - Figure 76 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - ii. Top 2% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 77 Heat map, the summary of pairwise comparison among all statistics
      - Figure 78 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 79 Heat map, the summary of pairwise comparison among all statistics

- Figure 80 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- iii. Top 10% genes are differential genes
  - A. Hyperparameters estimated by method of moment
    - Figure 81 Heat map, the summary of pairwise comparison among all statistics
    - Figure 82 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - B. True Hyperparameters
    - Figure 83 Heat map, the summary of pairwise comparison among all statistics
    - Figure 84 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- (d) Complete Conjugacy model
  - i. Top 1% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 85 Heat map, the summary of pairwise comparison among all statistics
      - Figure 86 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 87 Heat map, the summary of pairwise comparison among all statistics
      - Figure 88 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - ii. Top 2% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 89 Heat map, the summary of pairwise comparison among all statistics
      - Figure 90 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 91 Heat map, the summary of pairwise comparison among all statistics
      - Figure 92 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - iii. Top 10% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 93 Heat map, the summary of pairwise comparison among all statistics
      - Figure 94 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 95 Heat map, the summary of pairwise comparison among all statistics
      - Figure 96 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- 3. Number of replicates equal to 100
  - (a) Independence model
    - i. Top 1% genes are differential genes
      - A. Hyperparameters estimated by method of moment
        - Figure 97 Heat map, the summary of pairwise comparison among all statistics
        - Figure 98 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
      - B. True Hyperparameters

- Figure 99 Heat map, the summary of pairwise comparison among all statistics
  - Figure 100 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- ii. Top 2% genes are differential genes
- A. Hyperparameters estimated by method of moment
- Figure 101 Heat map, the summary of pairwise comparison among all statistics
  - Figure 102 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- B. True Hyperparameters
- Figure 103 Heat map, the summary of pairwise comparison among all statistics
  - Figure 104 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- iii. Top 10% genes are differential genes
- A. Hyperparameters estimated by method of moment
- Figure 105 Heat map, the summary of pairwise comparison among all statistics
  - Figure 106 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- B. True Hyperparameters
- Figure 107 Heat map, the summary of pairwise comparison among all statistics
  - Figure 108 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- (b) Independence of Signal and Noise model
- i. Top 1% genes are differential genes
- A. Hyperparameters estimated by method of moment
- Figure 109 Heat map, the summary of pairwise comparison among all statistics
  - Figure 110 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- B. True Hyperparameters
- Figure 111 Heat map, the summary of pairwise comparison among all statistics
  - Figure 112 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- ii. Top 2% genes are differential genes
- A. Hyperparameters estimated by method of moment
- Figure 113 Heat map, the summary of pairwise comparison among all statistics
  - Figure 114 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- B. True Hyperparameters
- Figure 115 Heat map, the summary of pairwise comparison among all statistics
  - Figure 116 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- iii. Top 10% genes are differential genes
- A. Hyperparameters estimated by method of moment
- Figure 117 Heat map, the summary of pairwise comparison among all statistics
  - Figure 118 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- B. True Hyperparameters
- Figure 119 Heat map, the summary of pairwise comparison among all statistics

- Figure 120 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- (c) Independence of Abundance and Noise model
- i. Top 1% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 121 Heat map, the summary of pairwise comparison among all statistics
      - Figure 122 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 123 Heat map, the summary of pairwise comparison among all statistics
      - Figure 124 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - ii. Top 2% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 125 Heat map, the summary of pairwise comparison among all statistics
      - Figure 126 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 127 Heat map, the summary of pairwise comparison among all statistics
      - Figure 128 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - iii. Top 10% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 129 Heat map, the summary of pairwise comparison among all statistics
      - Figure 130 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 131 Heat map, the summary of pairwise comparison among all statistics
      - Figure 132 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- (d) Complete Conjugacy model
- i. Top 1% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 133 Heat map, the summary of pairwise comparison among all statistics
      - Figure 134 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 135 Heat map, the summary of pairwise comparison among all statistics
      - Figure 136 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - ii. Top 2% genes are differential genes
    - A. Hyperparameters estimated by method of moment
      - Figure 137 Heat map, the summary of pairwise comparison among all statistics
      - Figure 138 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
    - B. True Hyperparameters
      - Figure 139 Heat map, the summary of pairwise comparison among all statistics

- Figure 140 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
- iii. Top 10% genes are differential genes
  - A. Hyperparameters estimated by method of moment
    - Figure 141 Heat map, the summary of pairwise comparison among all statistics
    - Figure 142 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model
  - B. True Hyperparameters
    - Figure 143 Heat map, the summary of pairwise comparison among all statistics
    - Figure 144 Pairwise comparison of T statistics, SAM and Best statistics motivated by multilevel model

## 1.2 Compare the difference between two definitions of CC.TP and CC.BF

1. Number of replicates equal to 3
  - (a) Figure 145 Top 1% genes are differential genes
  - (b) Figure 146 Top 2% genes are differential genes
  - (c) Figure 147 Top 10% genes are differential genes
2. Number of replicates equal to 10
  - (a) Figure 148 Top 1% genes are differential genes
  - (b) Figure 149 Top 2% genes are differential genes
  - (c) Figure 150 Top 10% genes are differential genes
3. Number of replicates equal to 100
  - (a) Figure 151 Top 1% genes are differential genes
  - (b) Figure 152 Top 2% genes are differential genes
  - (c) Figure 153 Top 10% genes are differential genes

## 1.3 Diagnostic plot of counter intuitive result from CC model

1. Compare CC.TP to IC.TP, CC.BF to IC.BF when data are simulated from CC model
  - (a) Number of replicates equal to 3
    - i. Figure 154 Top 1% genes are differential genes
    - ii. Figure 155 Top 2% genes are differential genes
    - iii. Figure 156 Top 10% genes are differential genes
  - (b) Number of replicates equal to 10
    - i. Figure 157 Top 1% genes are differential genes
    - ii. Figure 158 Top 2% genes are differential genes
    - iii. Figure 159 Top 10% genes are differential genes
  - (c) Number of replicates equal to 100
    - i. Figure 160 Top 1% genes are differential genes
    - ii. Figure 161 Top 2% genes are differential genes
    - iii. Figure 162 Top 10% genes are differential genes
2. Diagnostic plot of data simulated from CC model
  - (a) Figure 163 IC.TP does better than CC.TP
  - (b) Figure 164 CC.TP does better than IC.TP
3. Diagnostic plot of data simulated from IC model
  - (a) Figure 165 IC.TP does better than CC.TP
  - (b) Figure 166 CC.TP does better than IC.TP

## 2 Figures

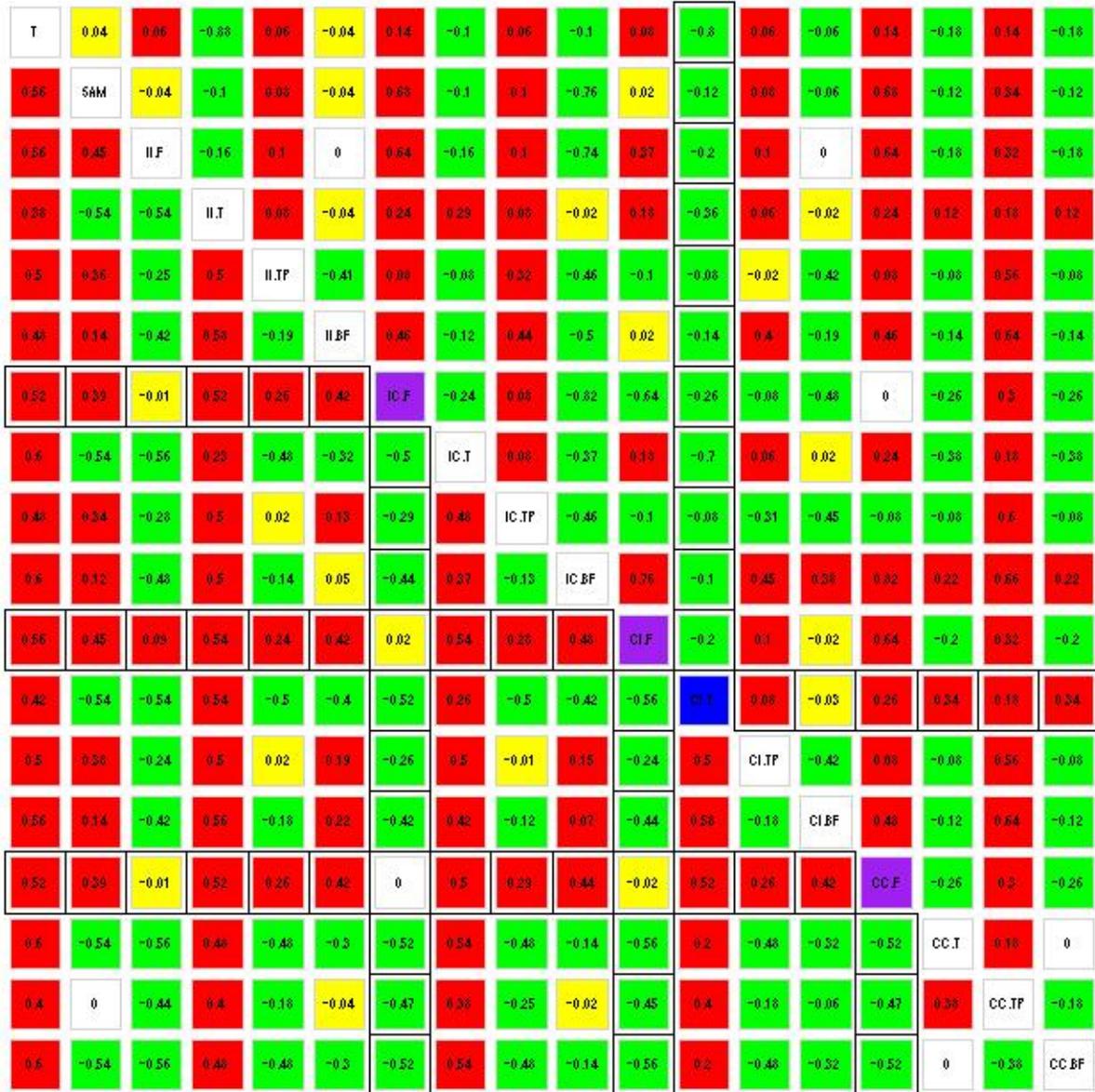


Figure 1: Heat map summary for simulations based on Independence model with number of replicates equal to 3 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

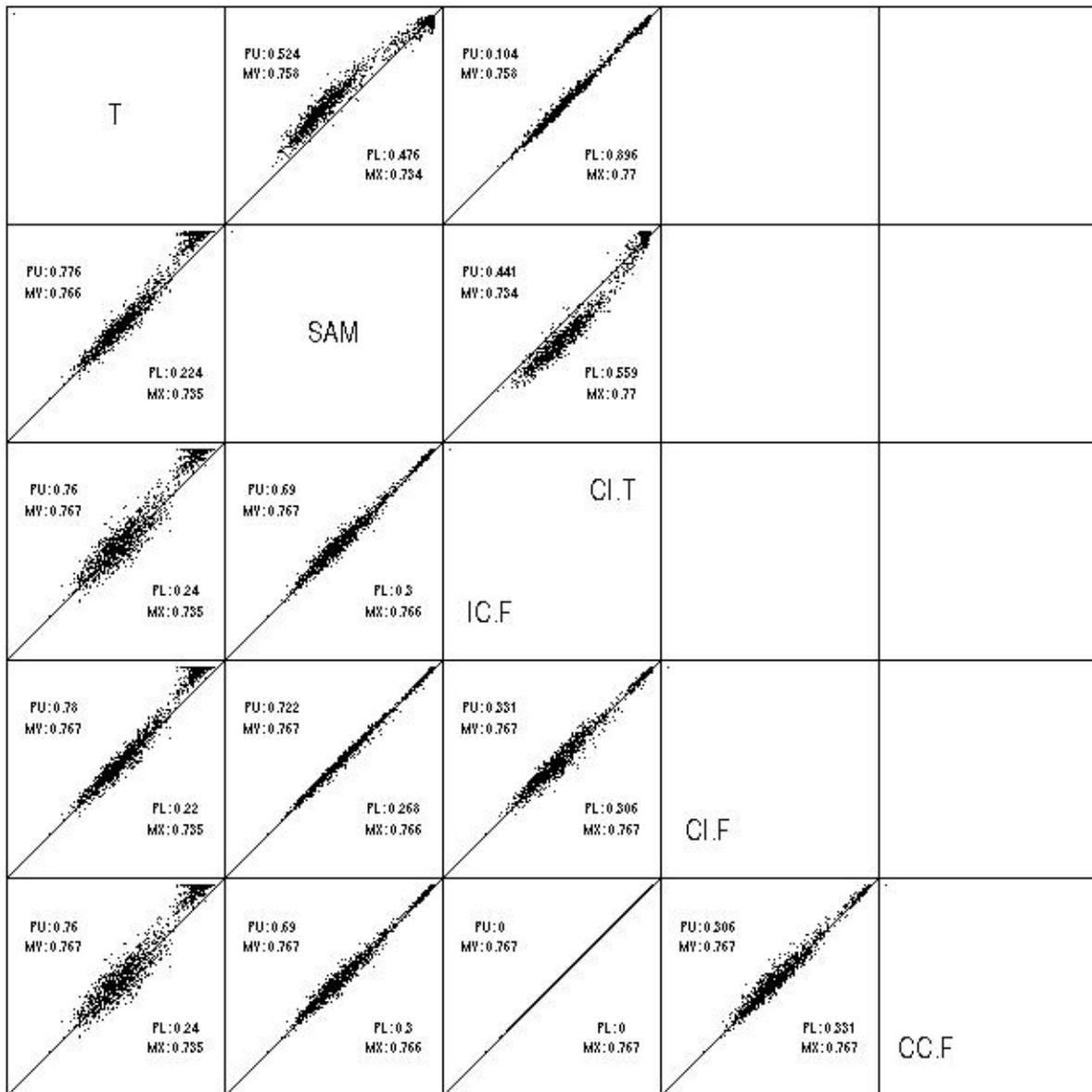


Figure 2: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 3 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

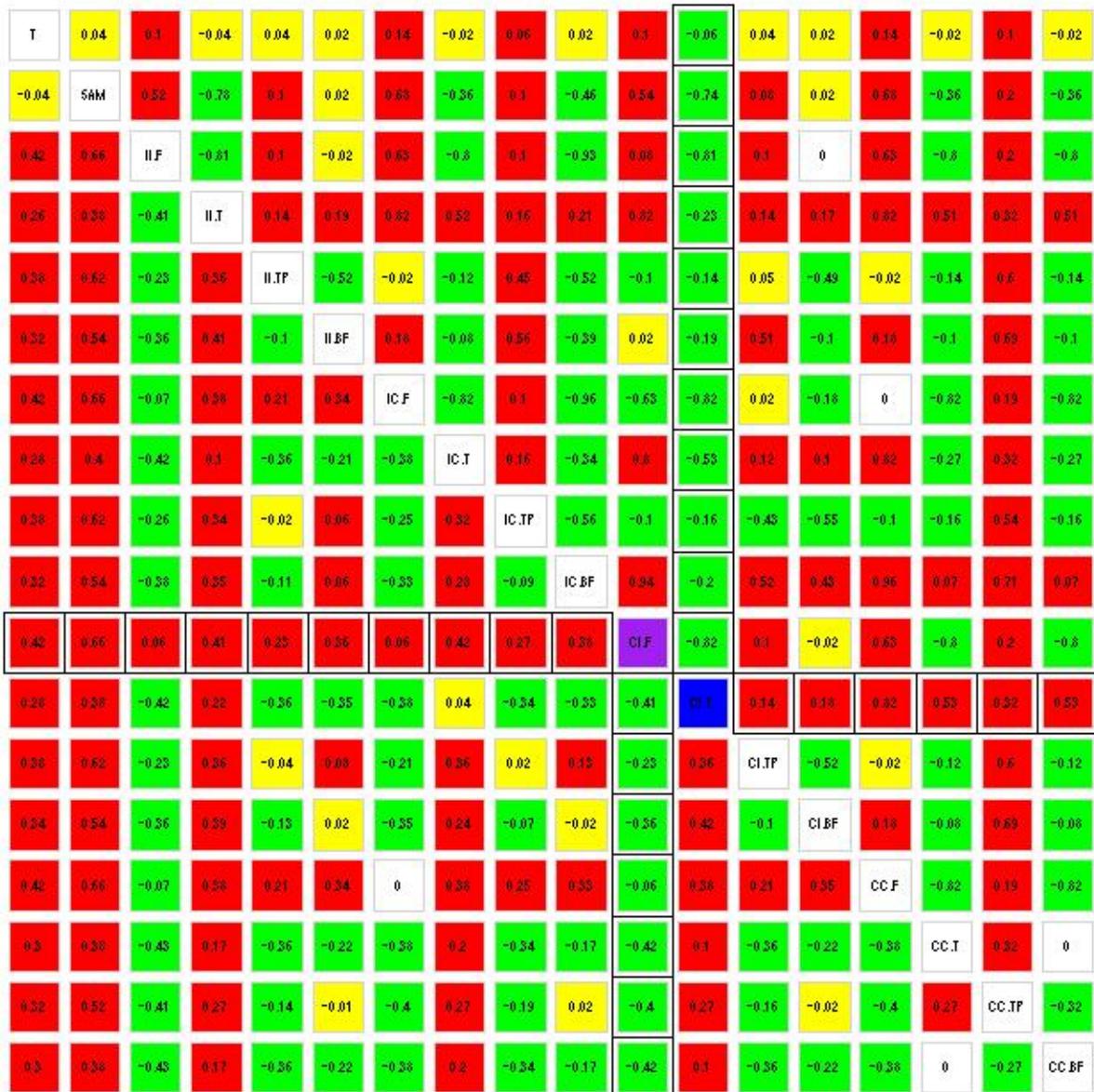


Figure 3: Heat map summary for simulations based on Independence model with number of replicates equal to 3 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

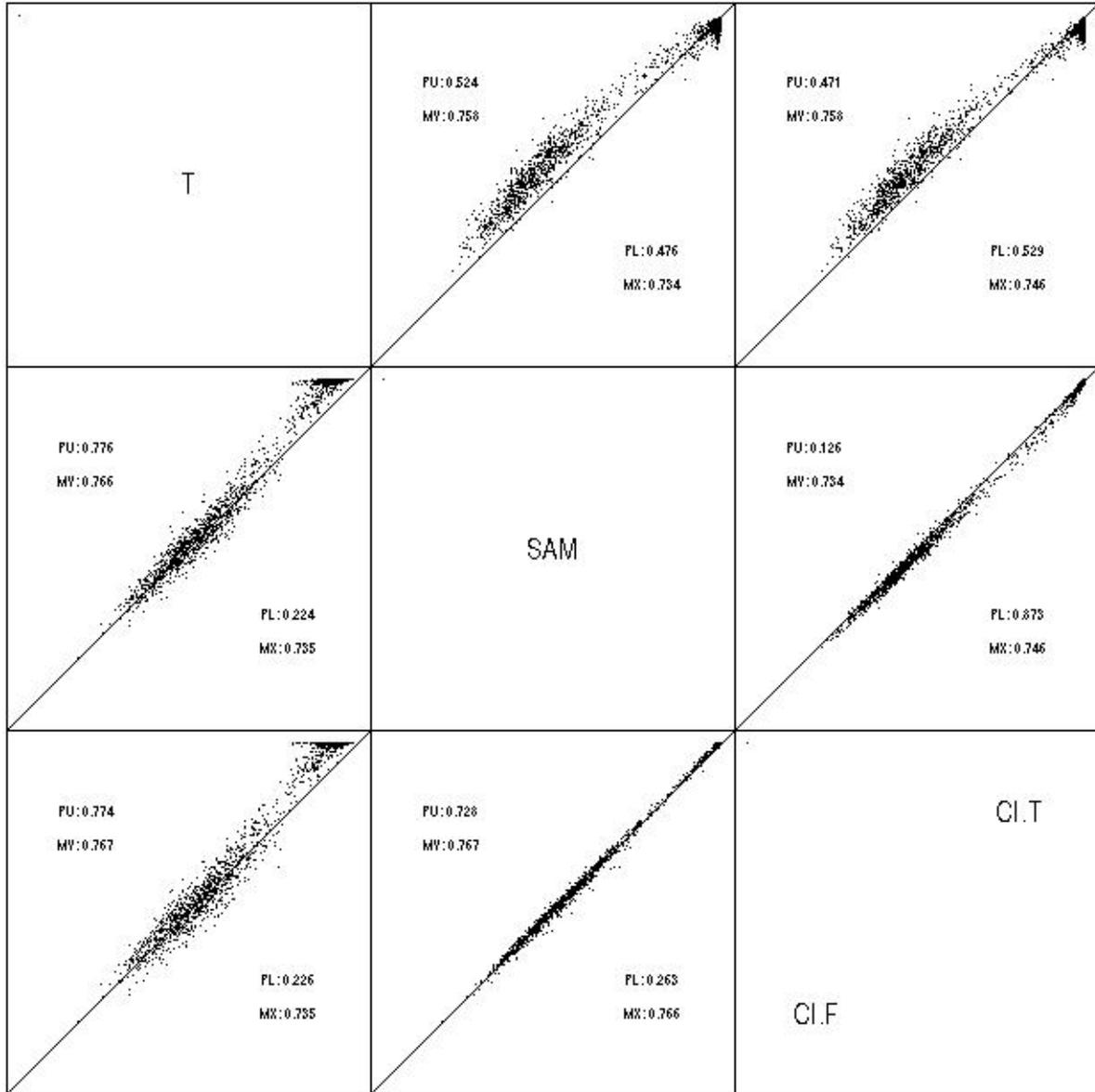


Figure 4: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 3 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

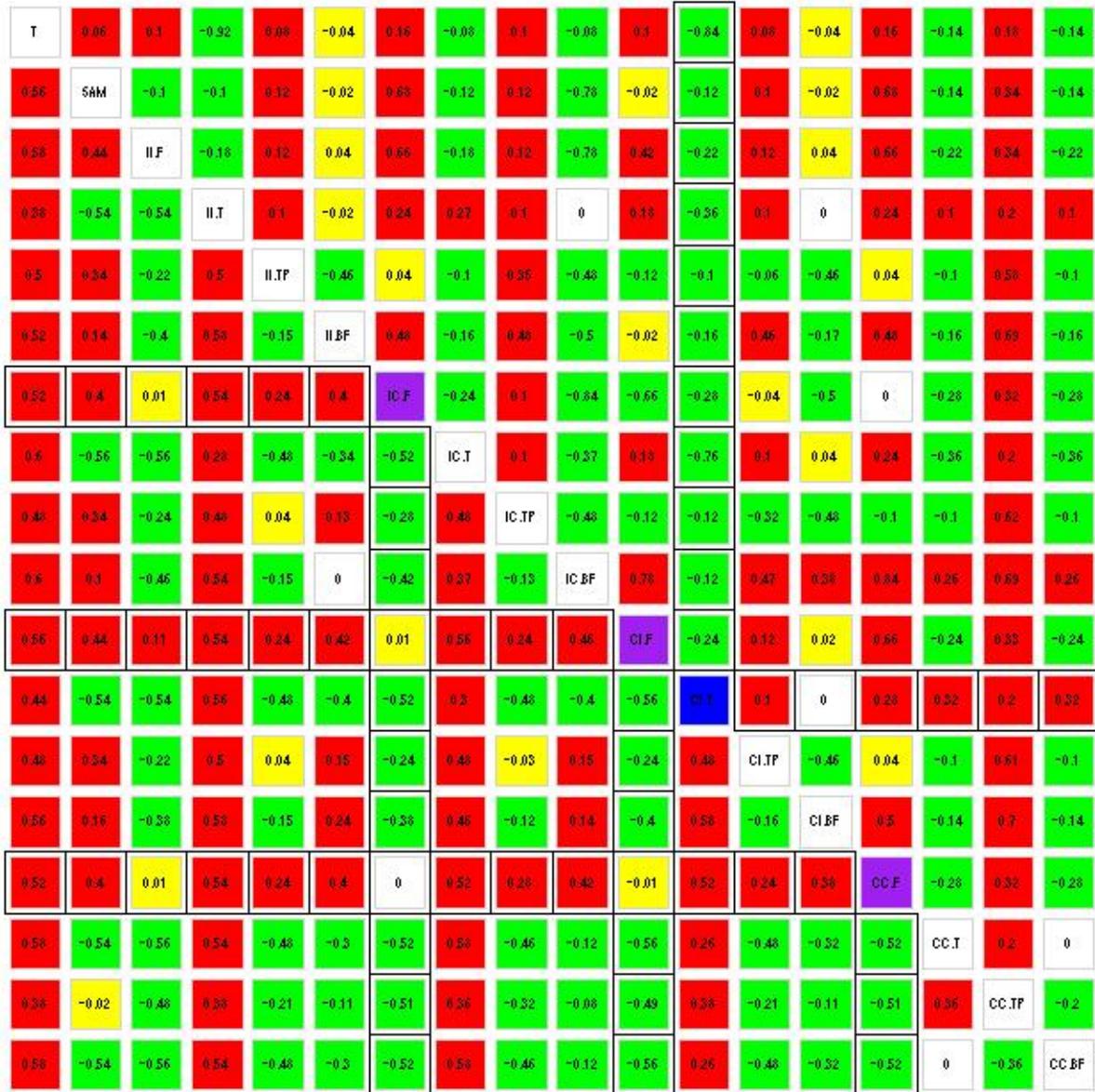


Figure 5: Heat map summary for simulations based on Independence model with number of replicates equal to 3 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

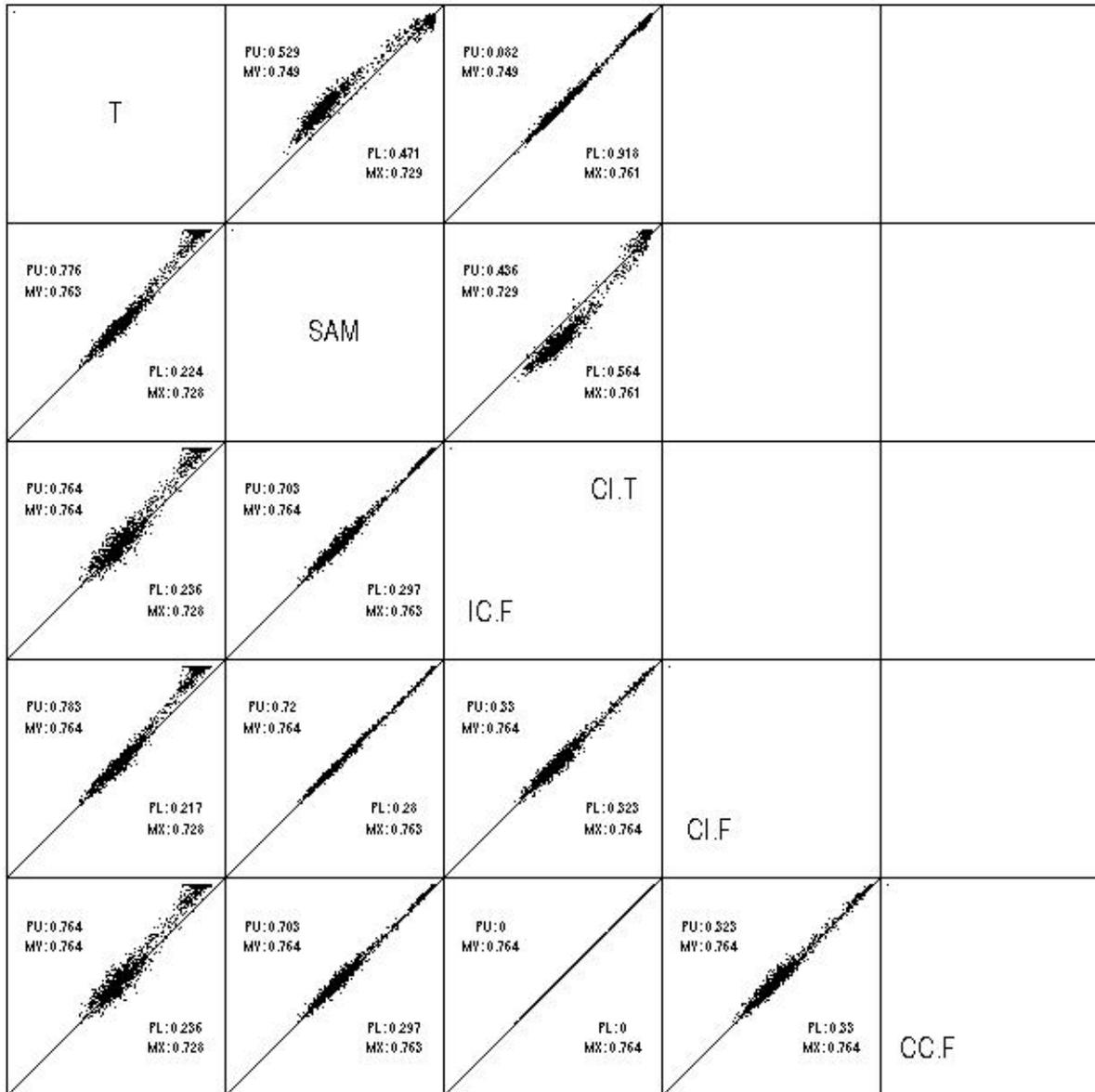


Figure 6: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 3 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

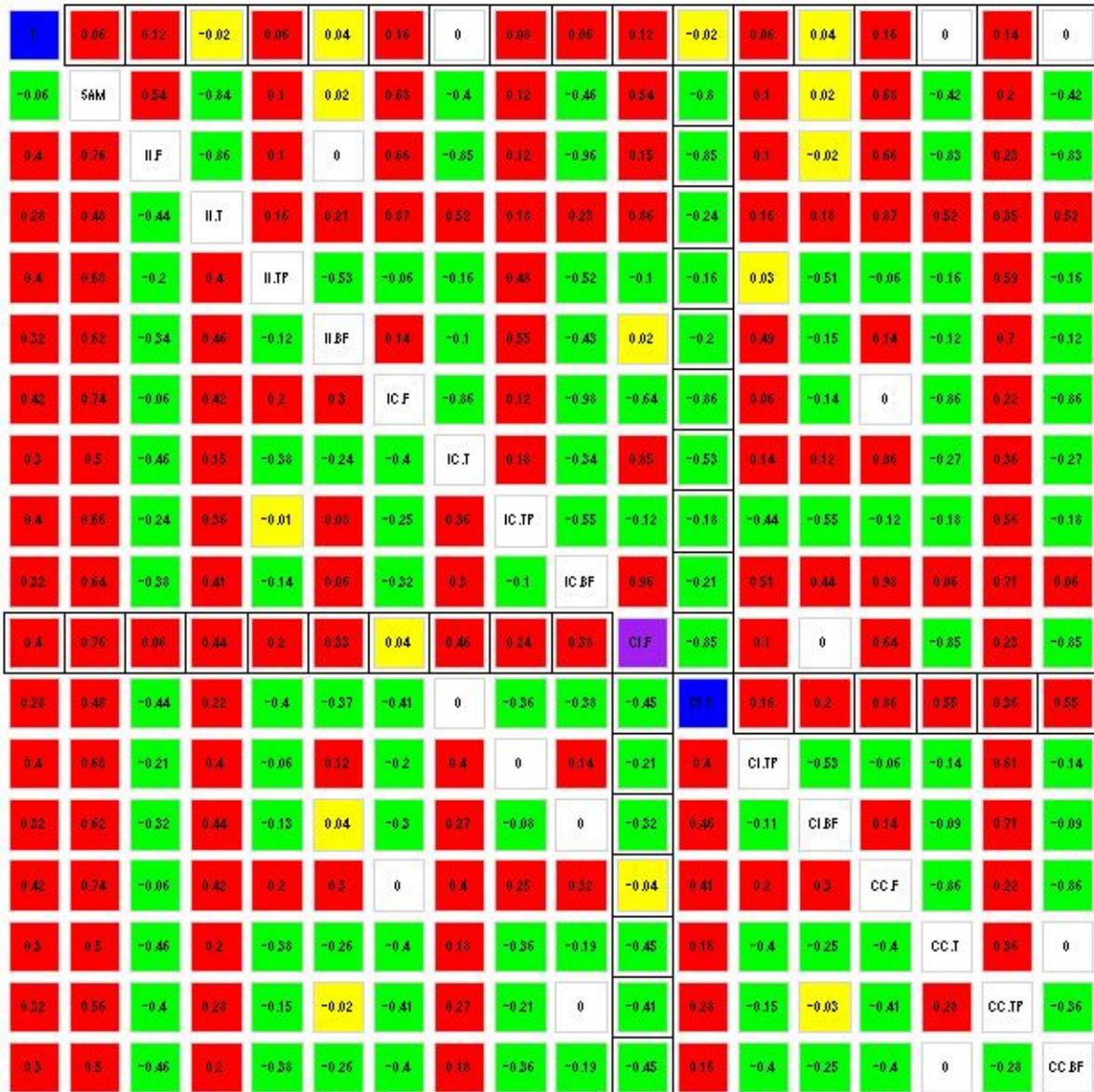


Figure 7: Heat map summary for simulations based on Independence model with number of replicates equal to 3 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

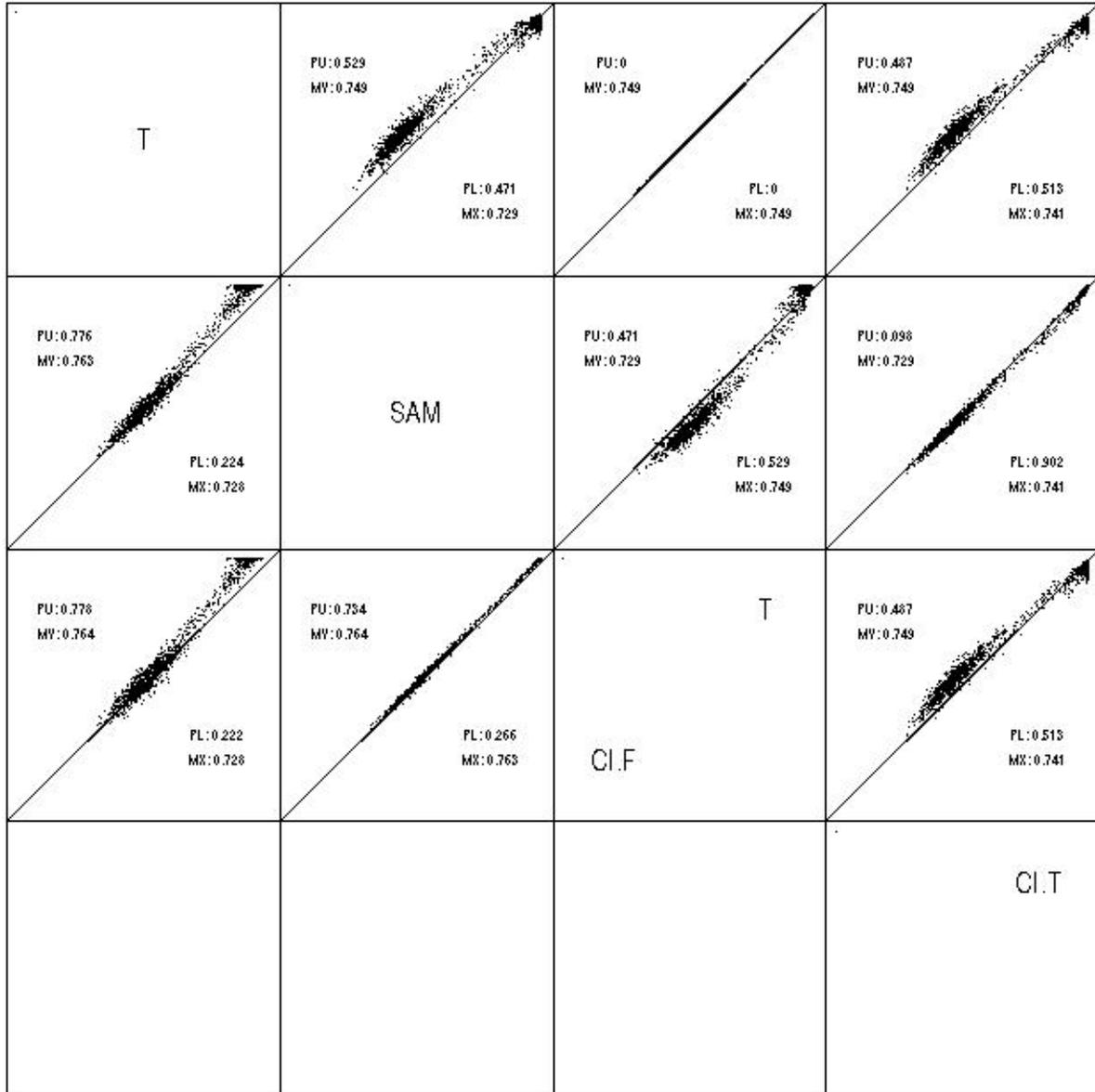


Figure 8: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 3 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

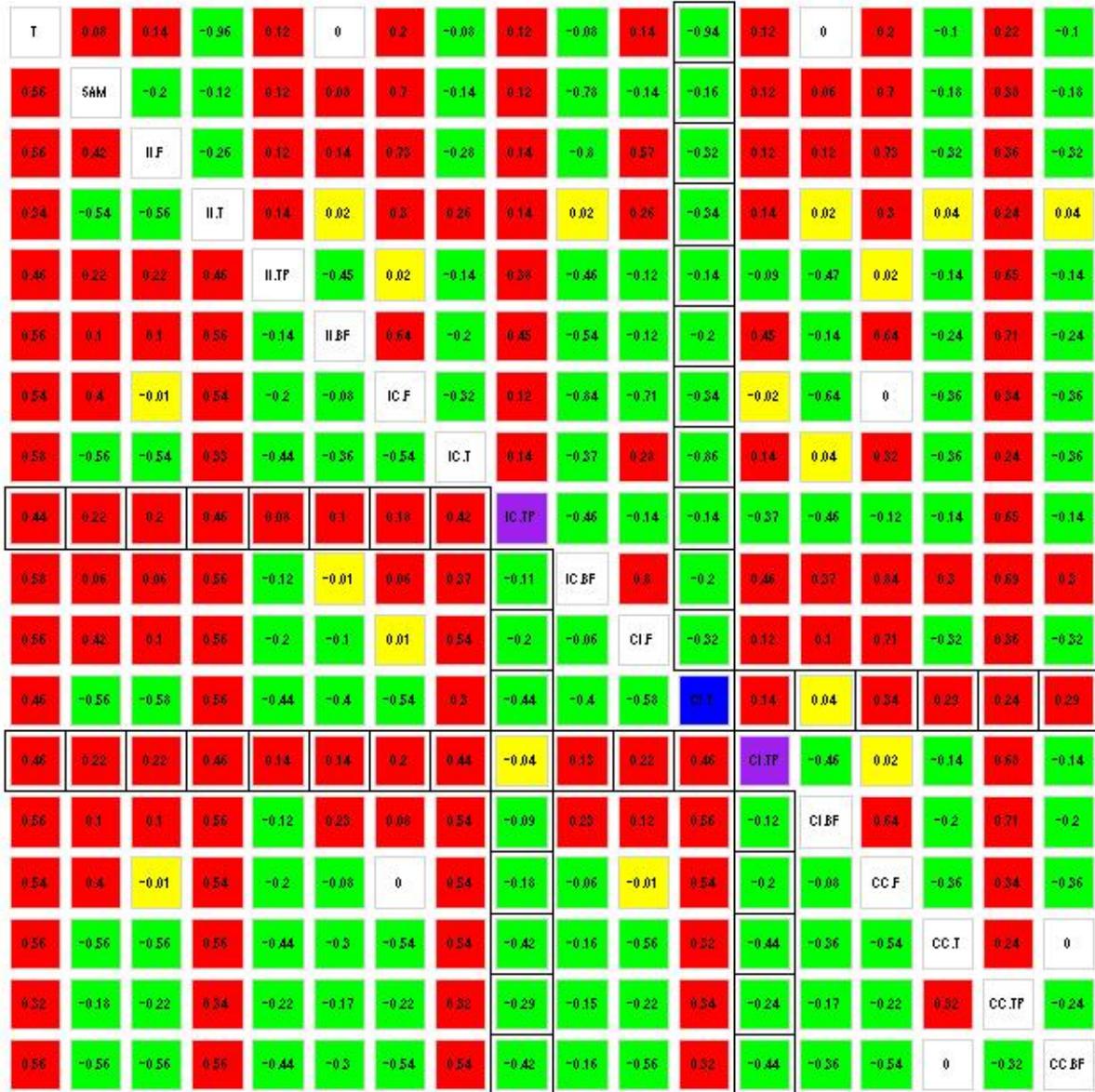


Figure 9: Heat map summary for simulations based on Independence model with number of replicates equal to 3 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

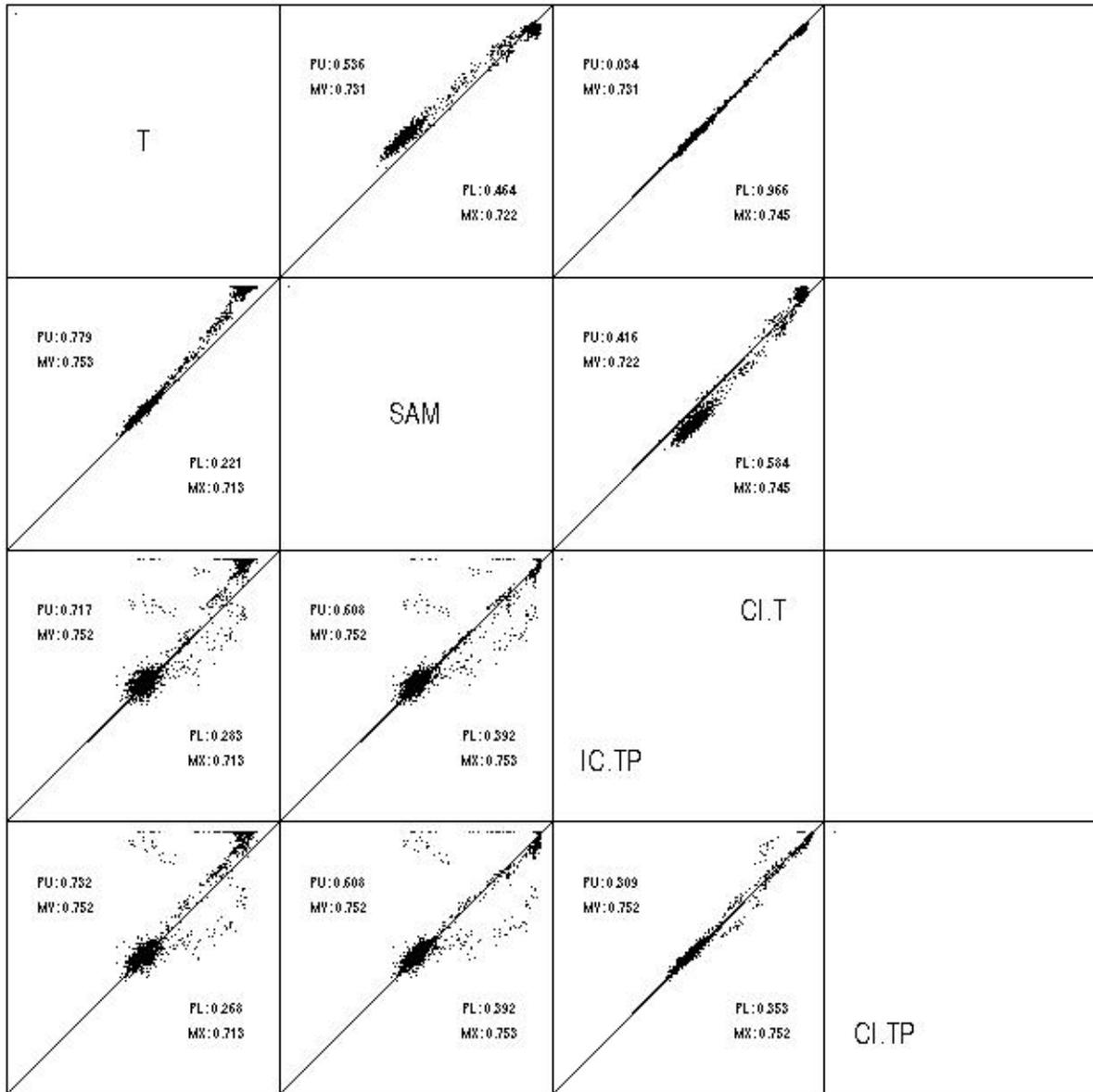


Figure 10: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 3 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

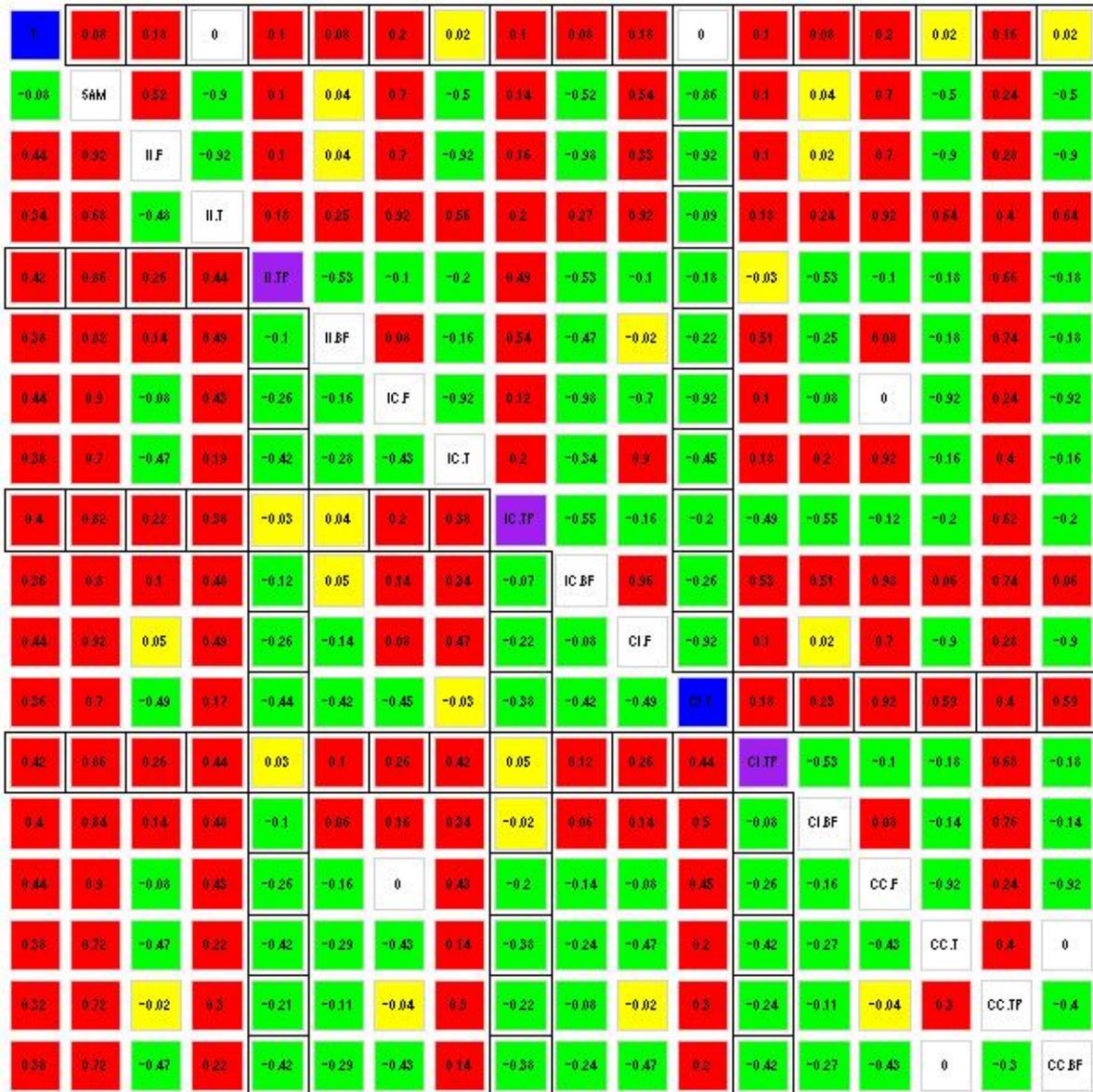


Figure 11: Heat map summary for simulations based on Independence model with number of replicates equal to 3 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

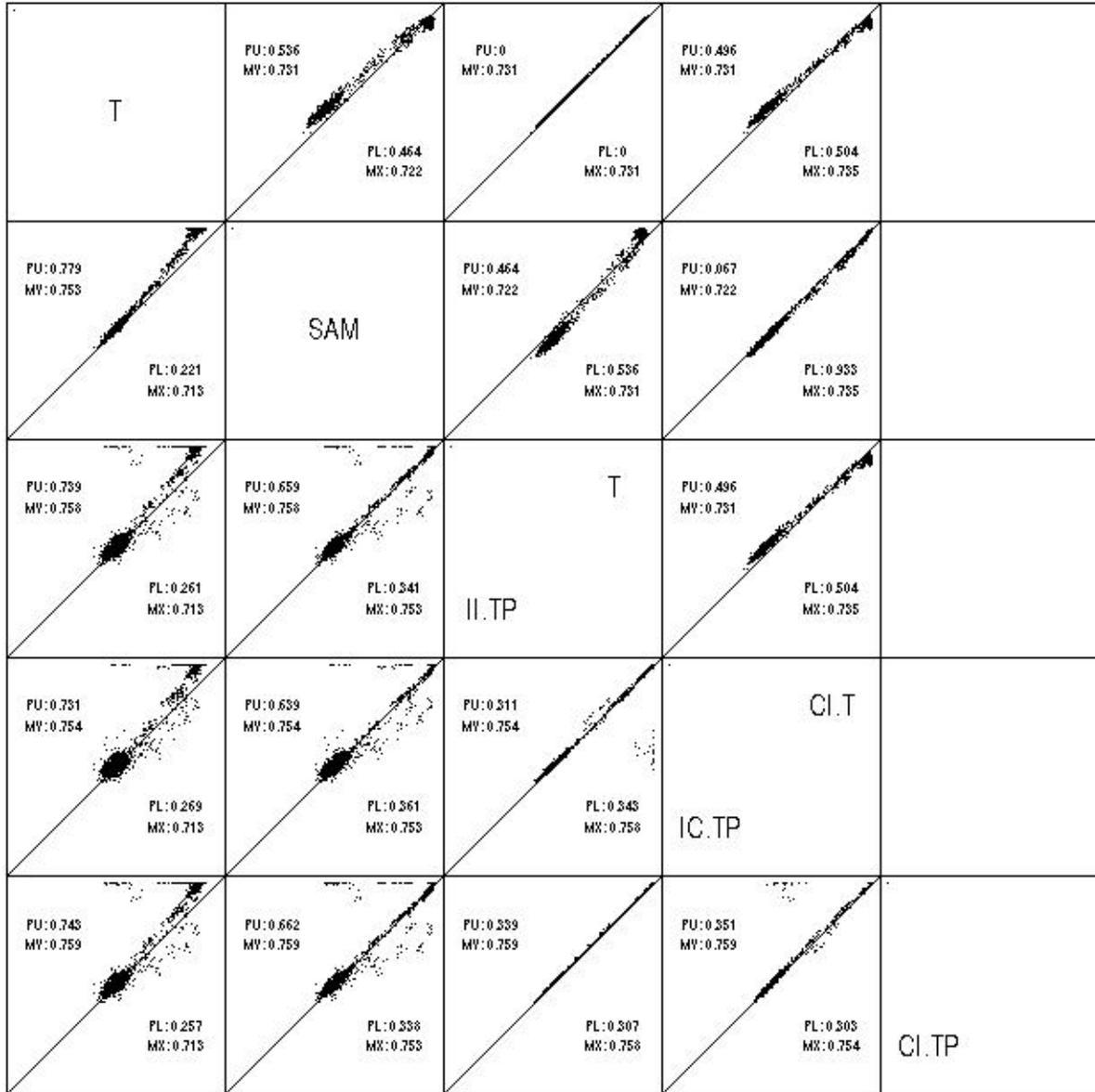


Figure 12: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 3 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

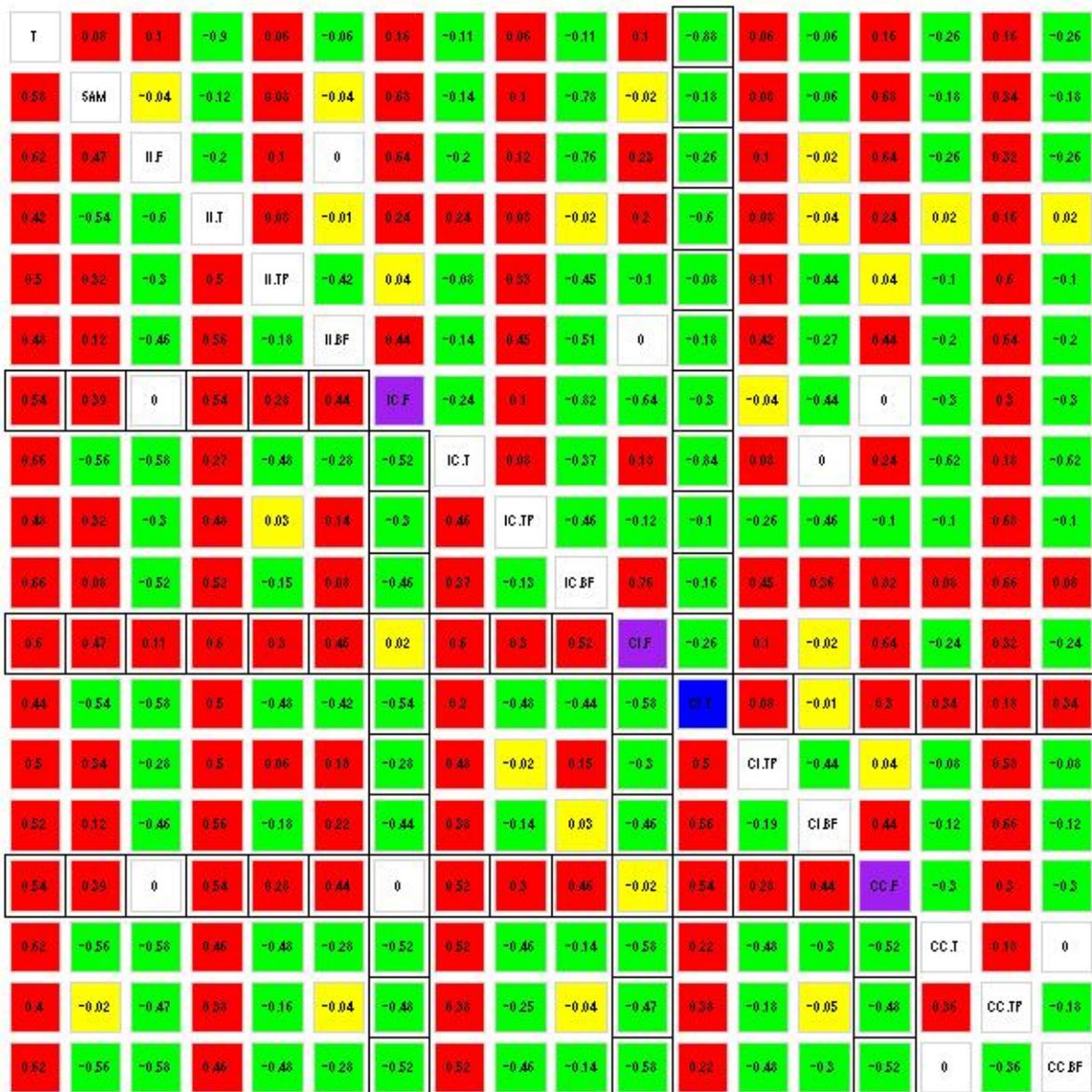


Figure 13: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 3 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

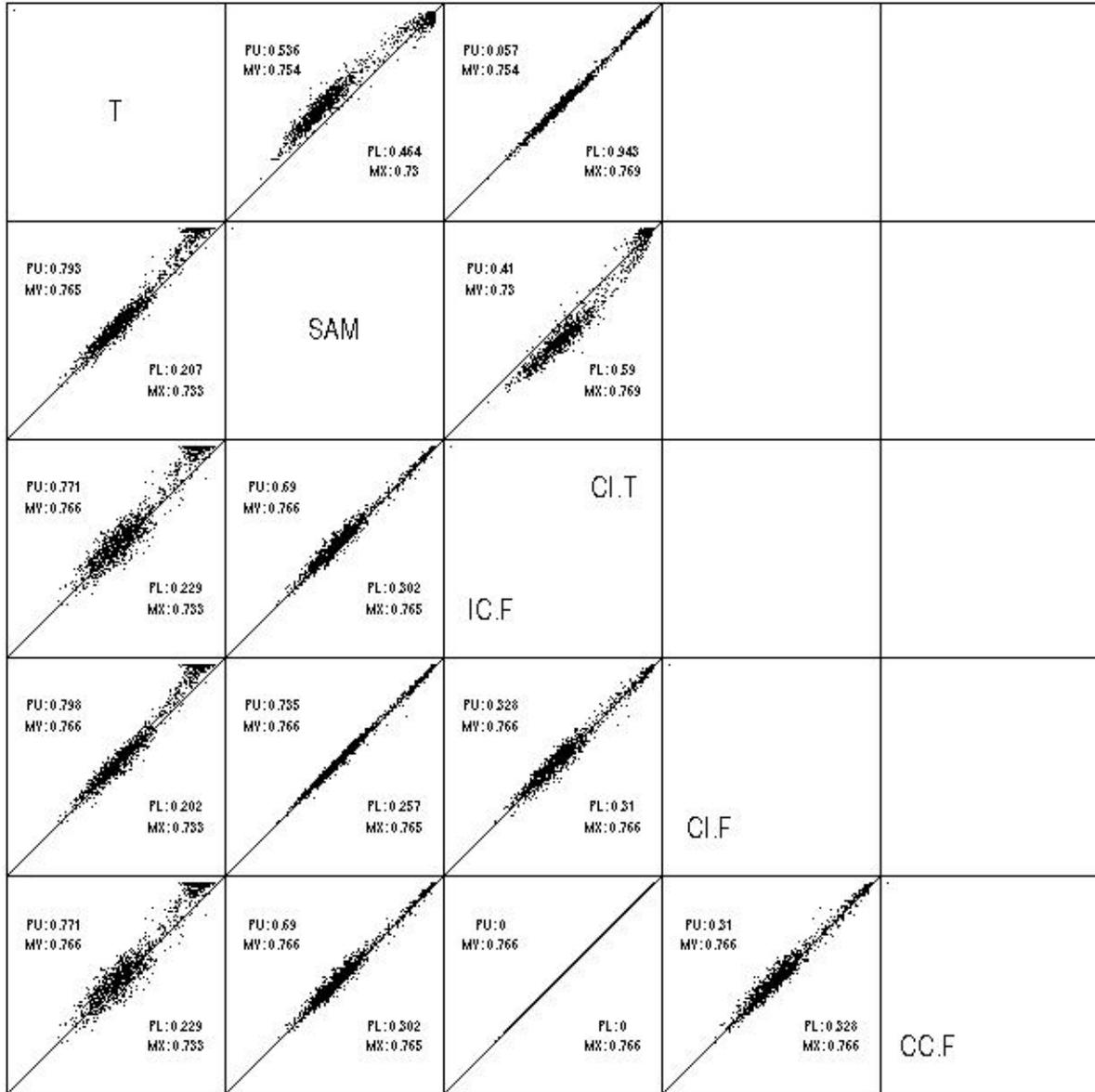


Figure 14: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 3 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

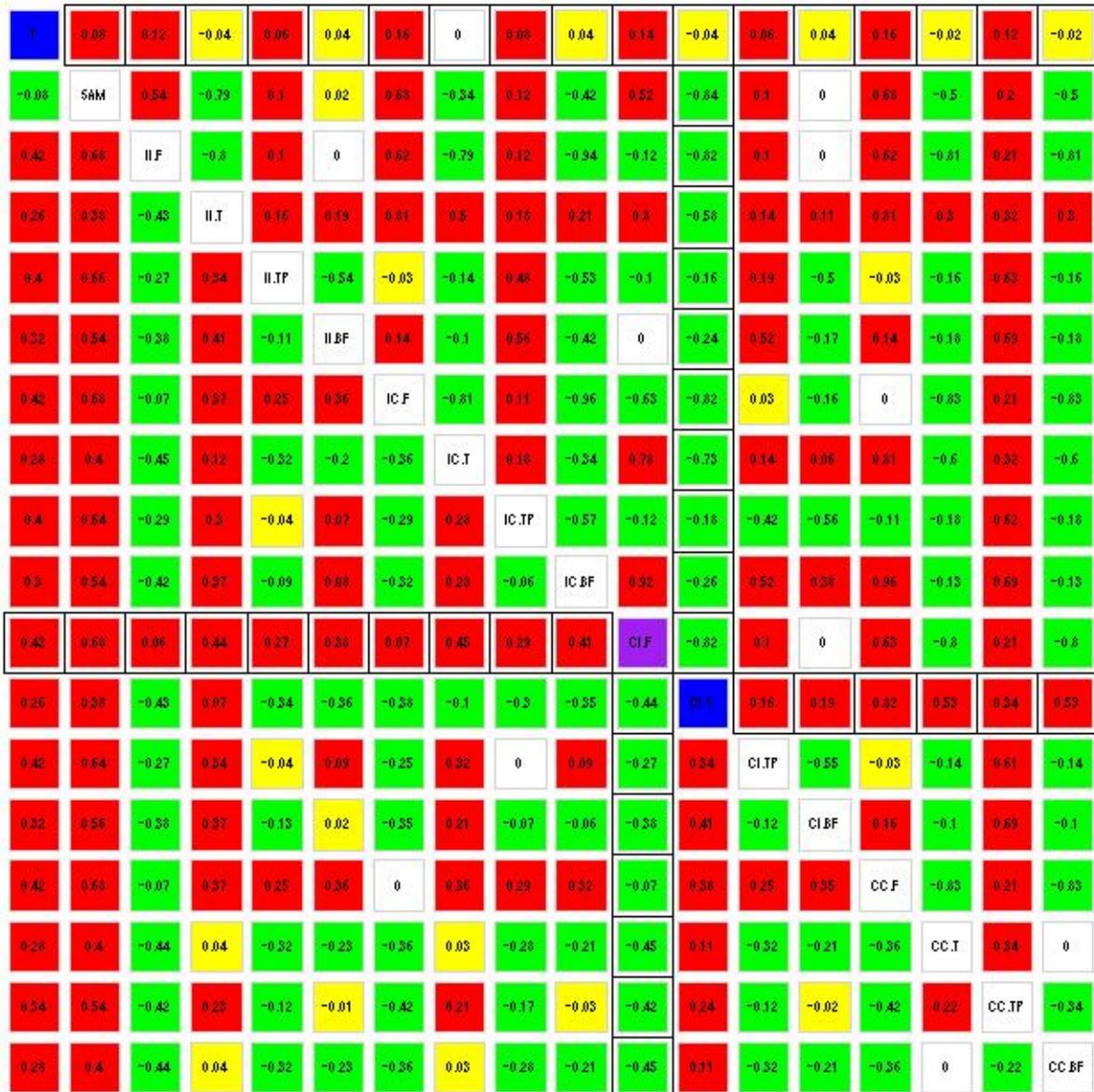


Figure 15: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 3 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

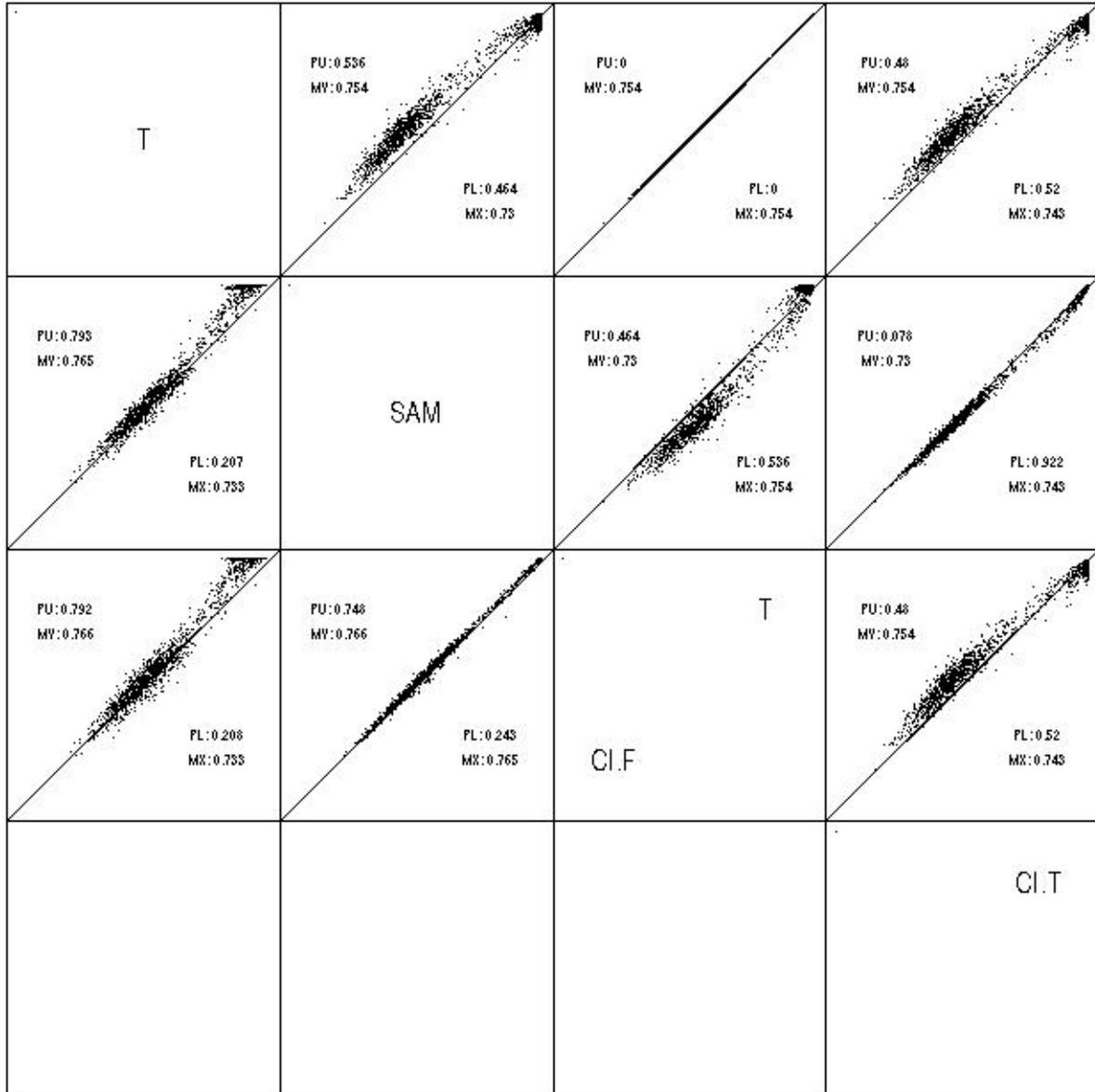


Figure 16: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 3 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

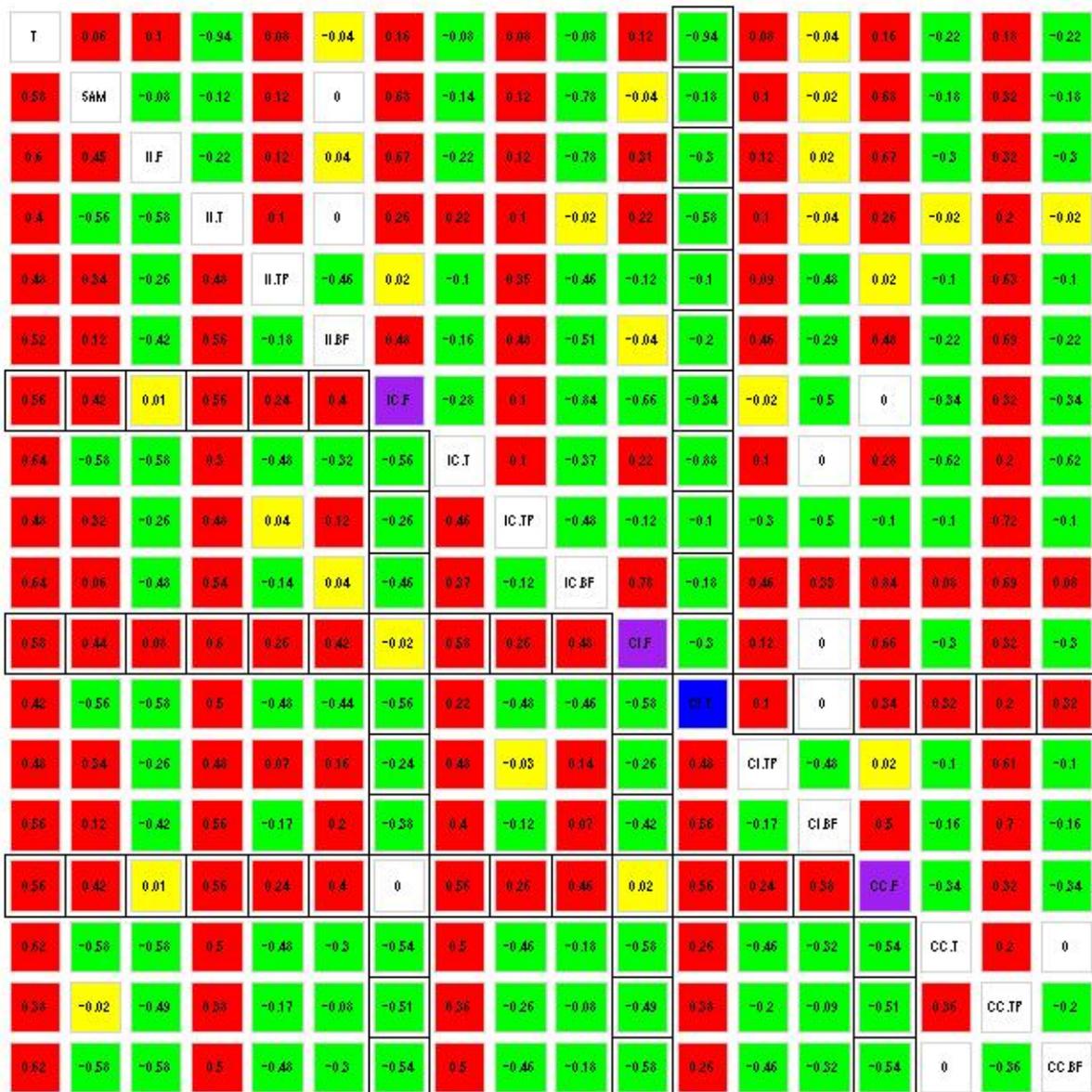


Figure 17: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 3 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

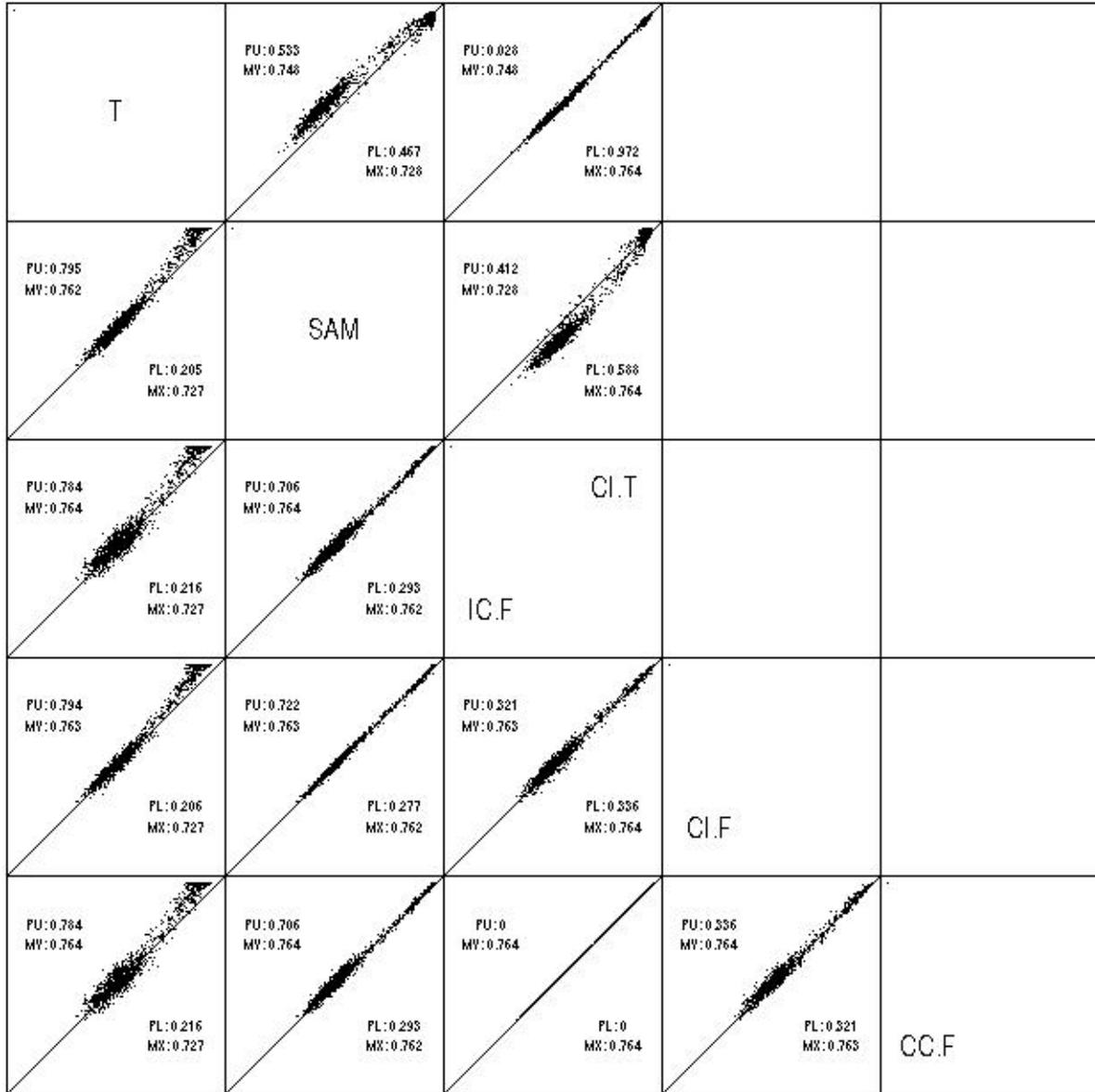


Figure 18: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 3 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

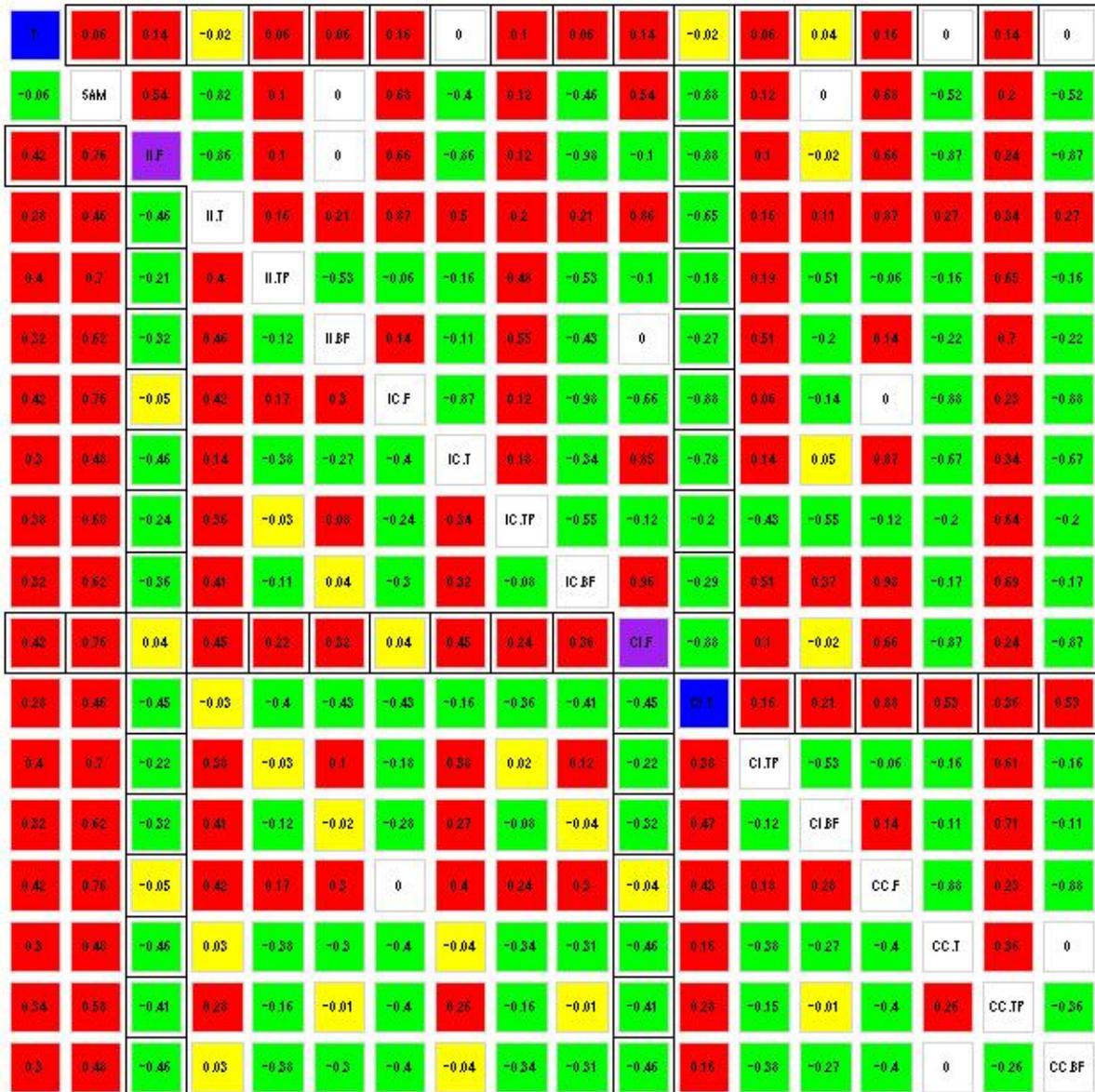


Figure 19: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 3 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

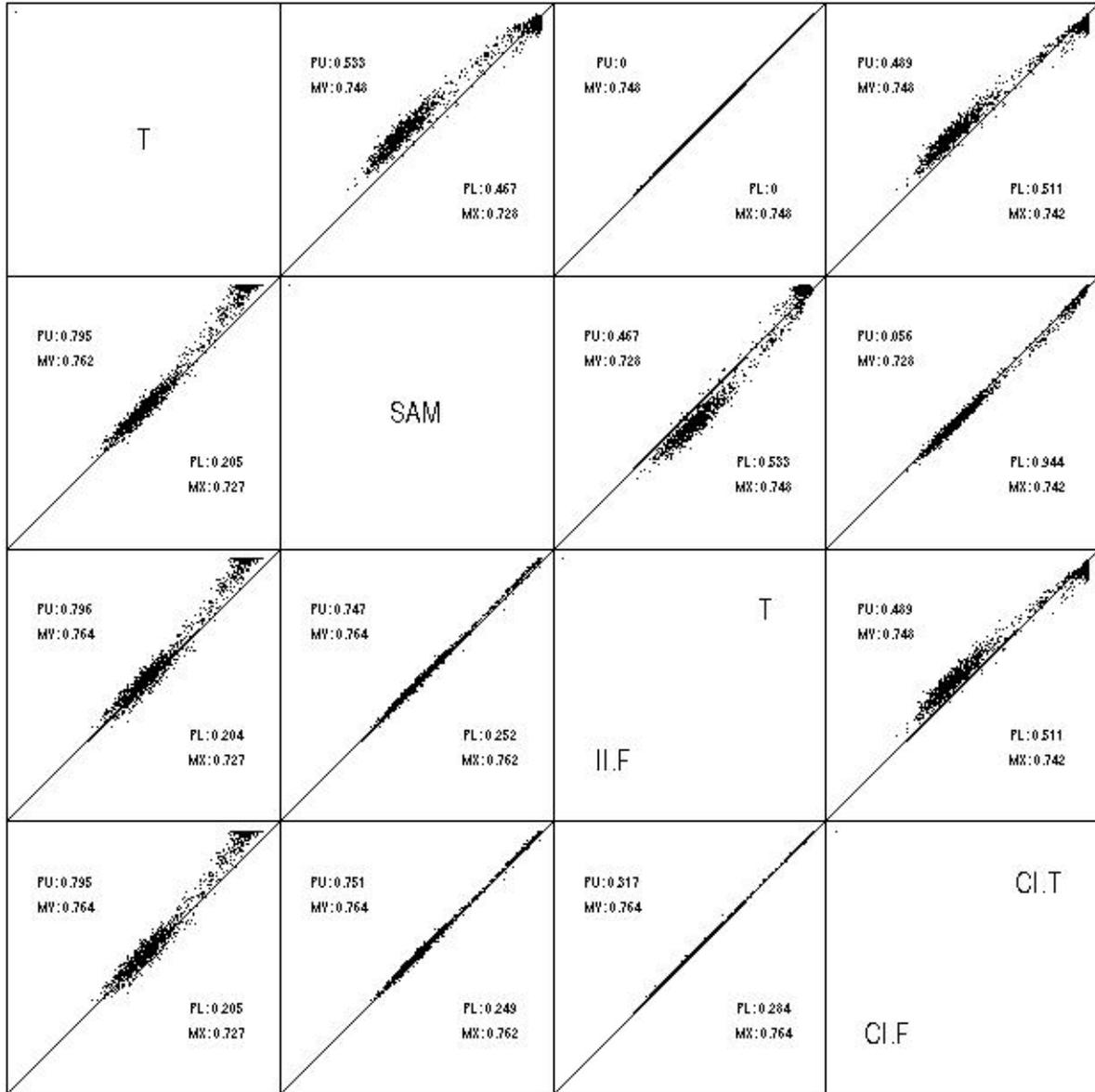


Figure 20: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 3 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.



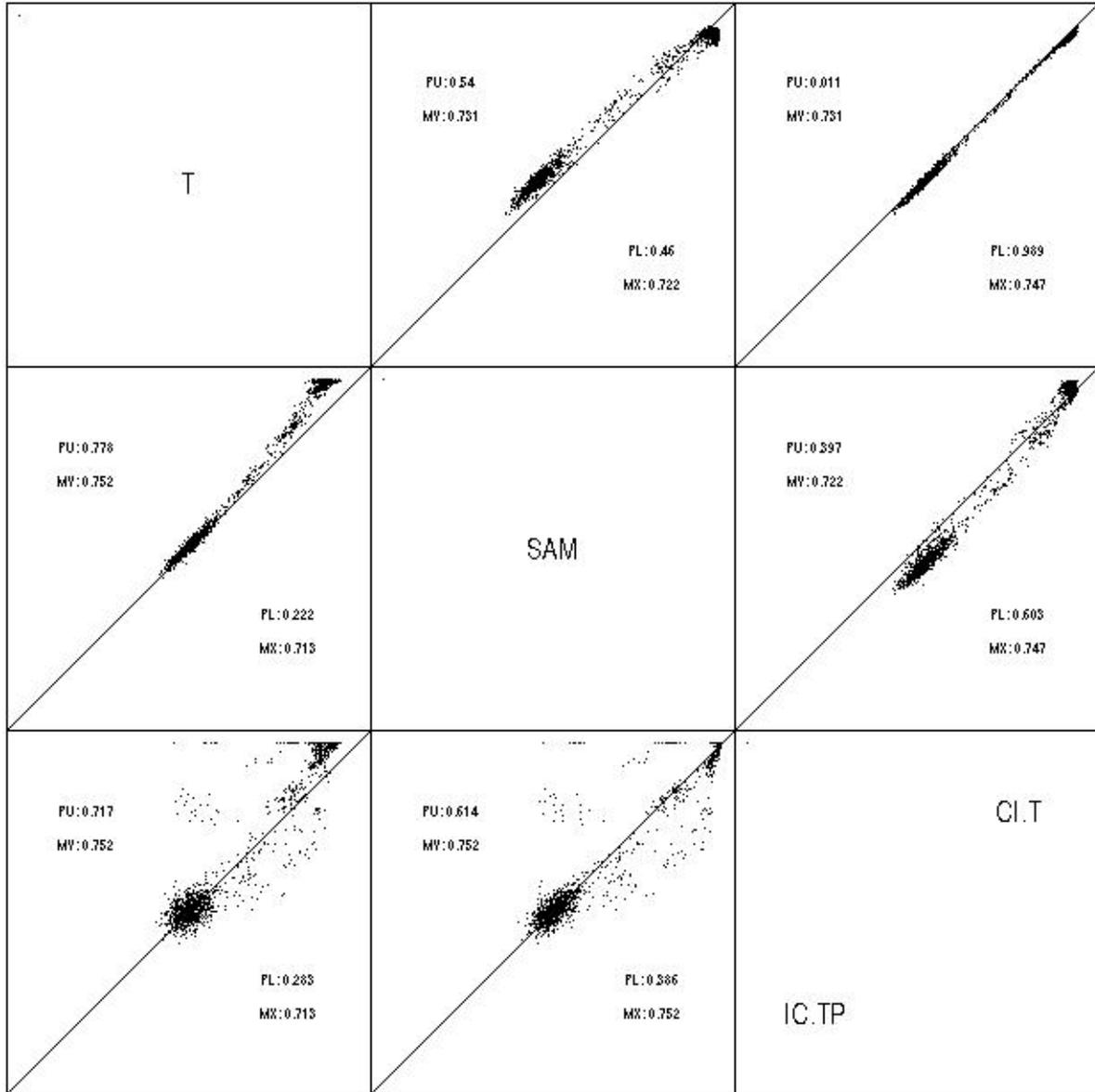


Figure 22: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 3 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

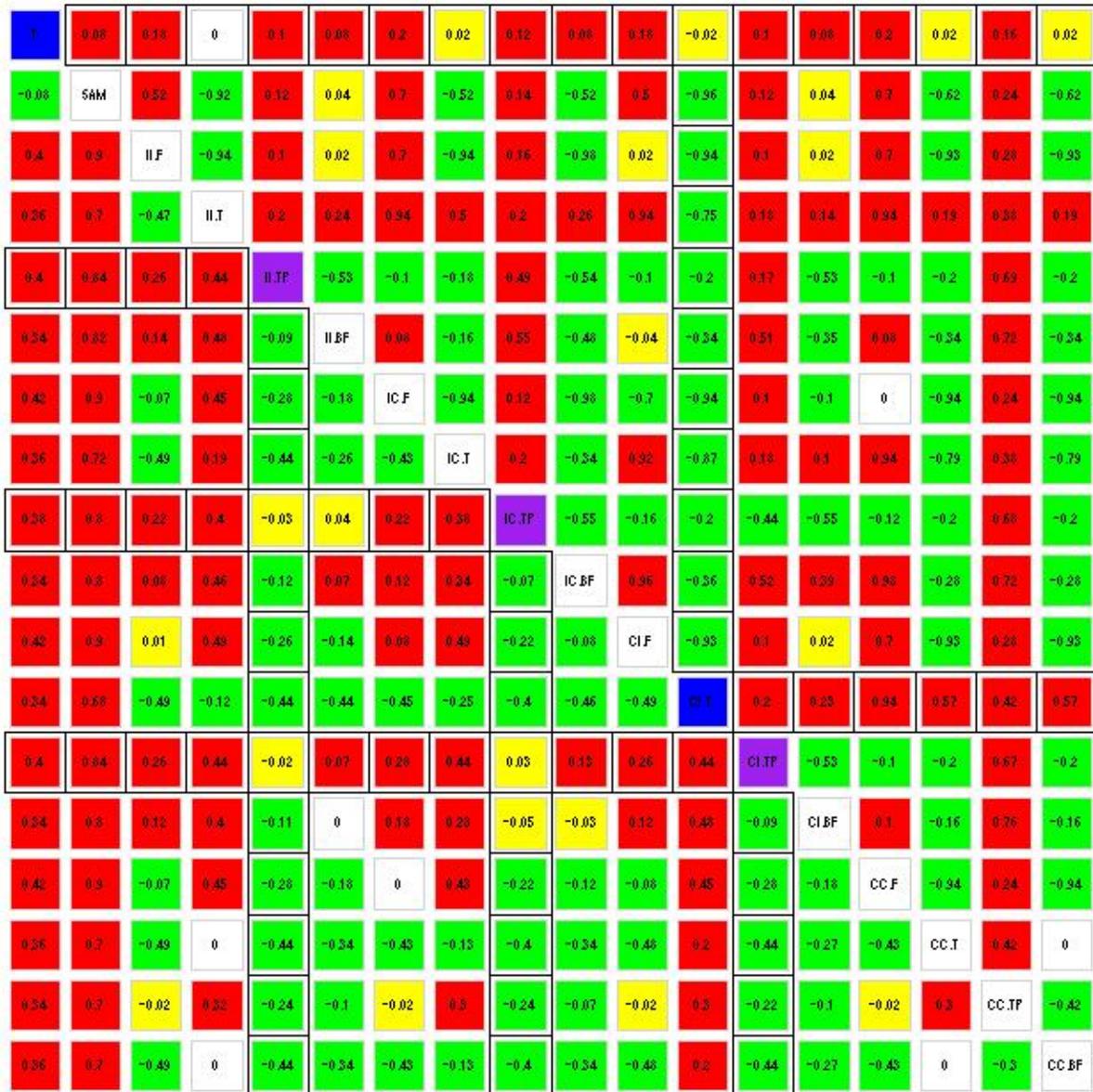


Figure 23: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 3 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

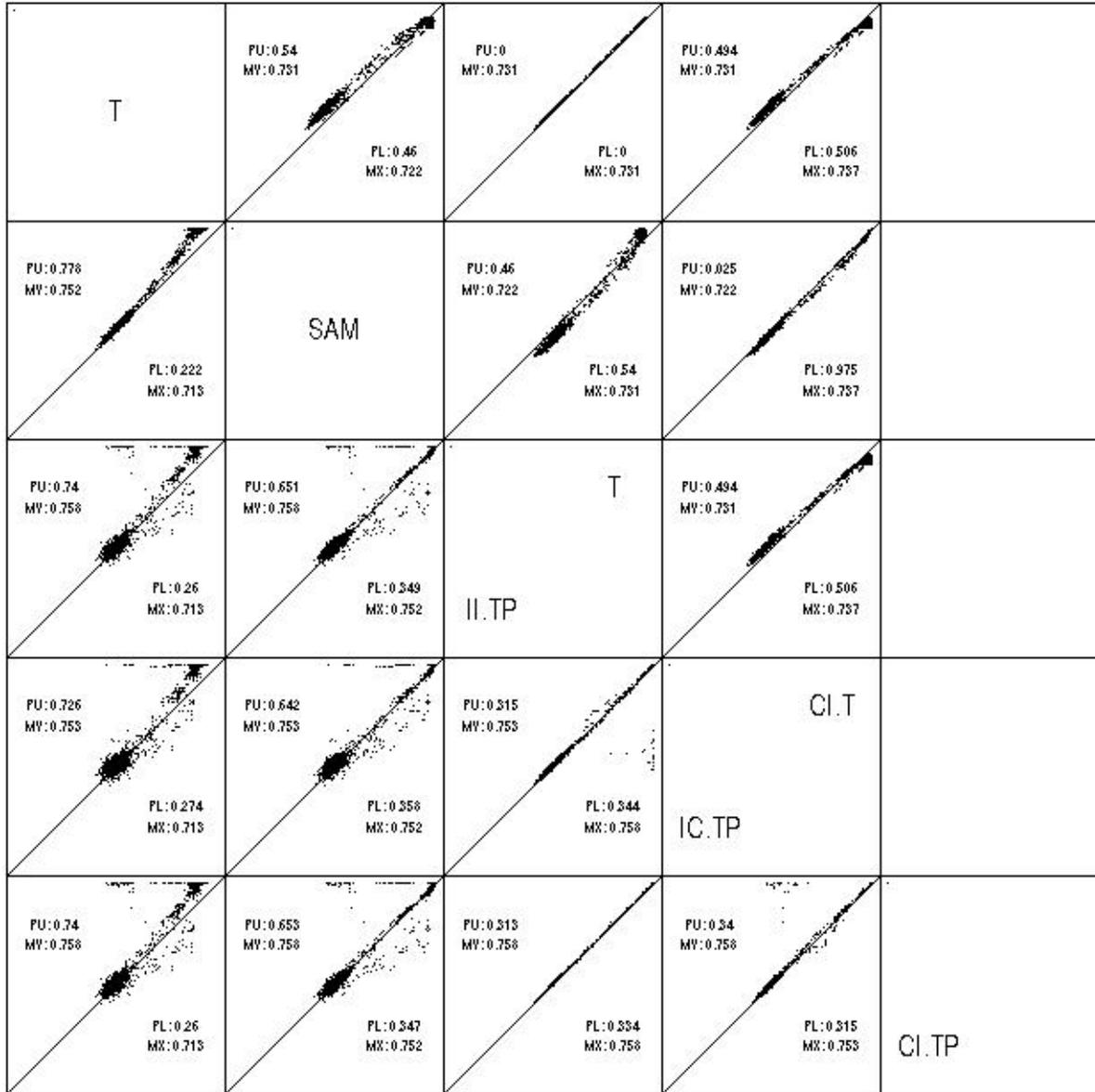


Figure 24: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 3 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

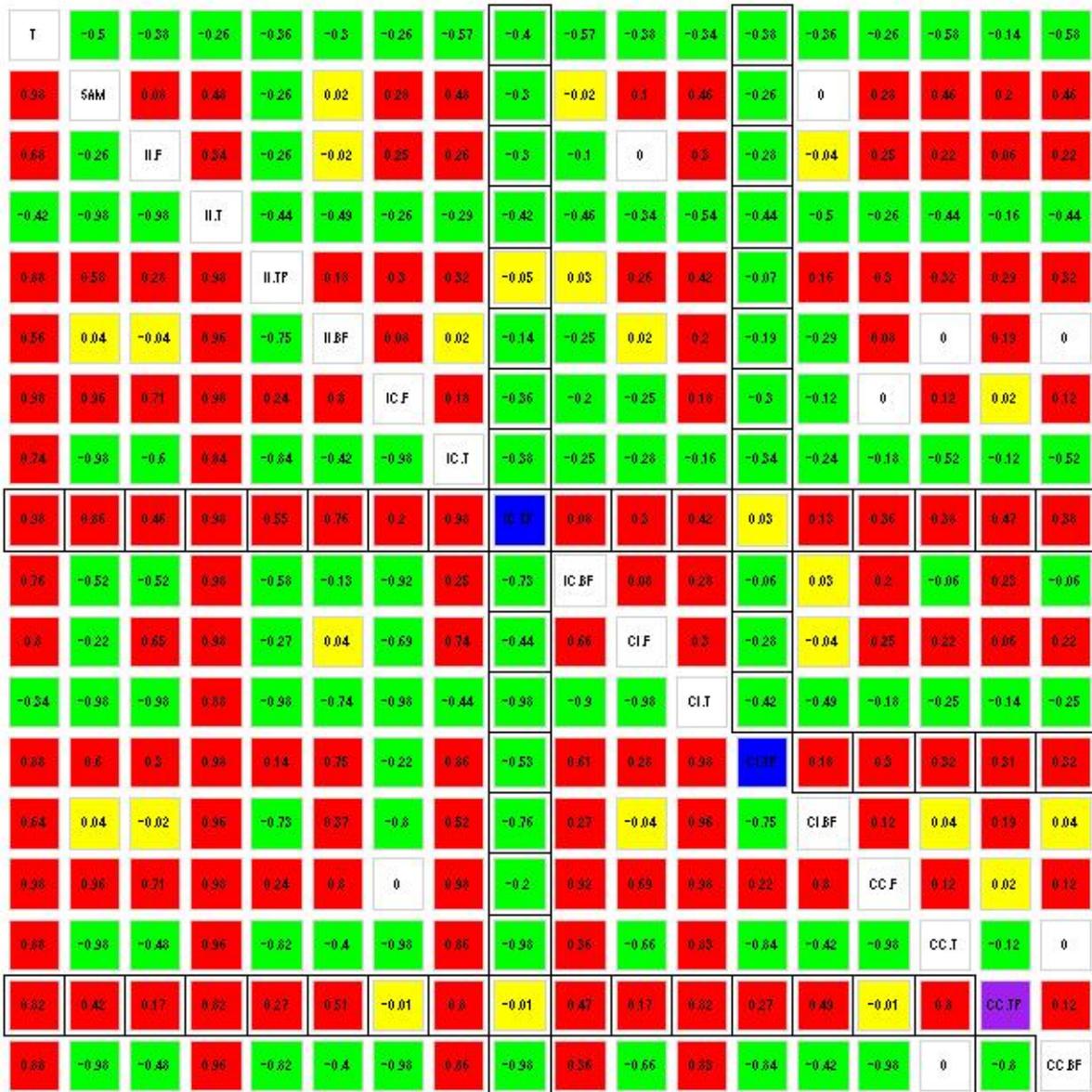


Figure 25: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

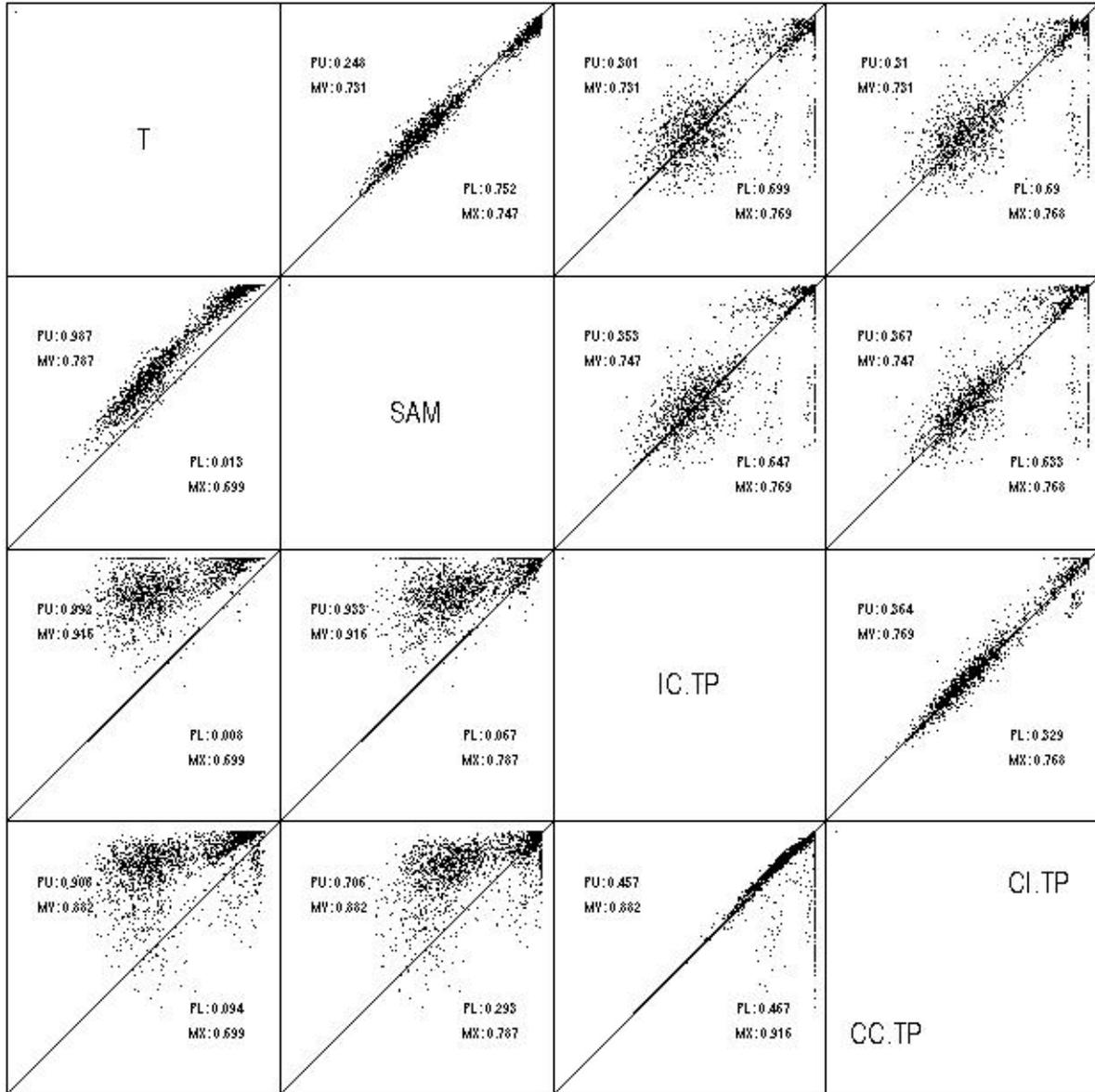


Figure 26: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.



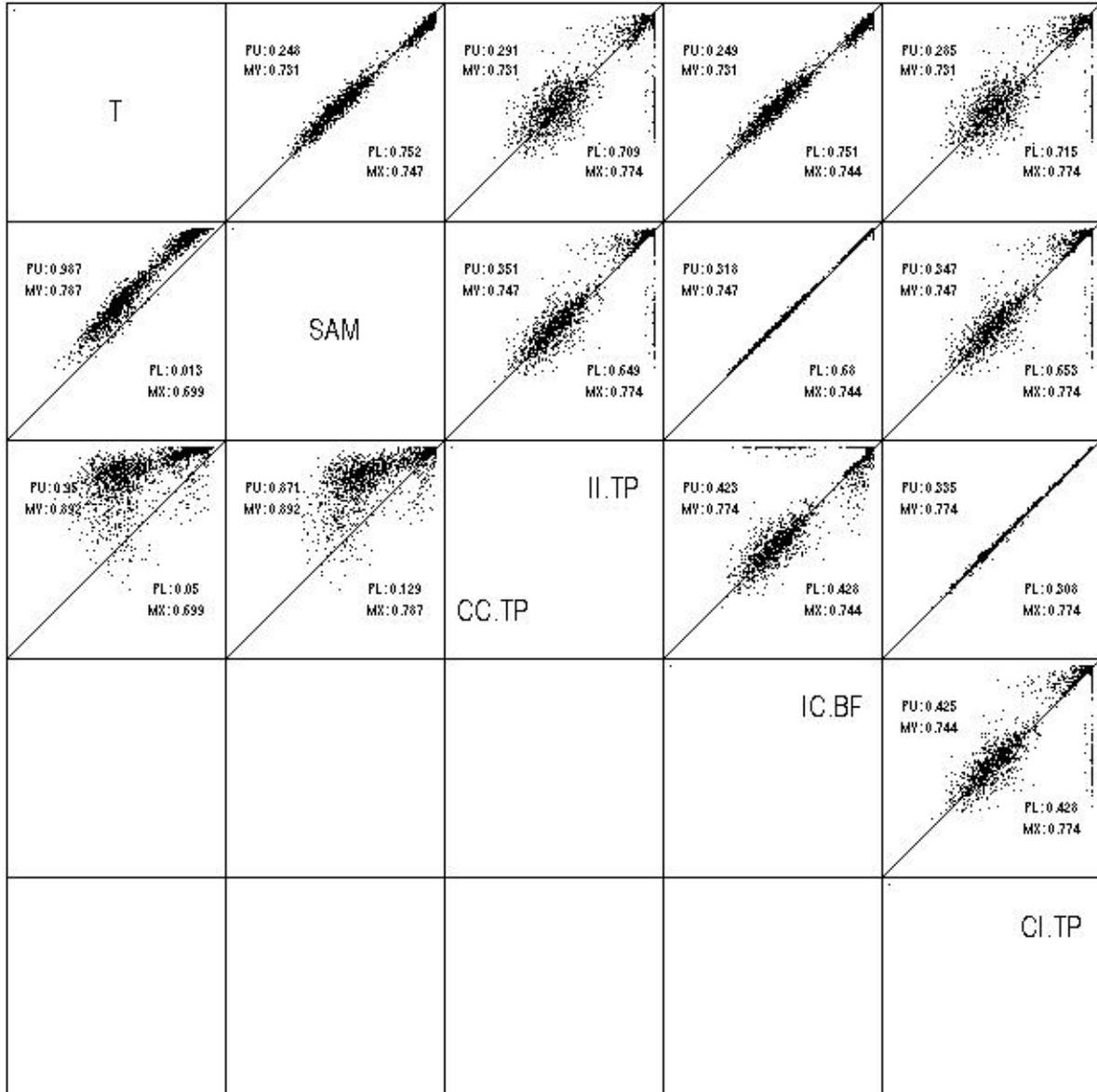


Figure 28: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

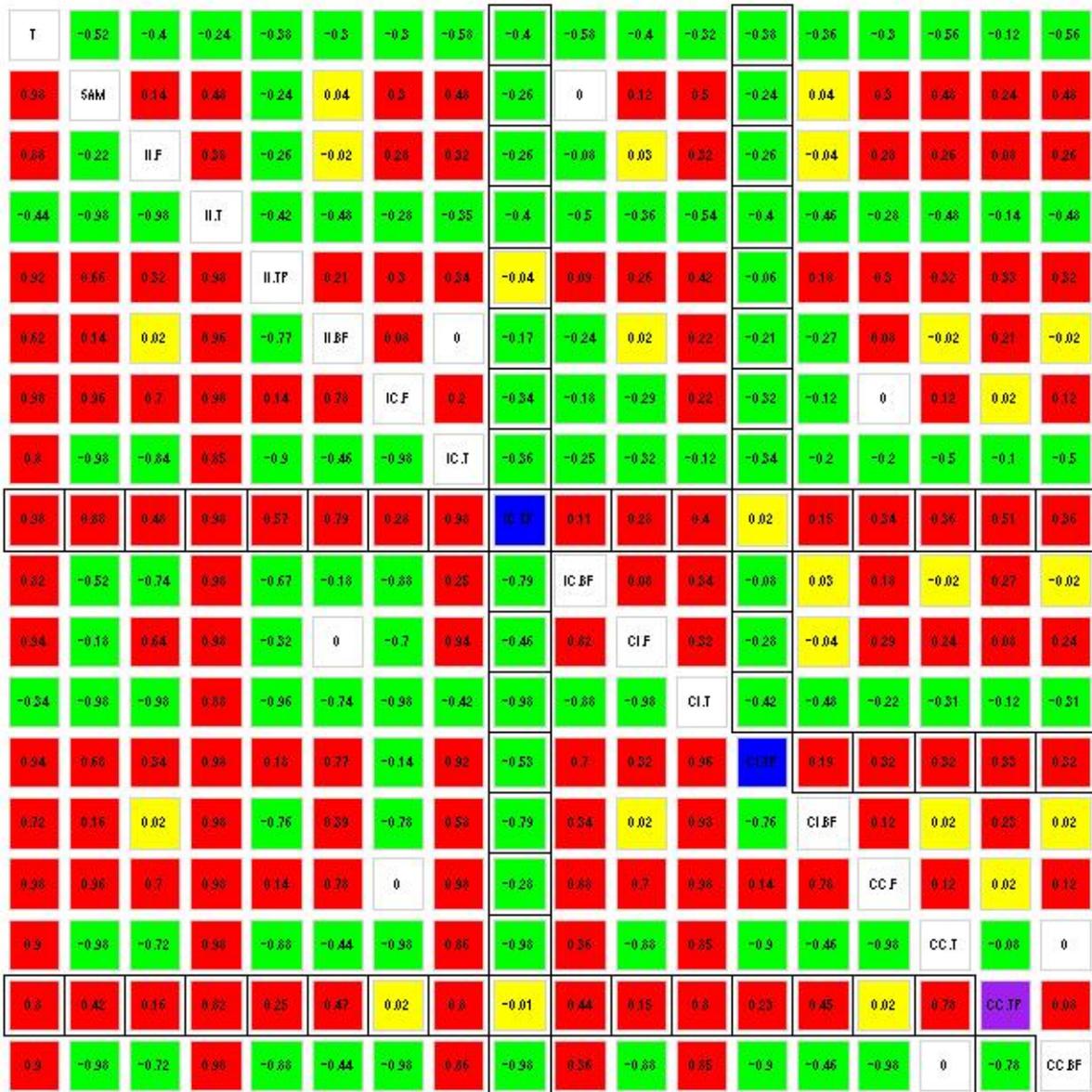


Figure 29: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

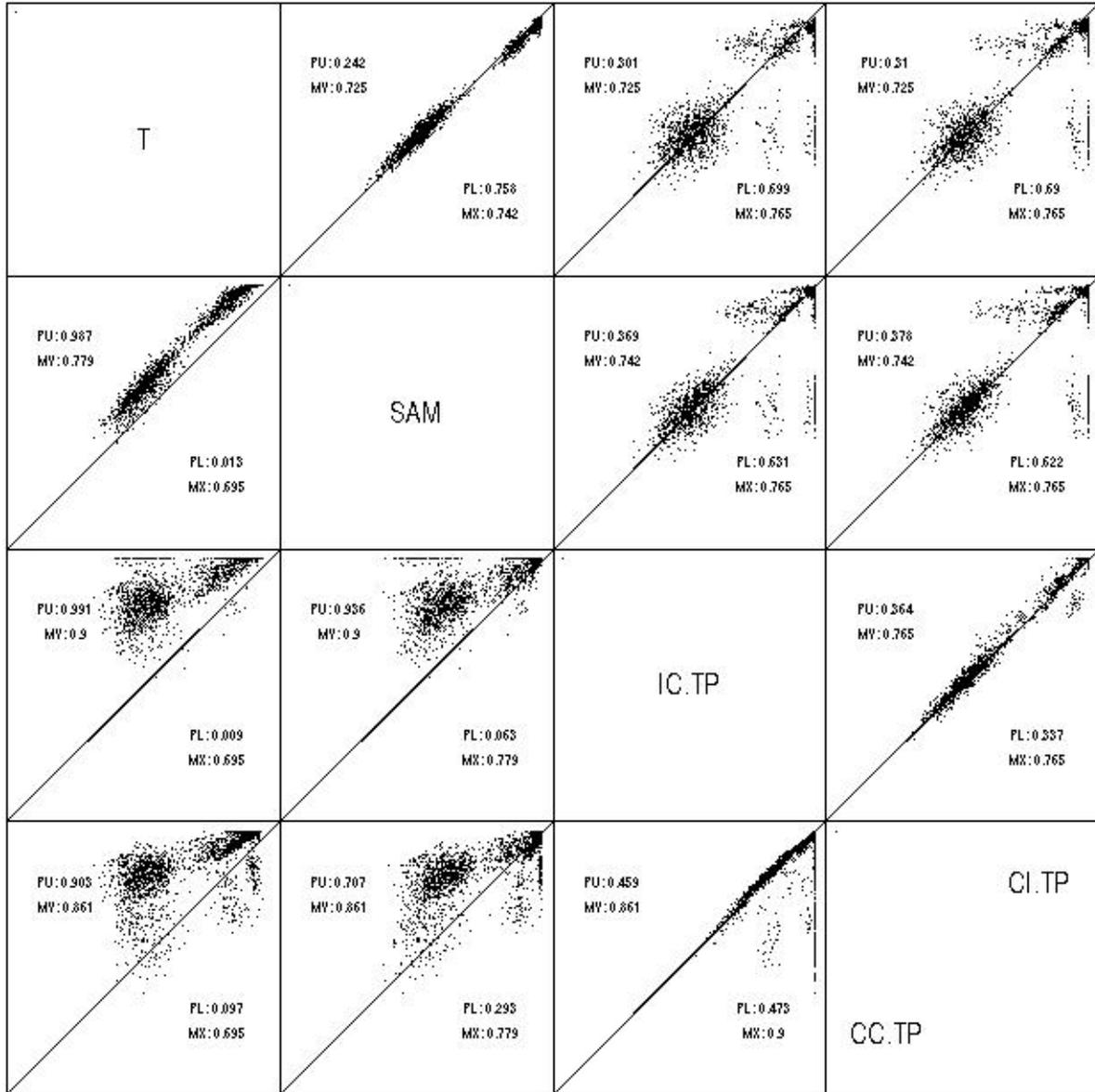


Figure 30: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

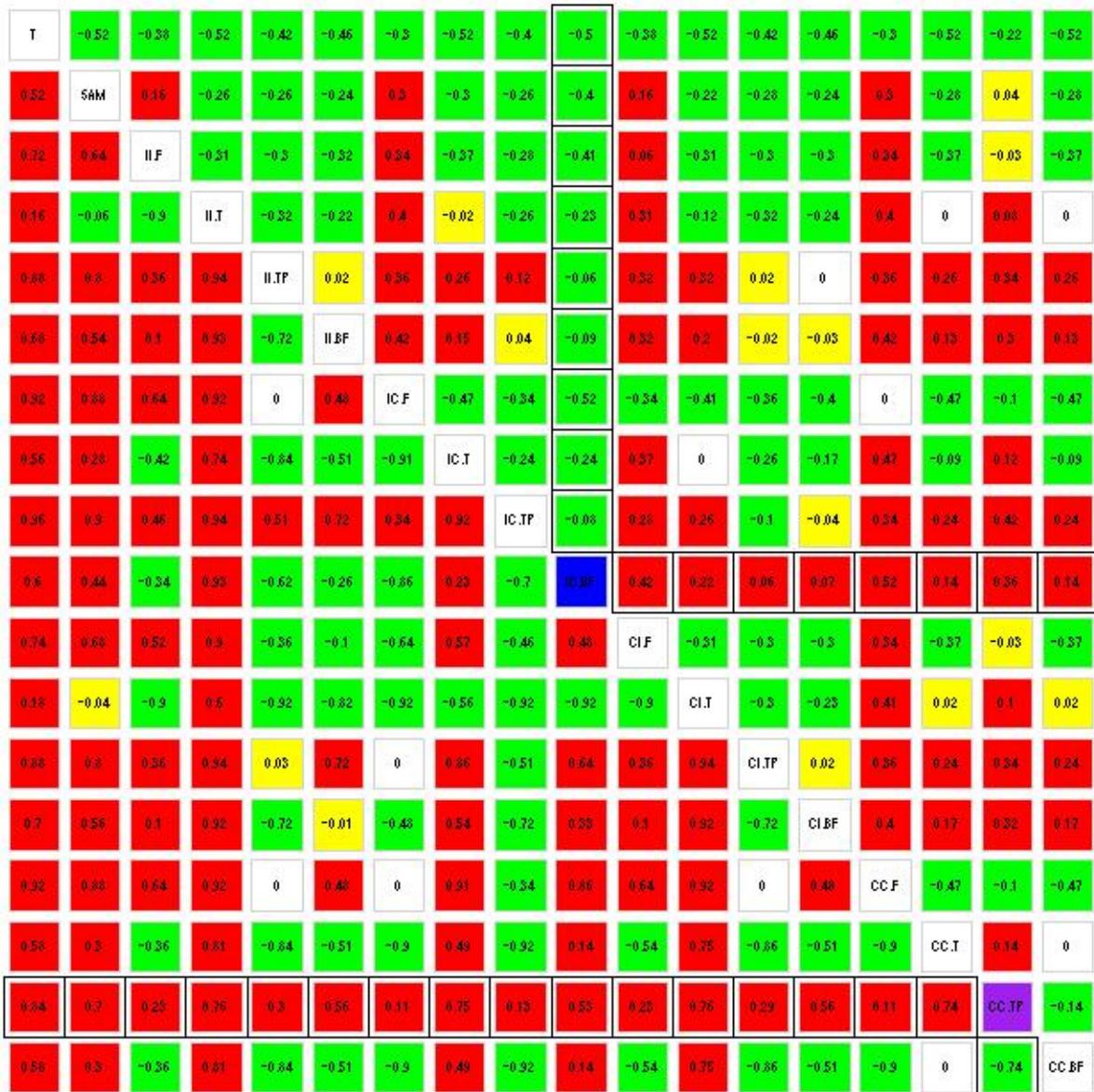


Figure 31: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

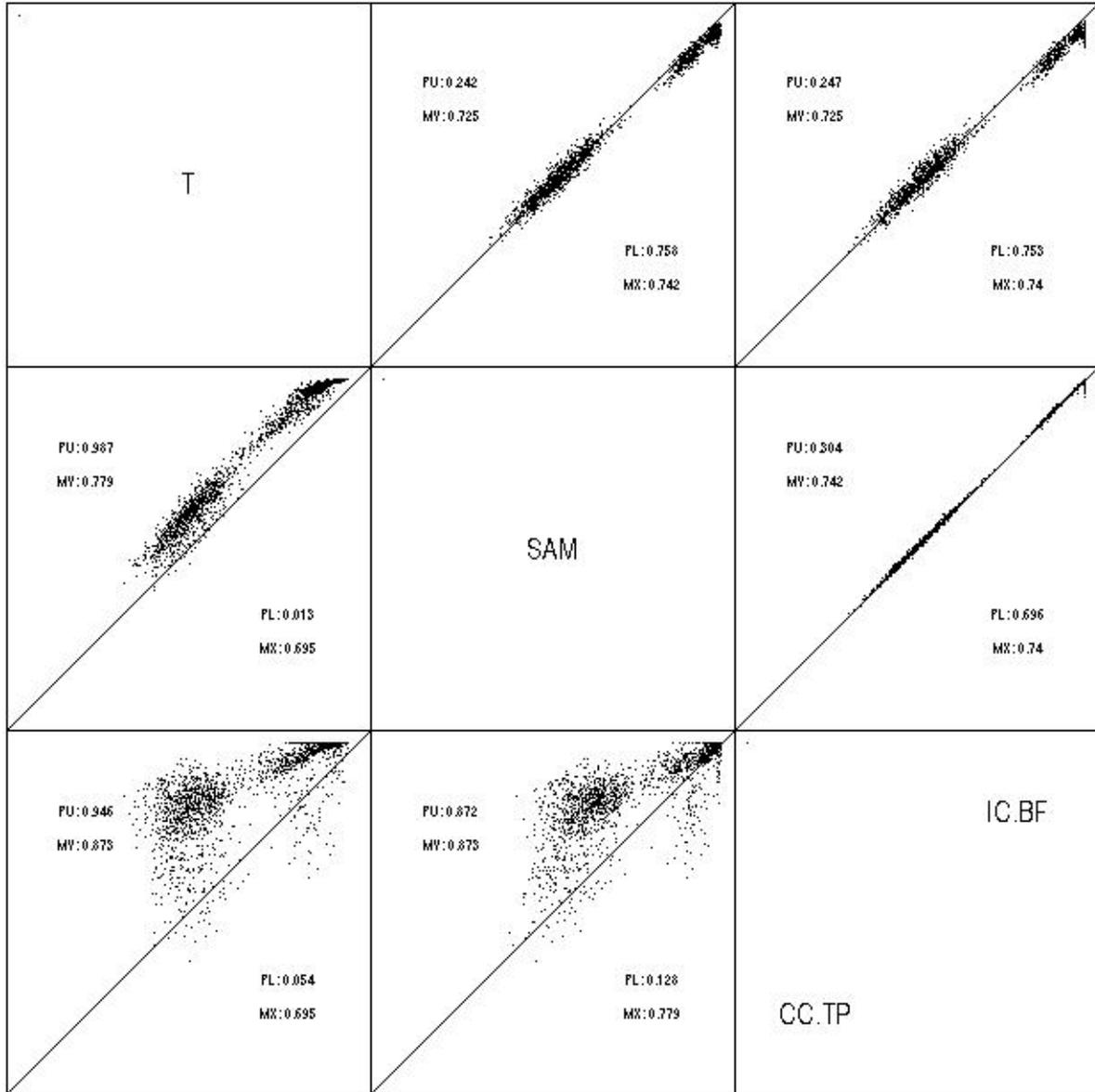


Figure 32: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 2% genes to be differential genes. Statistics based on multilevel genes are calculated with true hyperparameters plugged in.

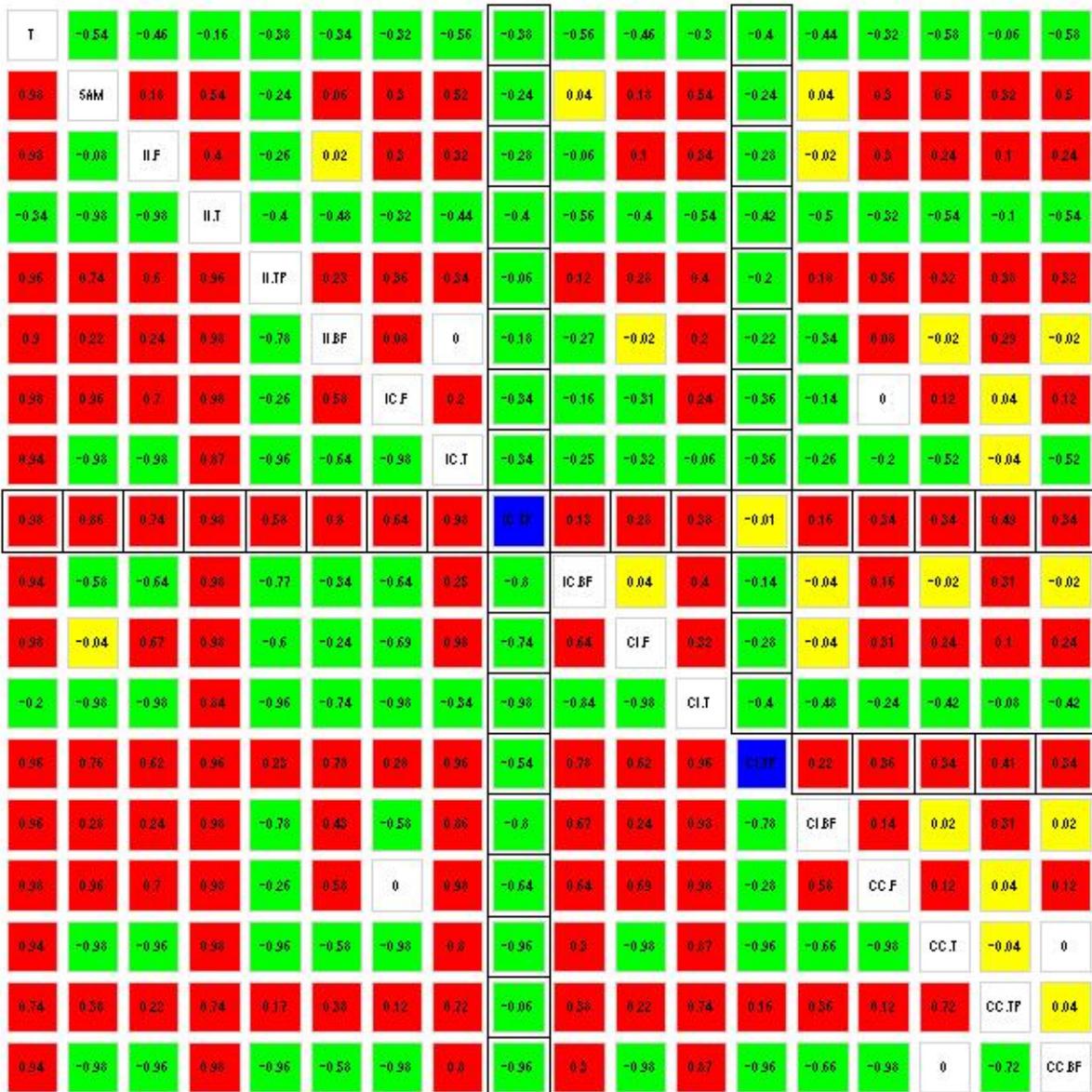


Figure 33: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

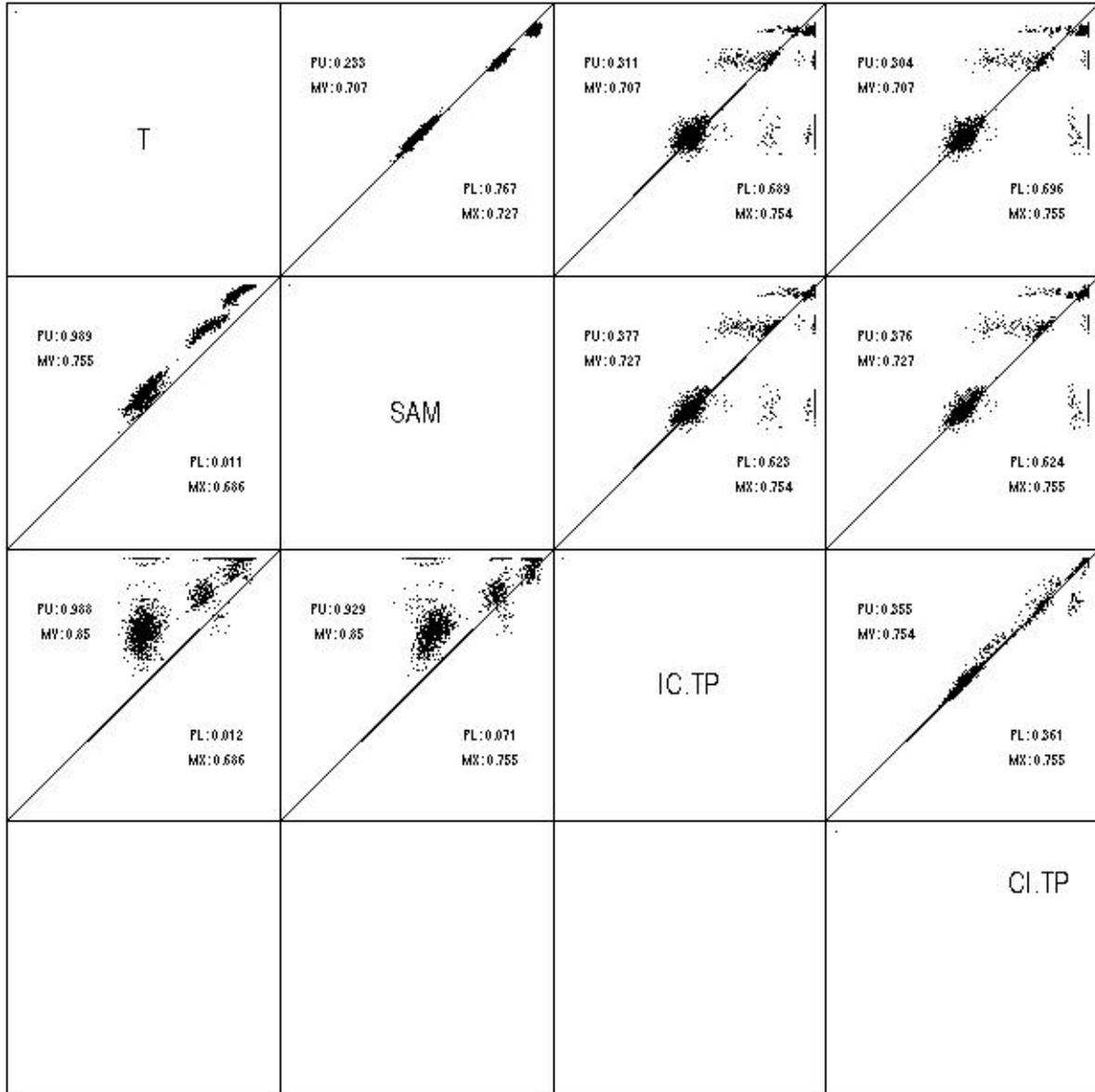


Figure 34: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

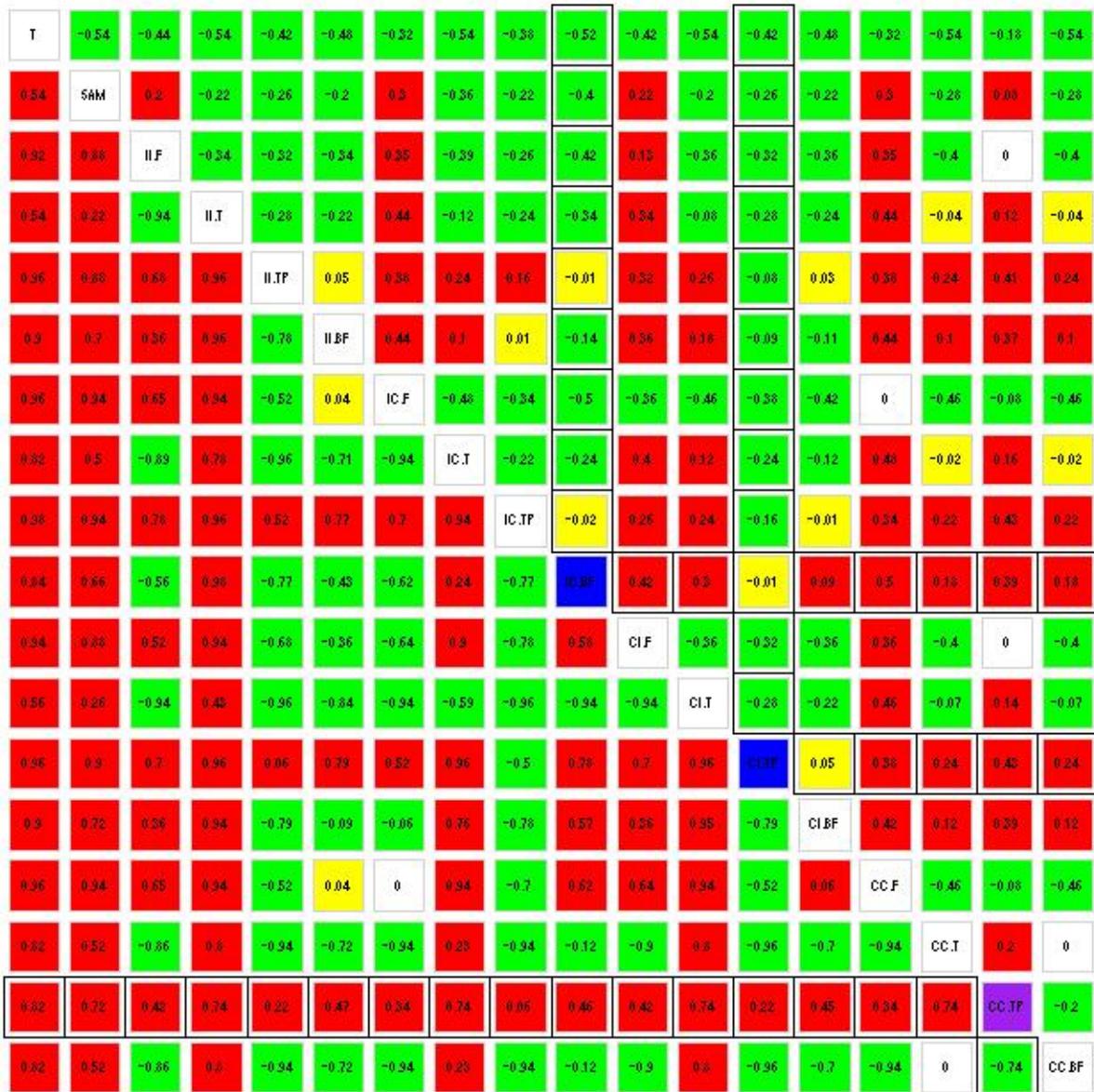


Figure 35: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

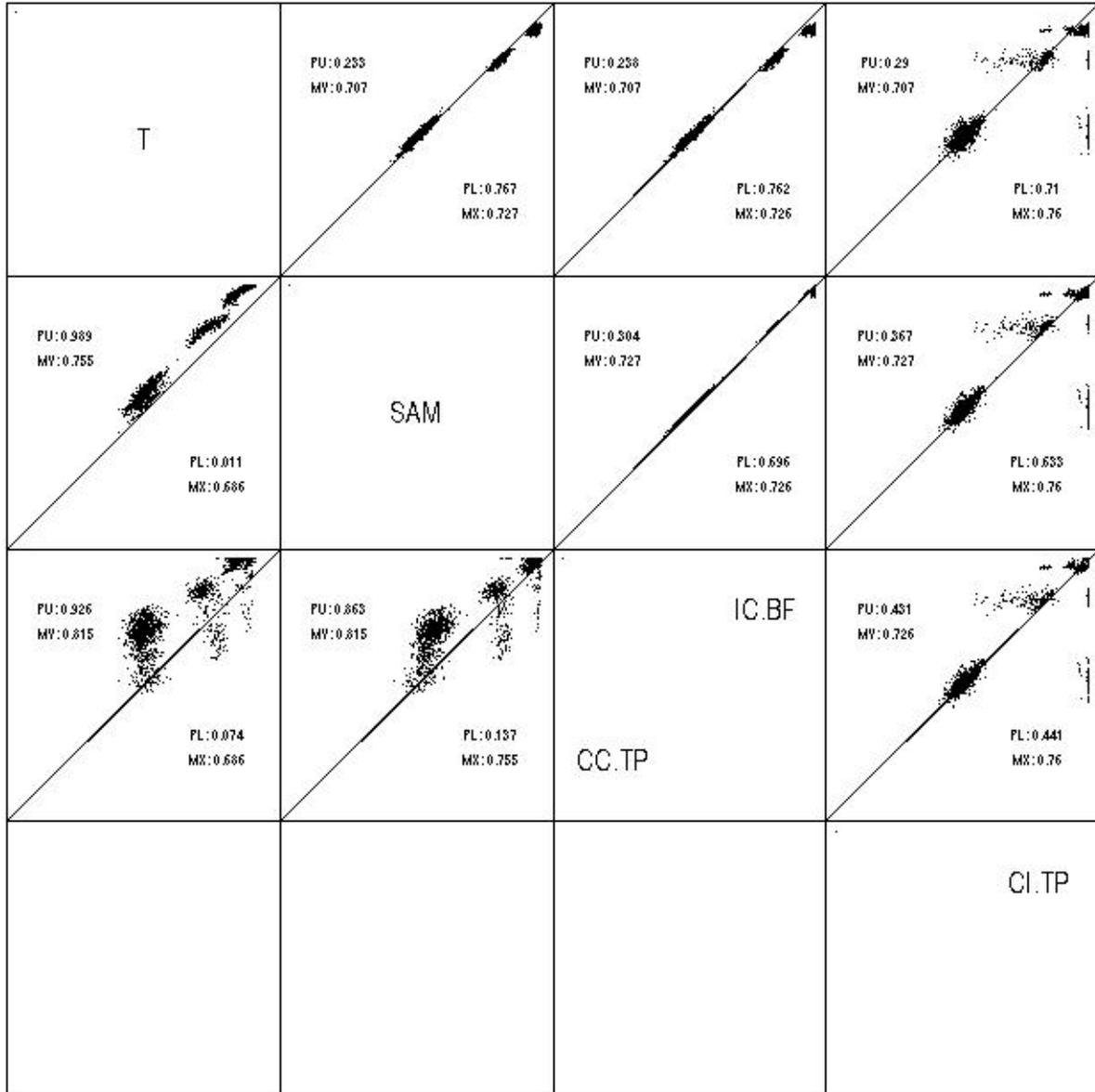


Figure 36: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 3 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

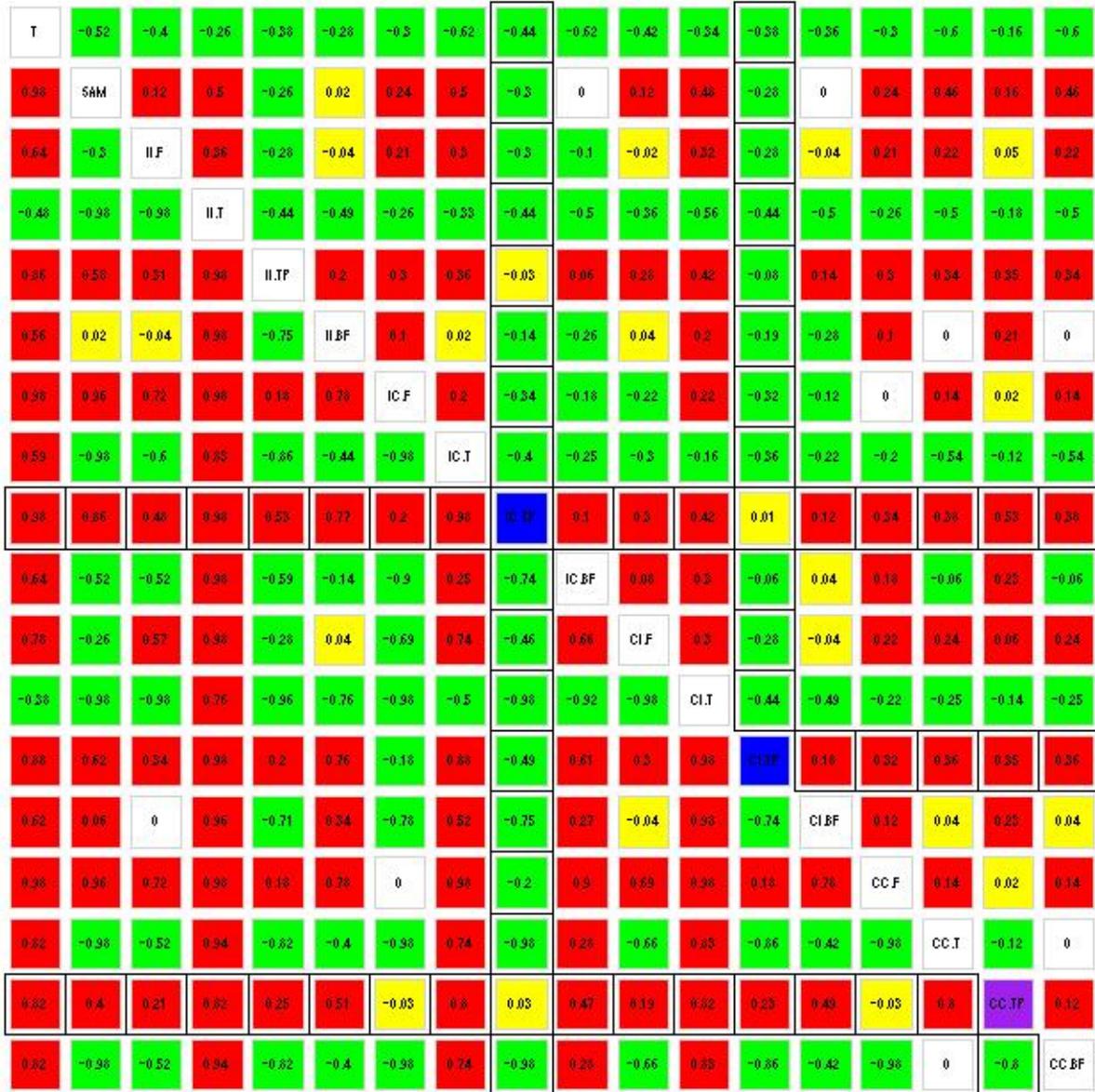


Figure 37: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 3 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

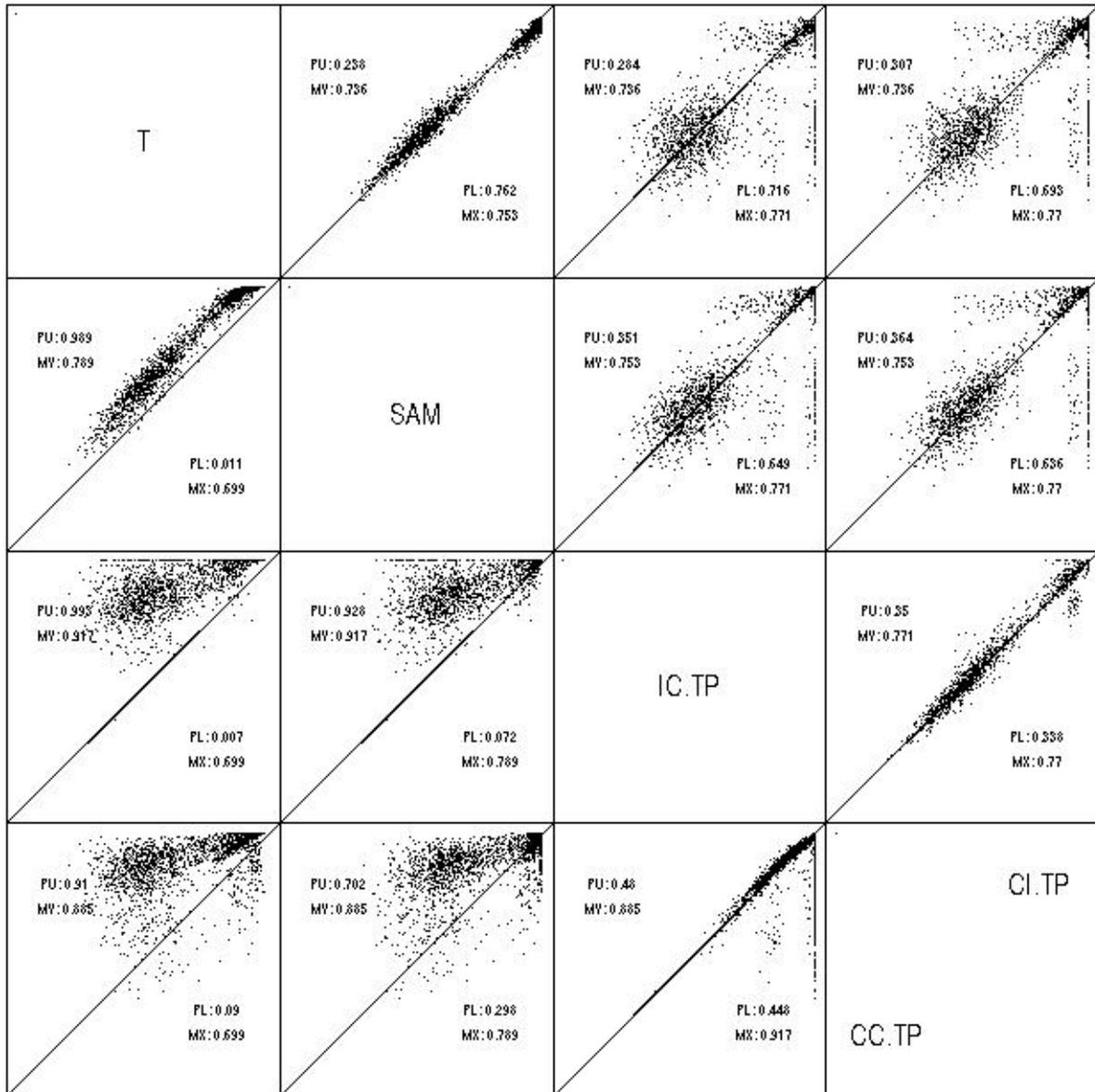


Figure 38: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 3 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

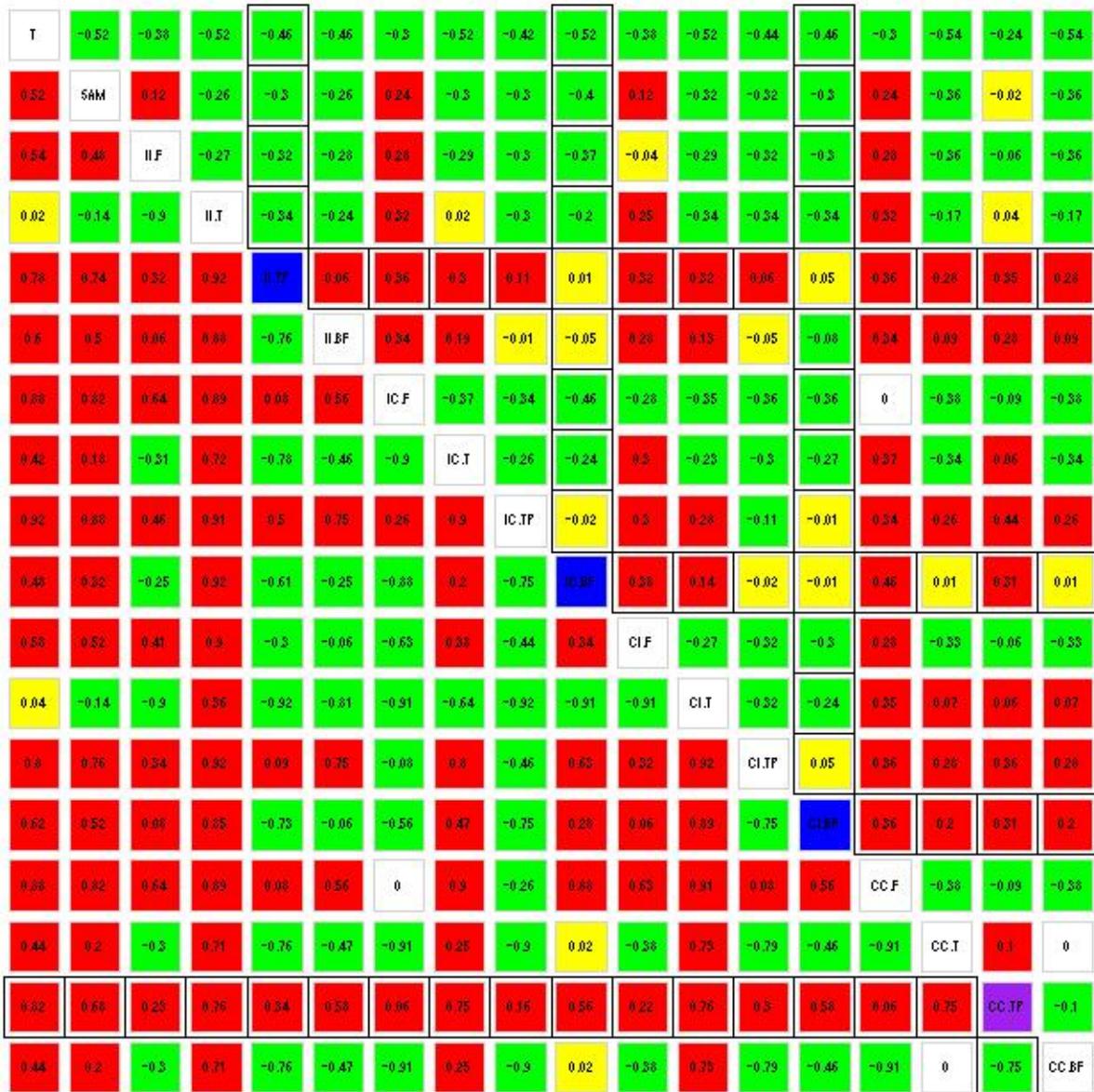


Figure 39: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 3 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

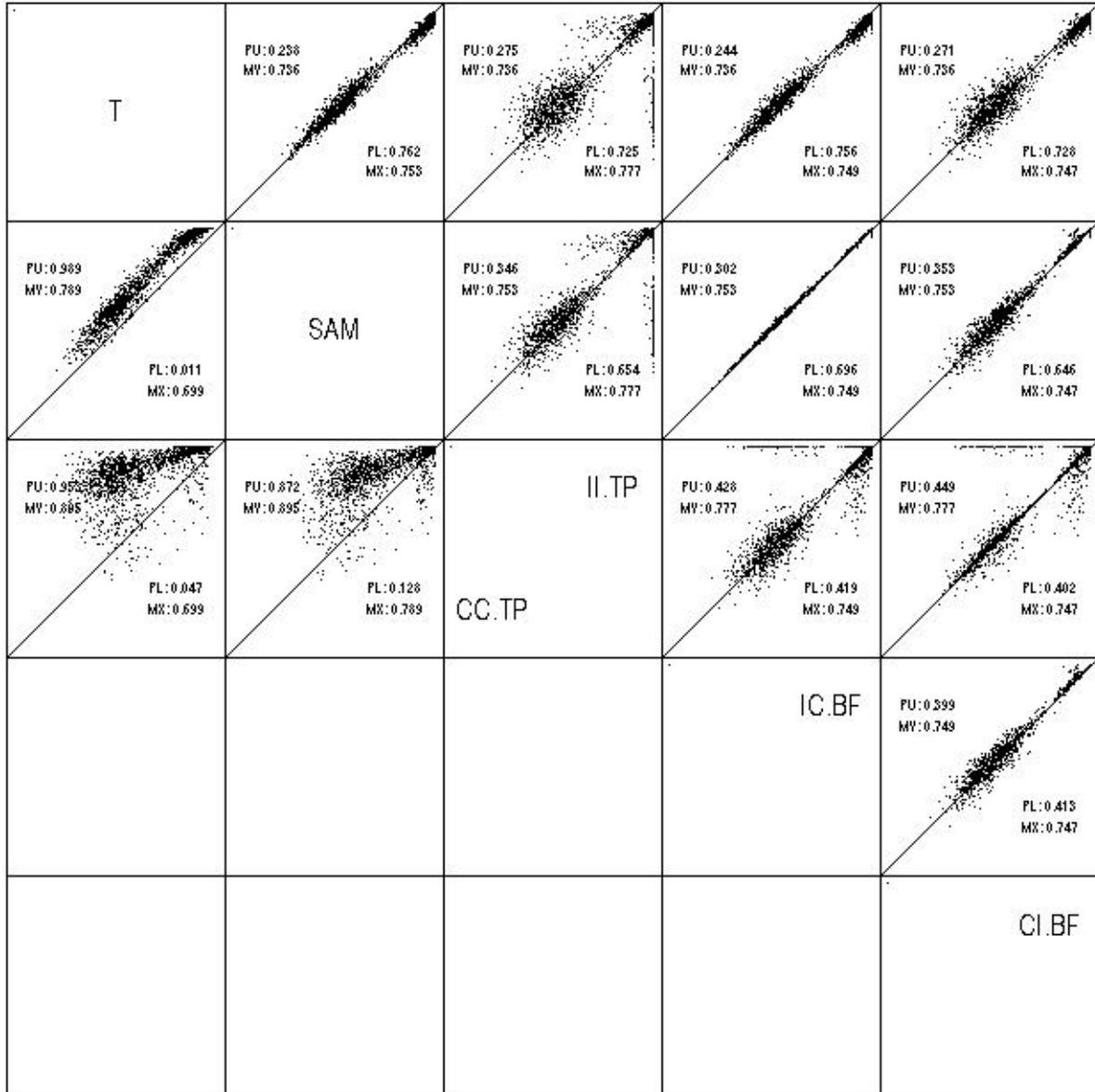


Figure 40: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 3 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

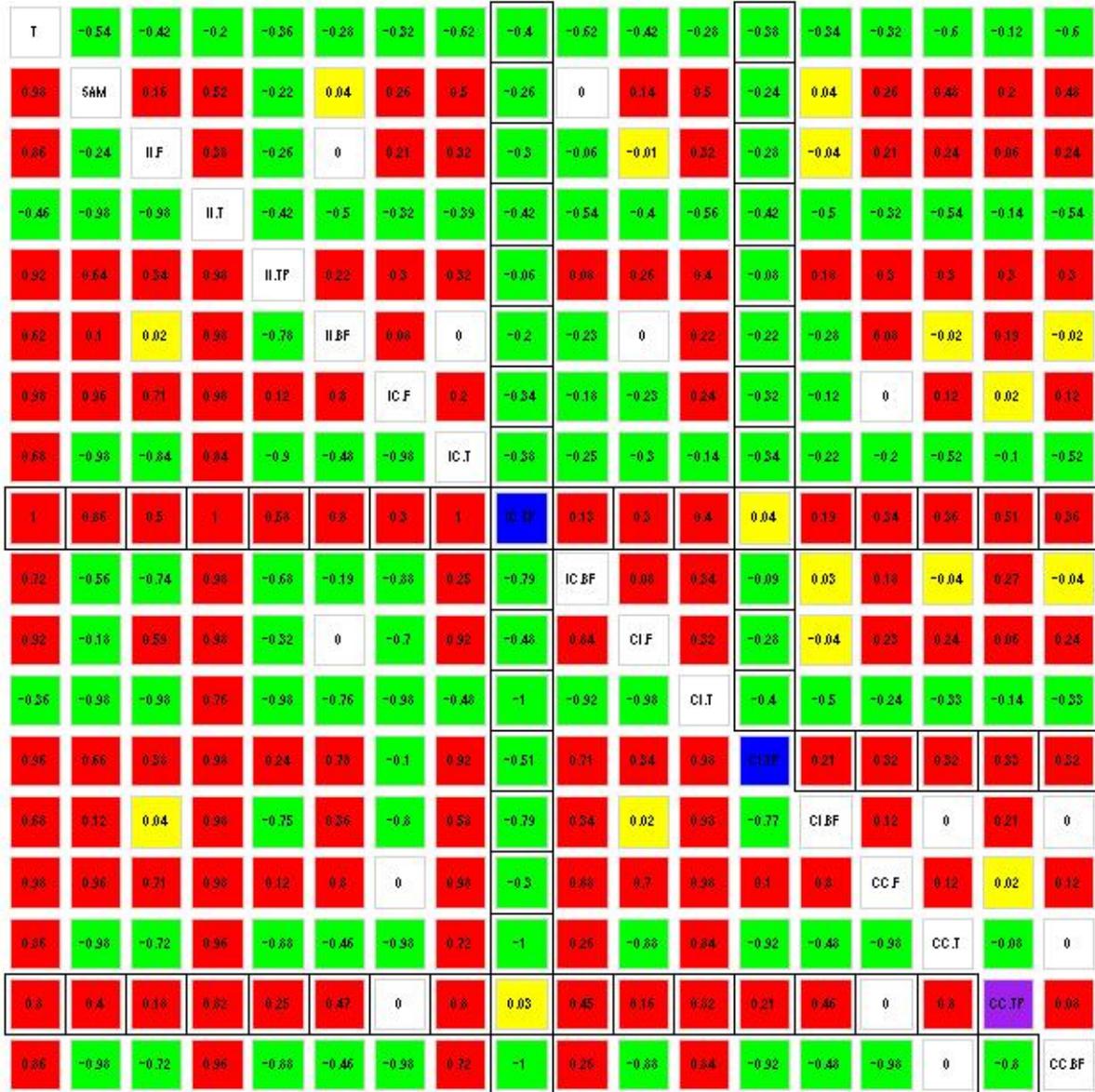


Figure 41: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 3 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

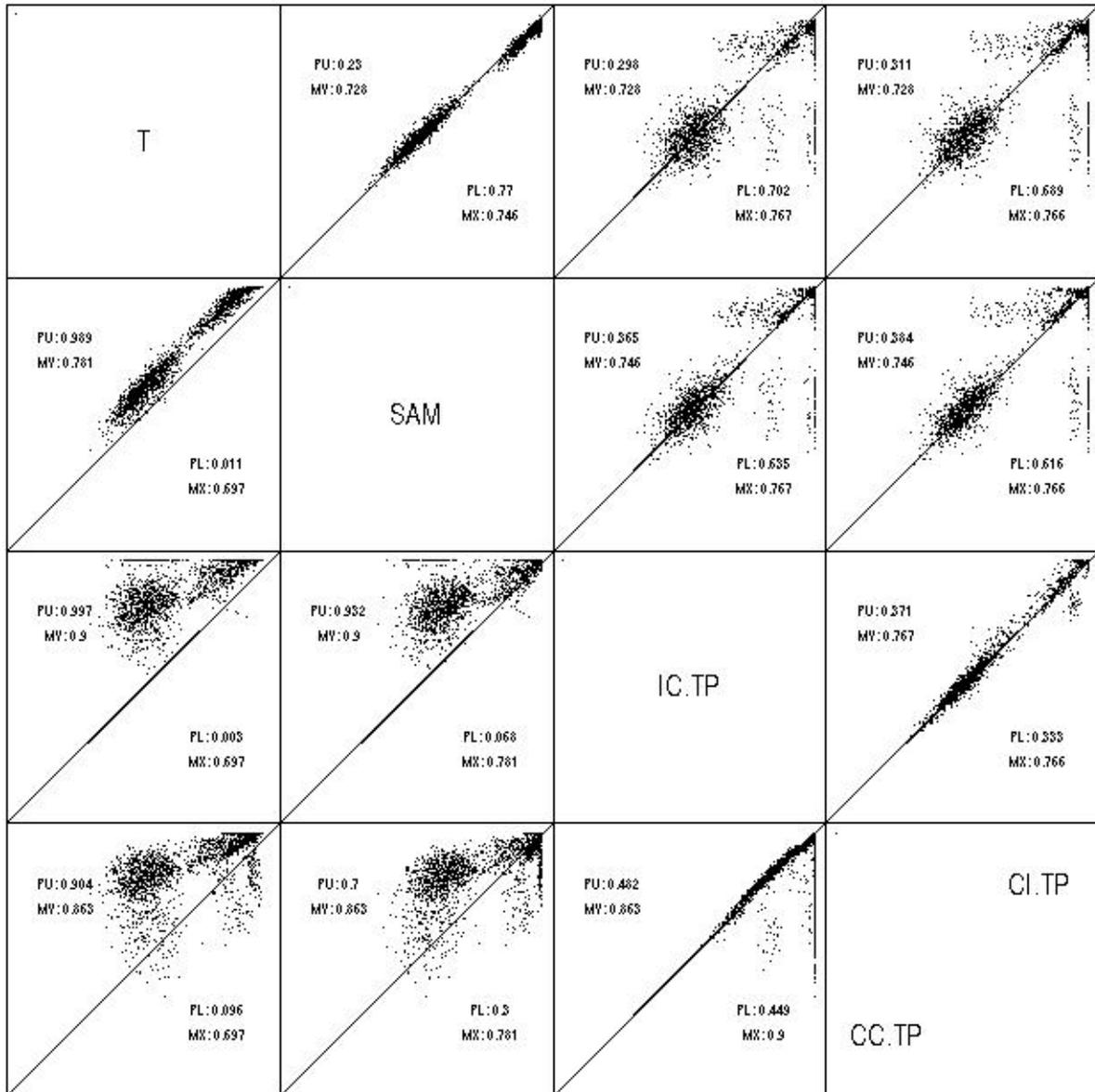


Figure 42: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 3 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

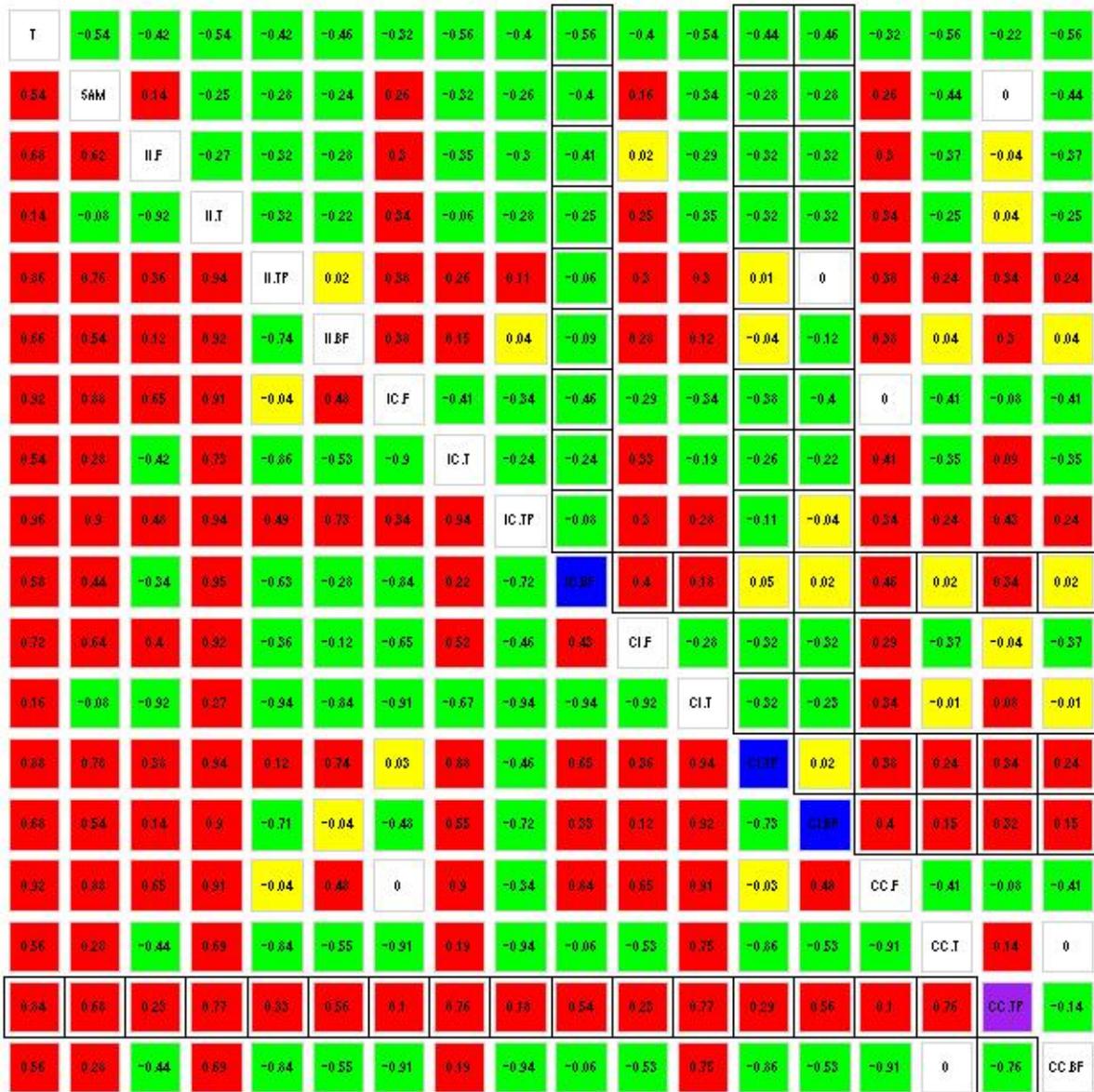


Figure 43: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 3 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

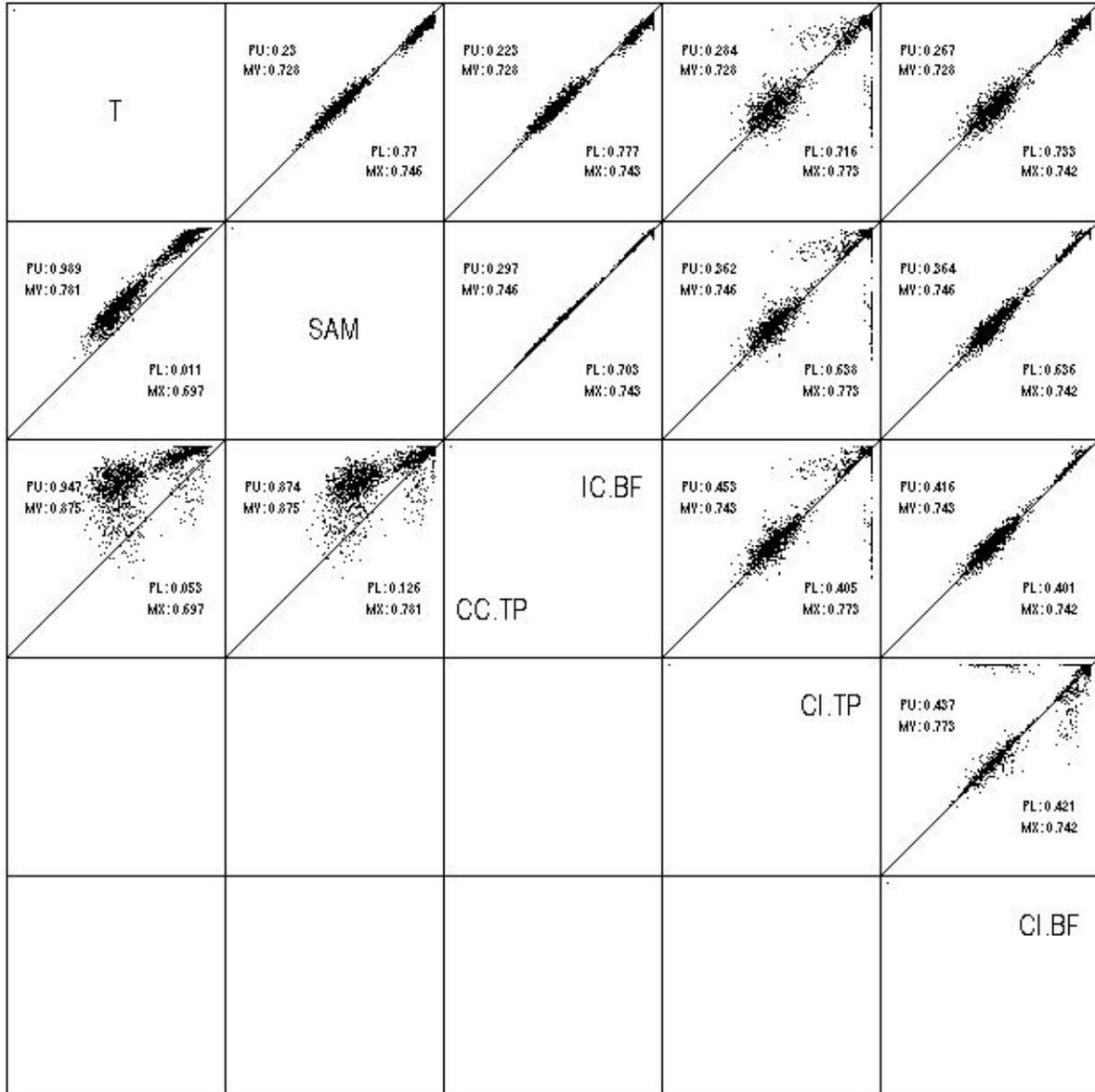


Figure 44: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 3 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

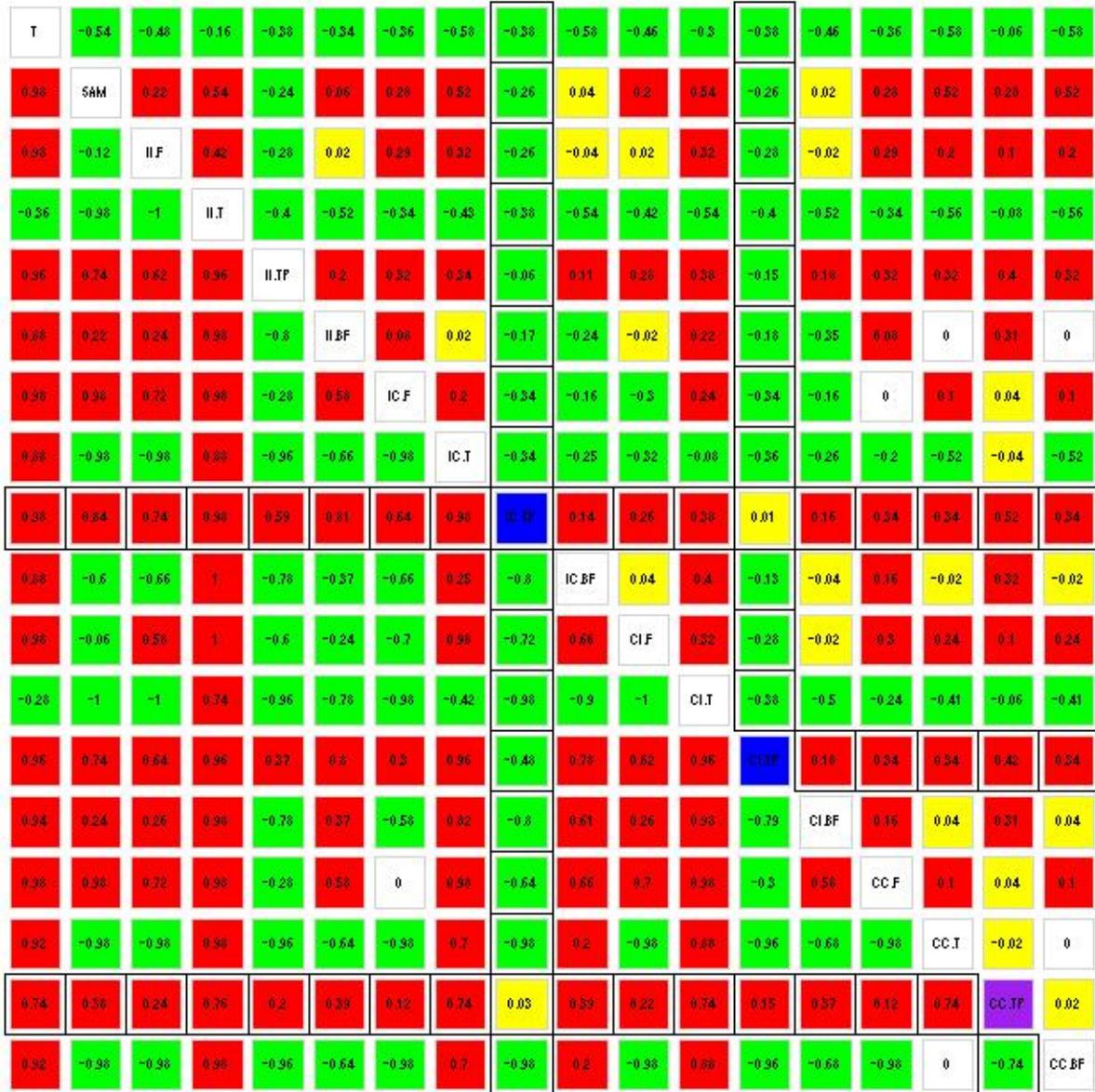


Figure 45: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 3 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

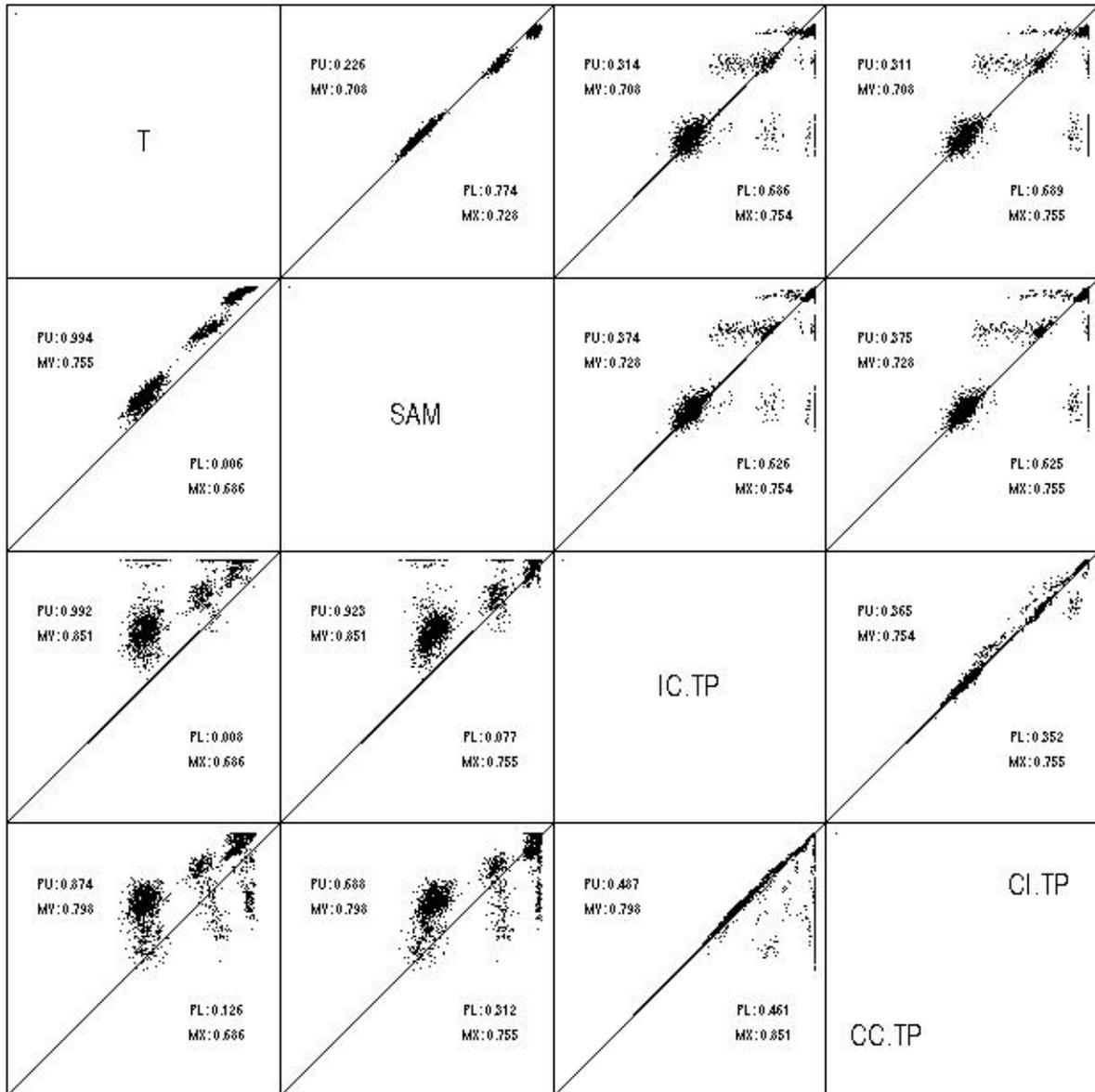


Figure 46: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 3 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

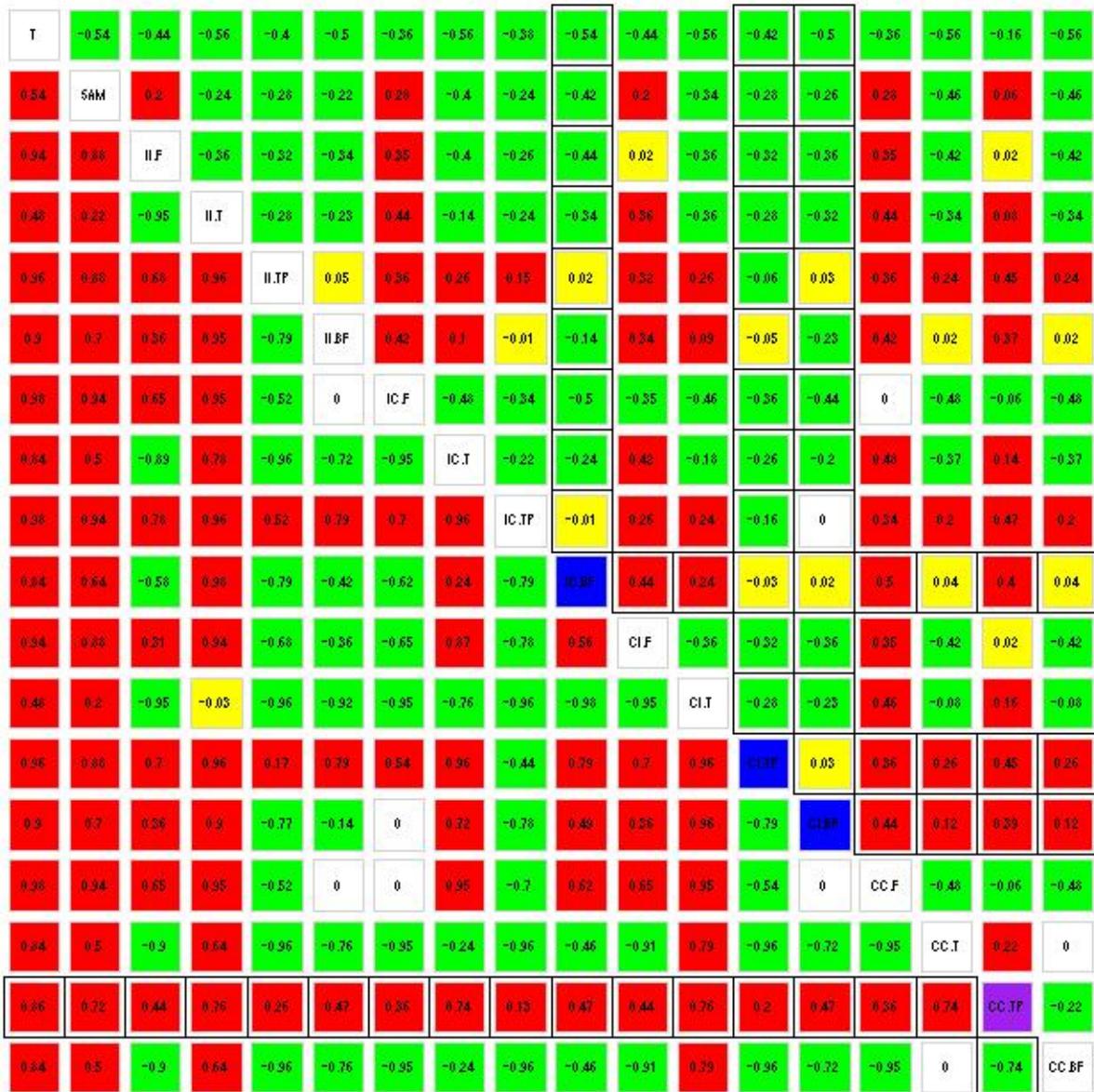


Figure 47: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 3 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

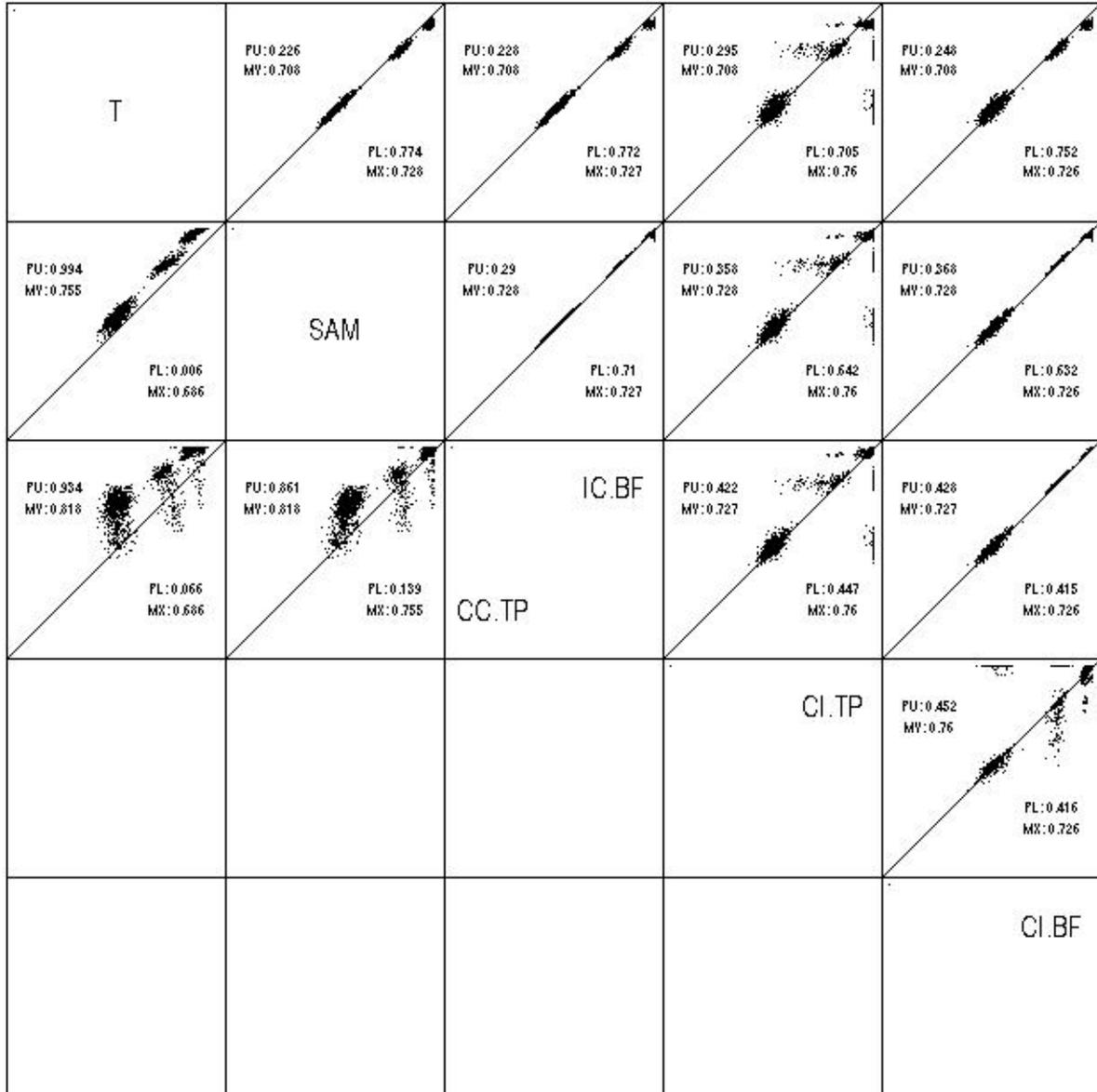


Figure 48: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 3 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

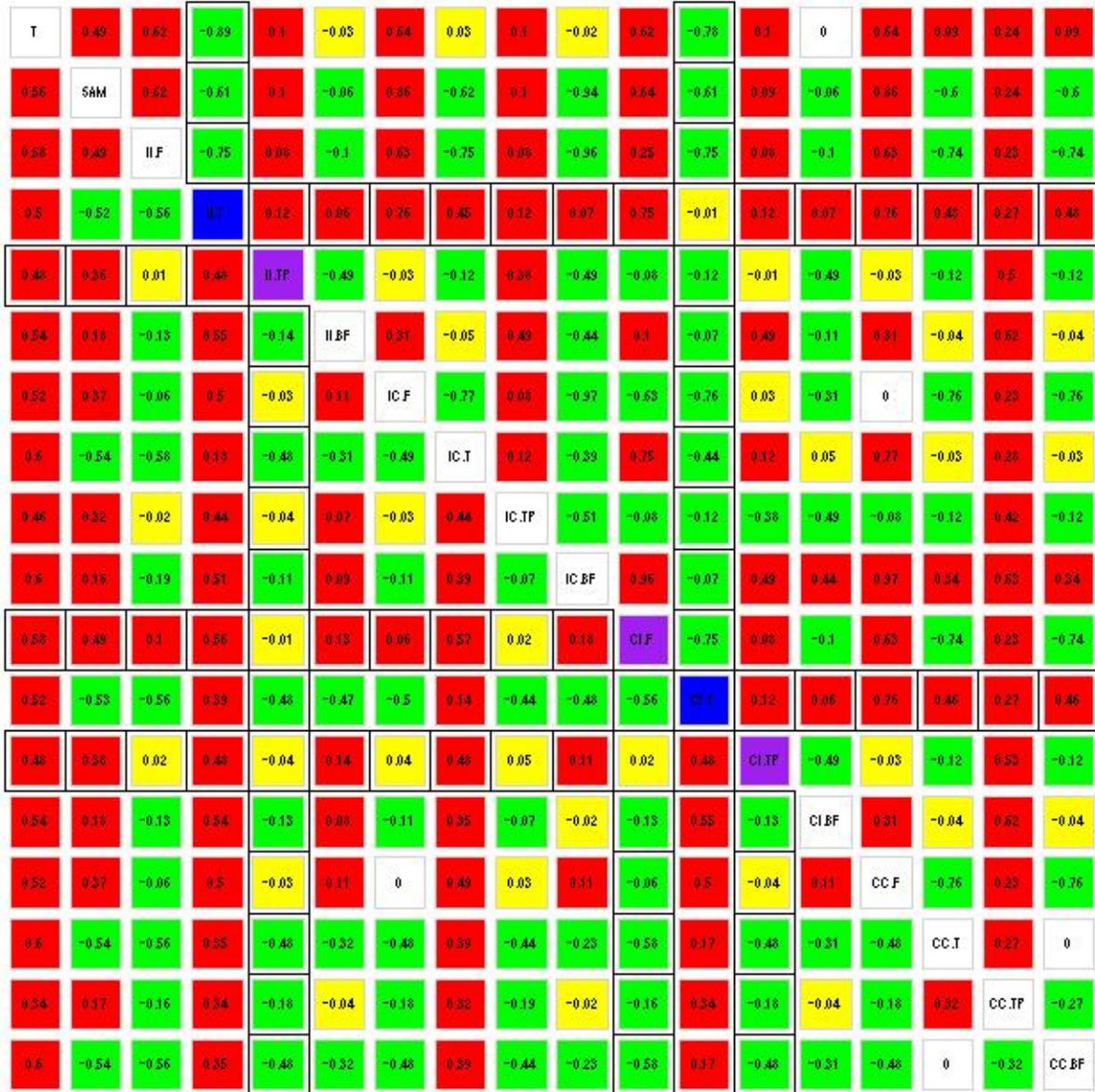


Figure 49: Heat map summary for simulations based on Independence model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

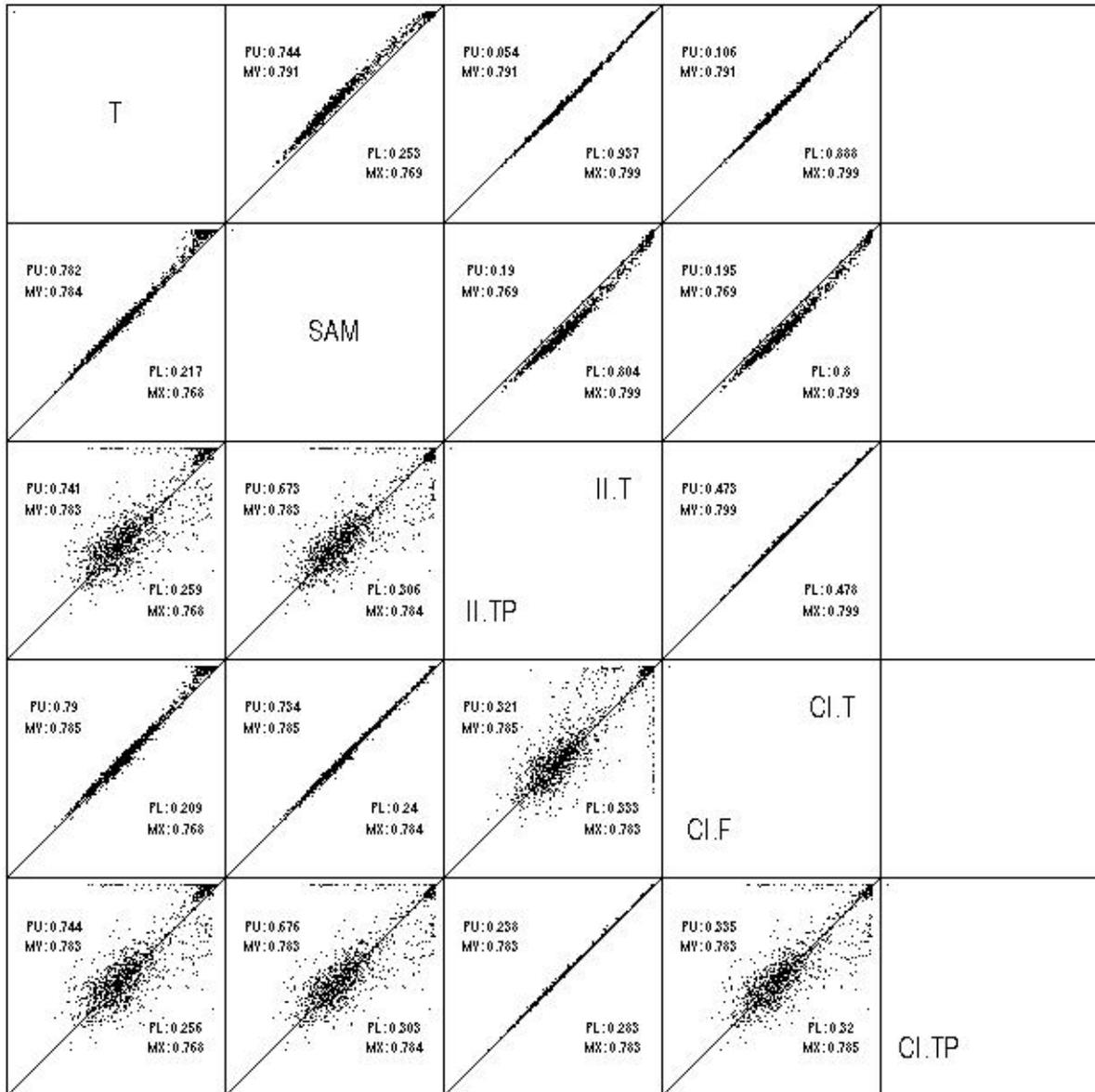


Figure 50: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

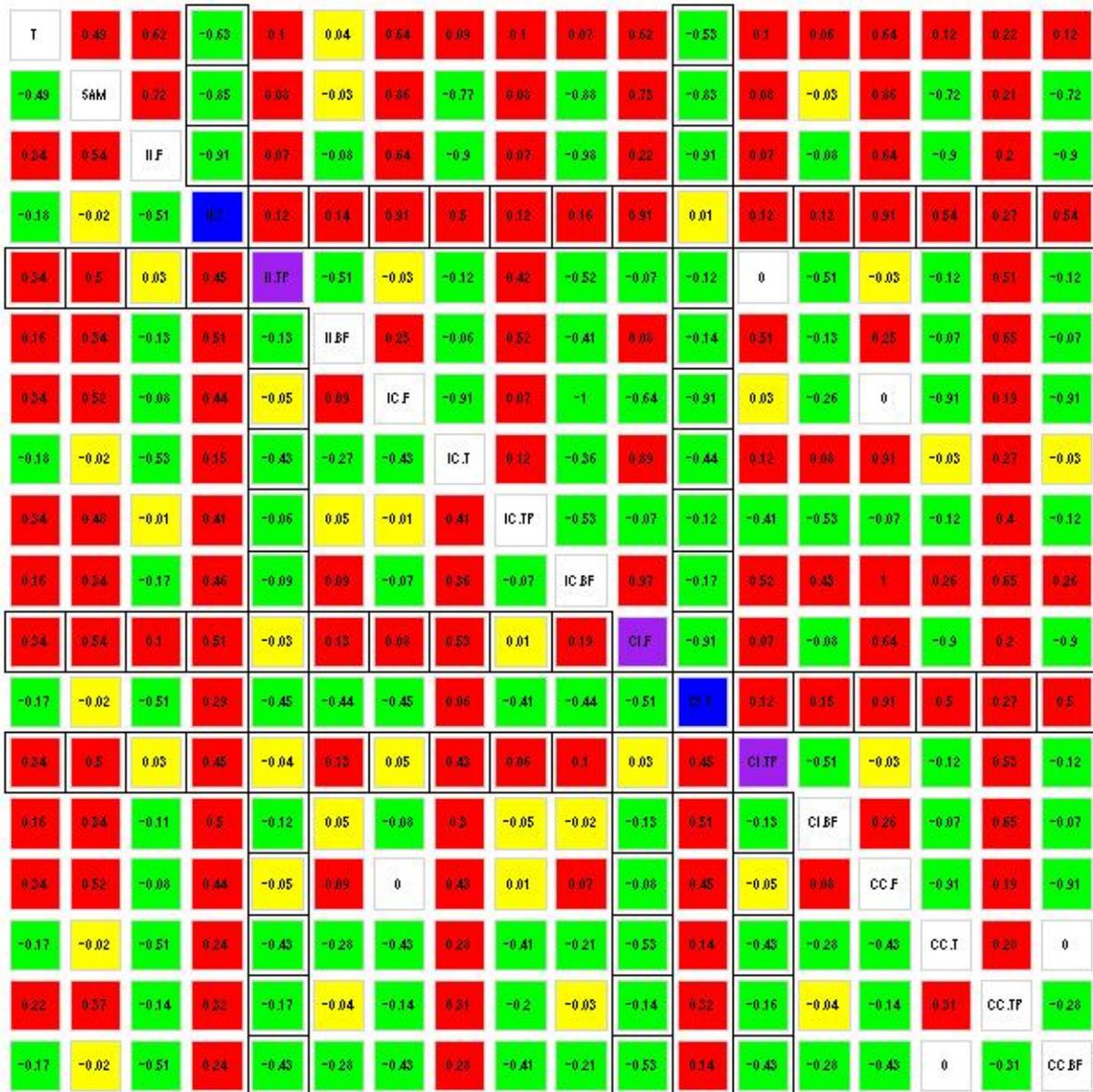


Figure 51: Heat map summary for simulations based on Independence model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

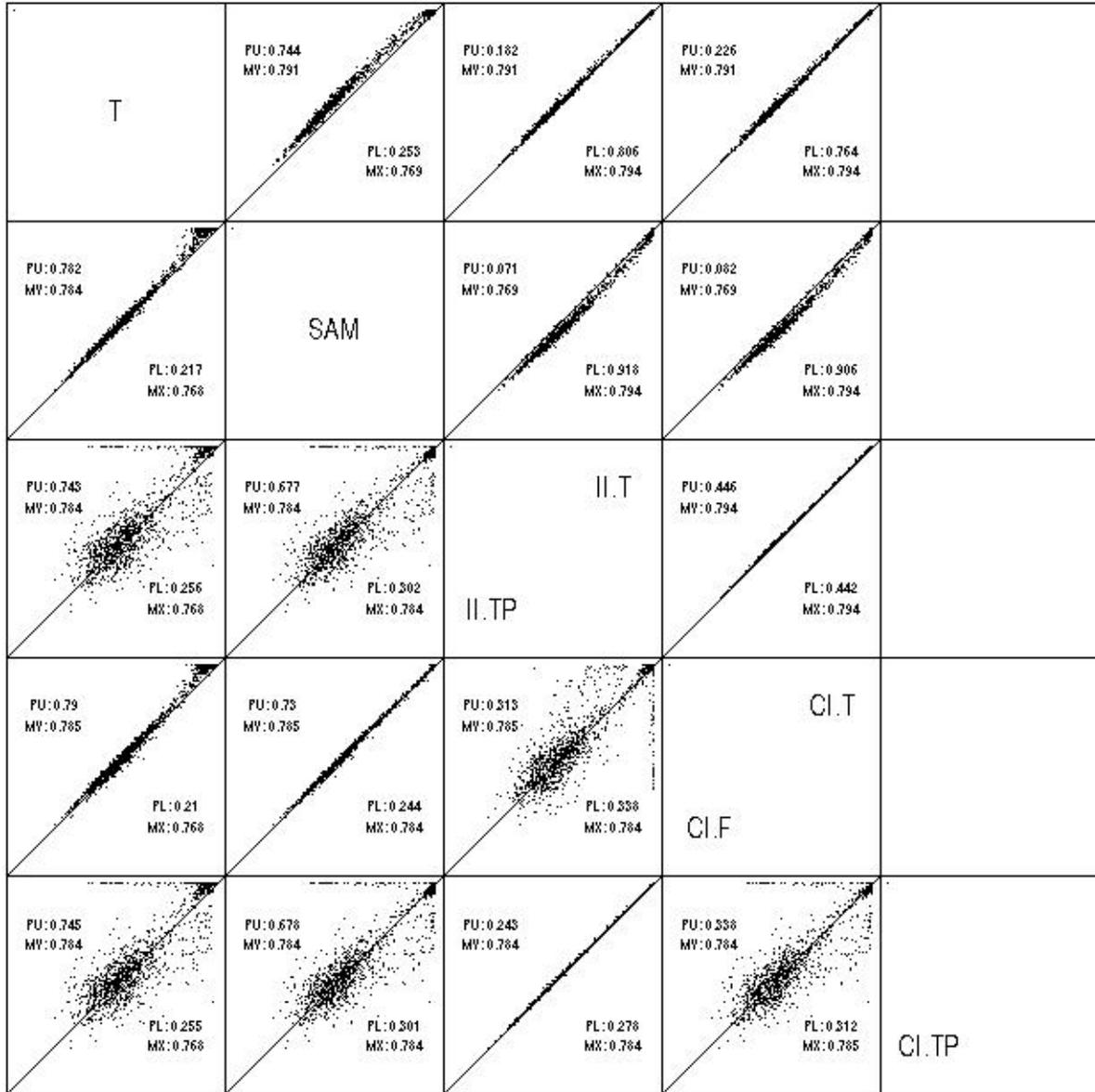


Figure 52: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

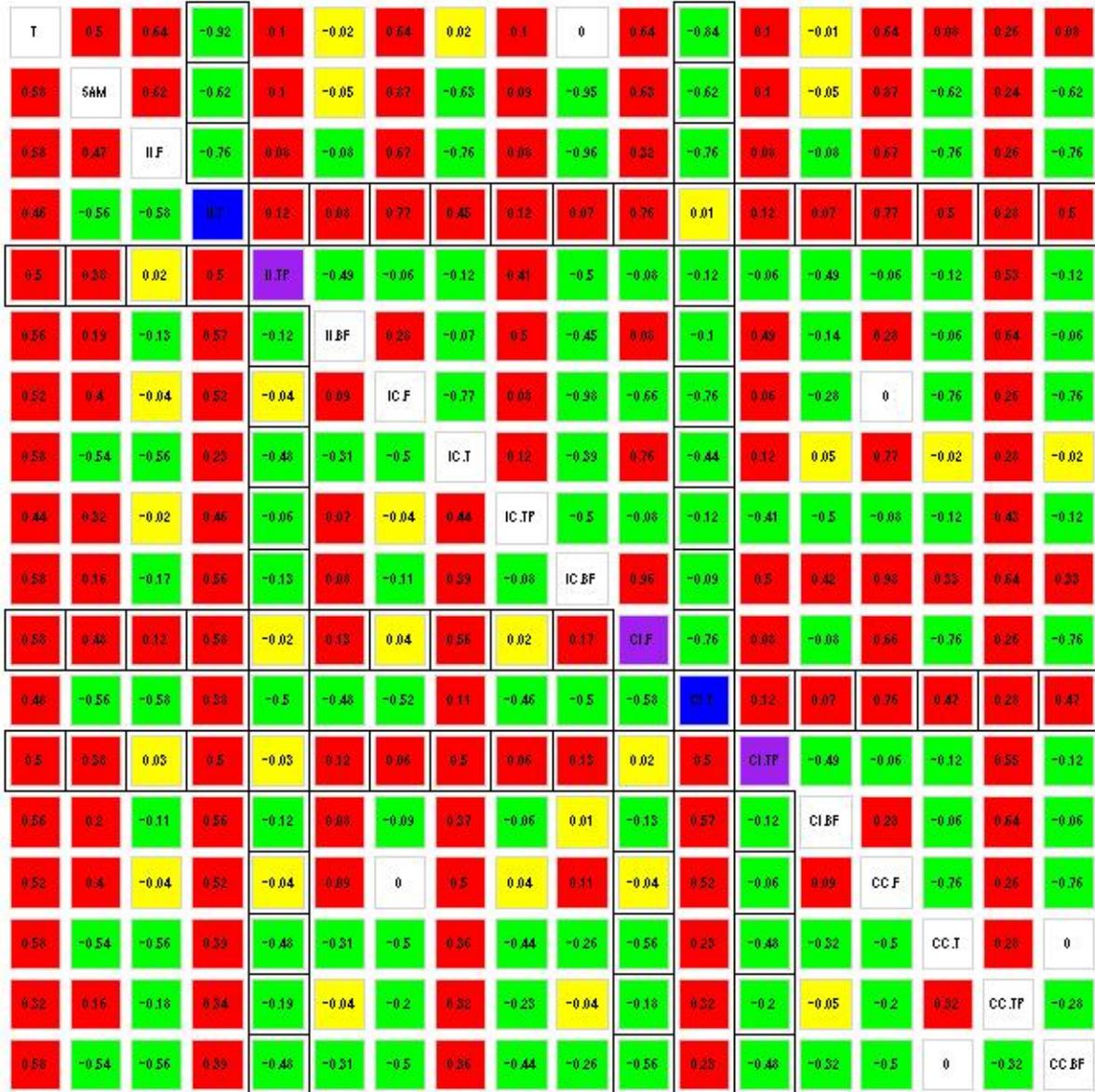


Figure 53: Heat map summary for simulations based on Independence model with number of replicates equal to 10 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

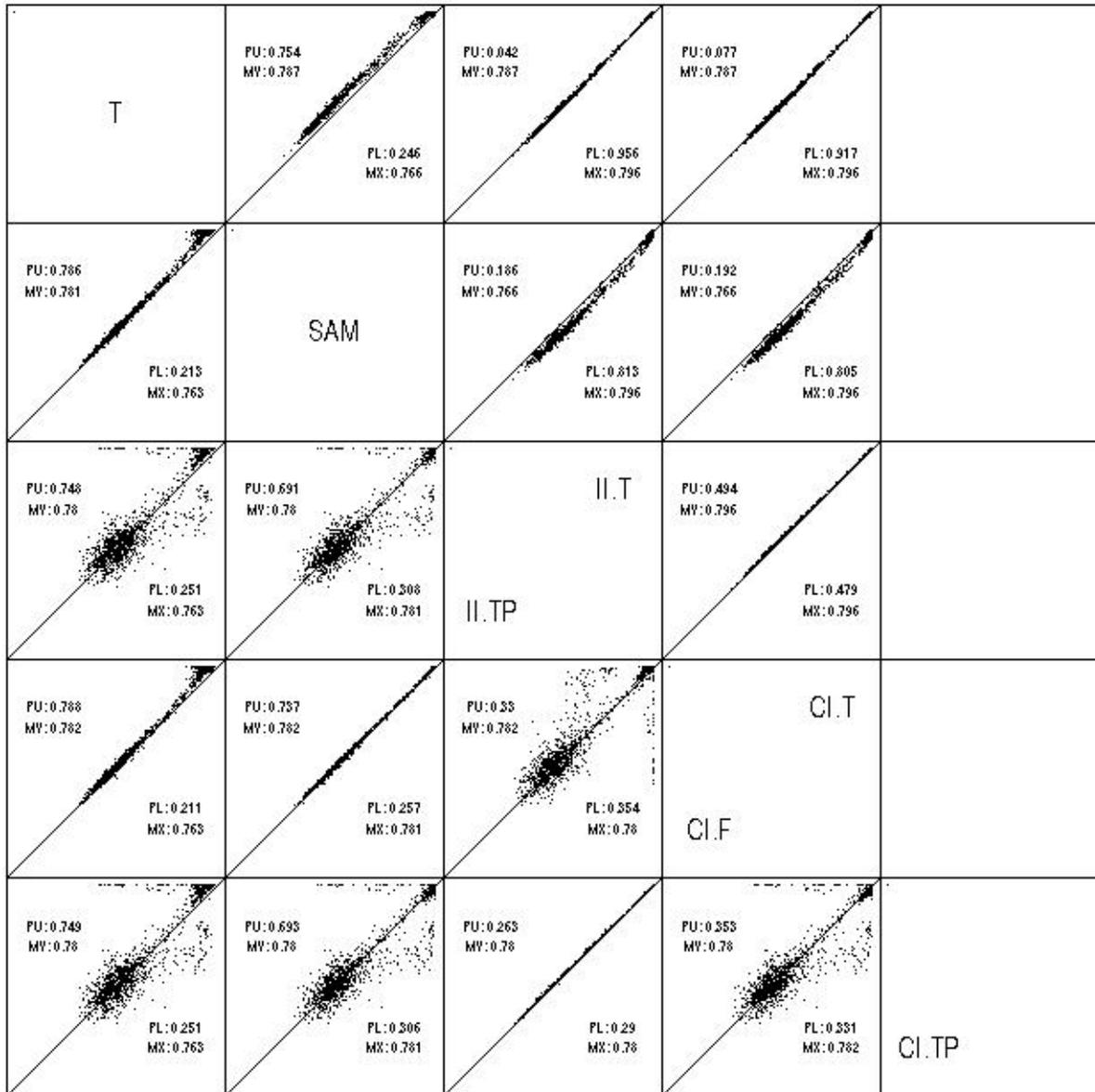


Figure 54: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 10 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

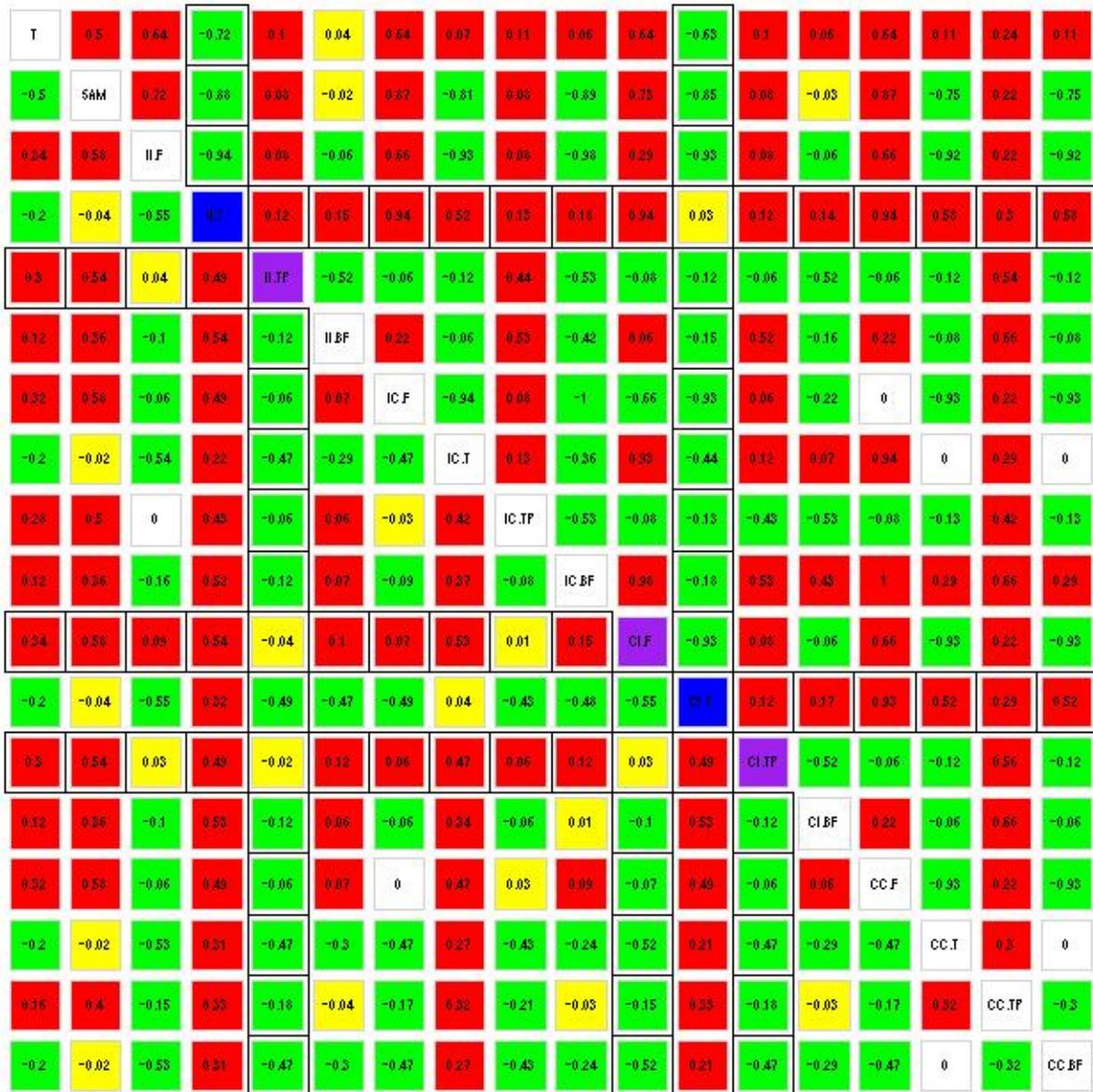


Figure 55: Heat map summary for simulations based on Independence model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

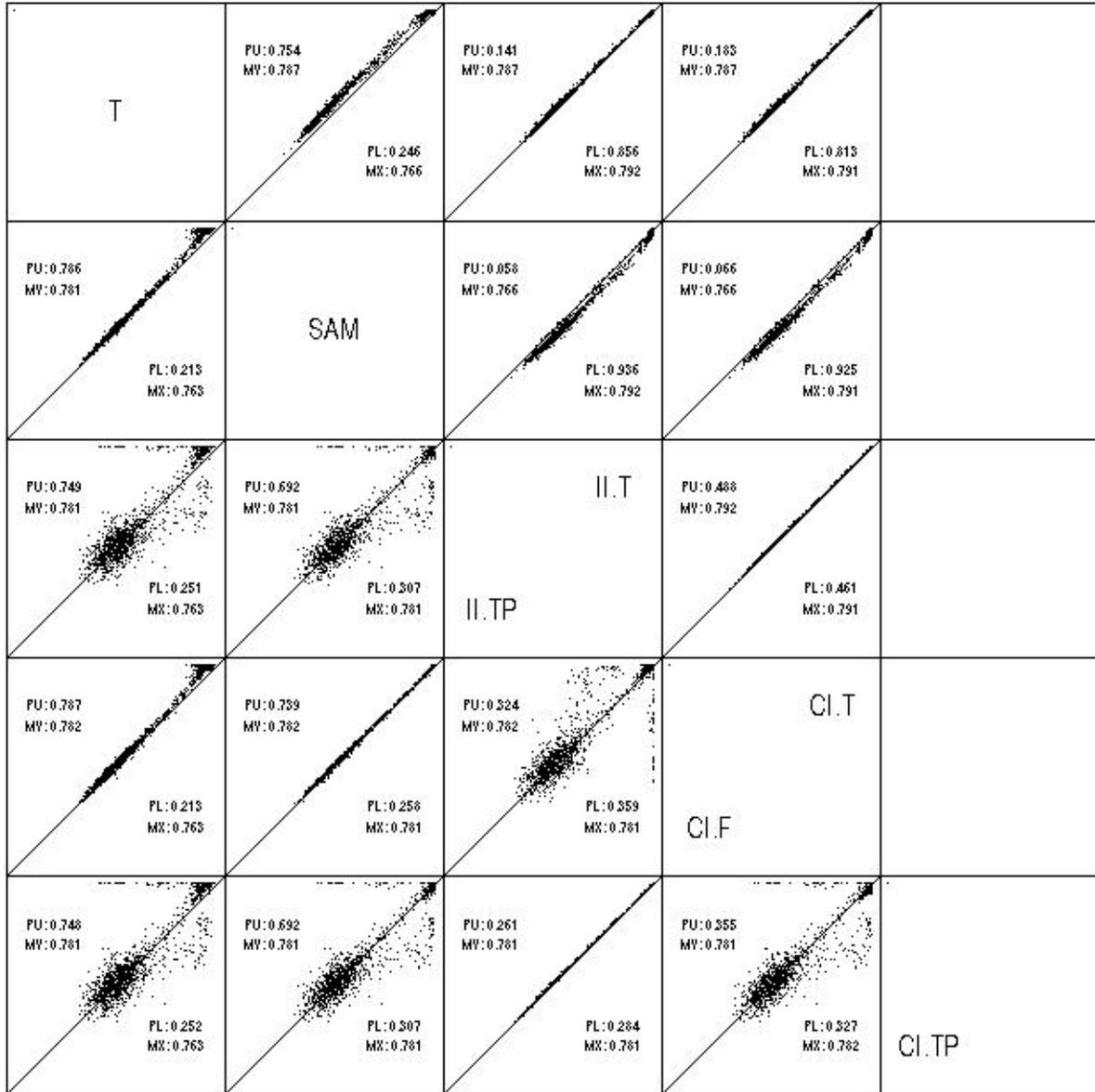


Figure 56: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

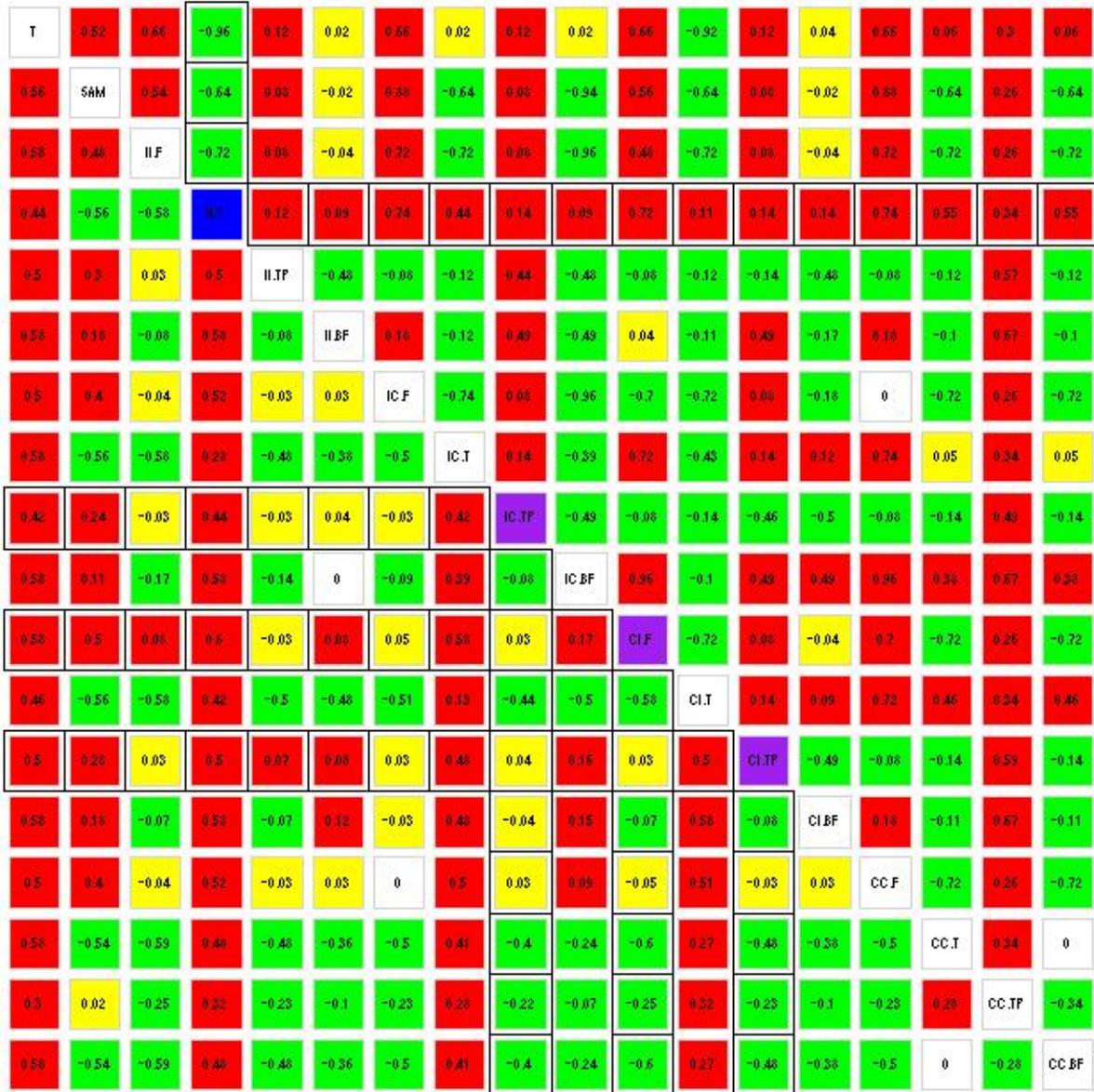


Figure 57: Heat map summary for simulations based on Independence model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

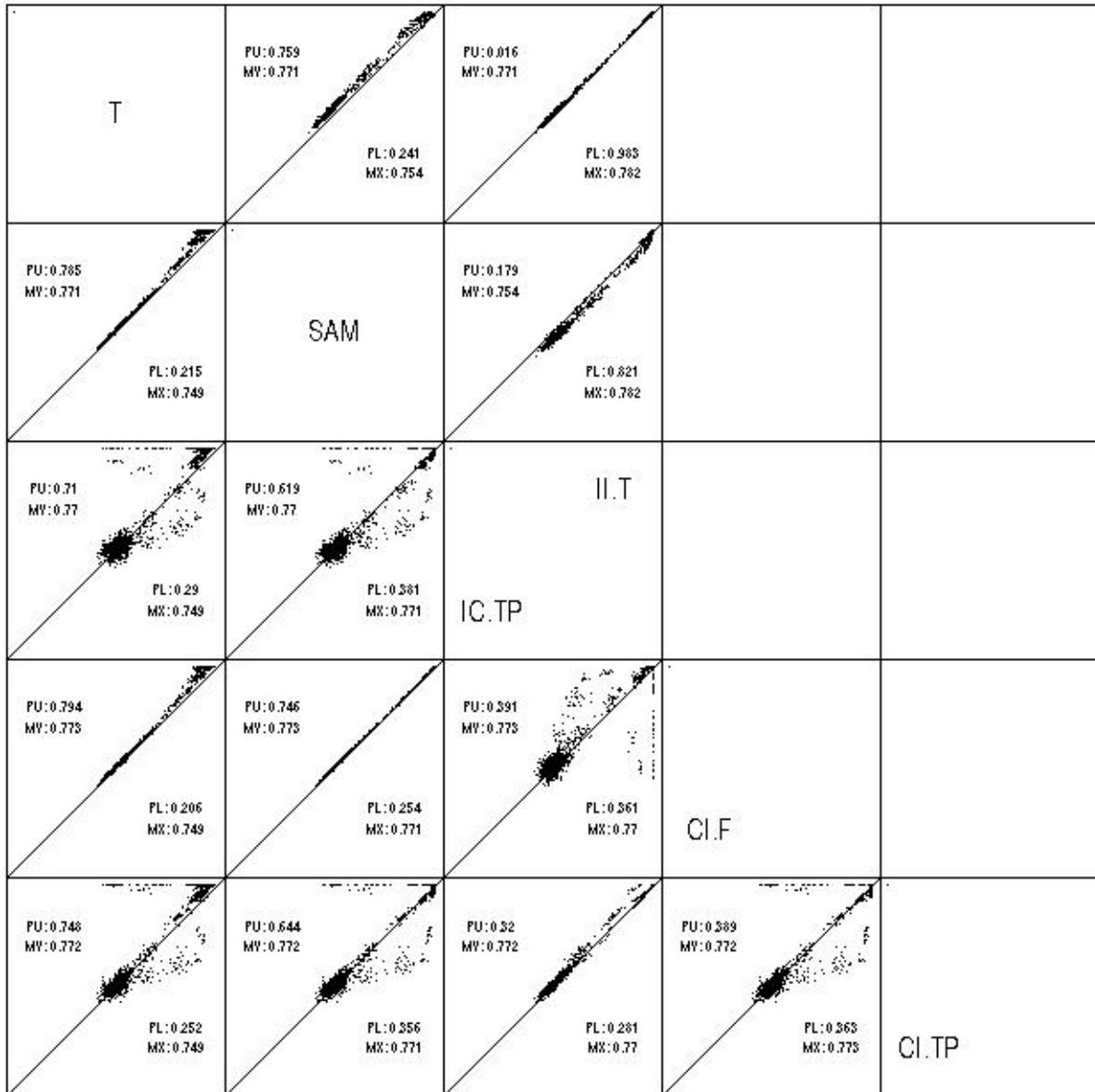


Figure 58: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

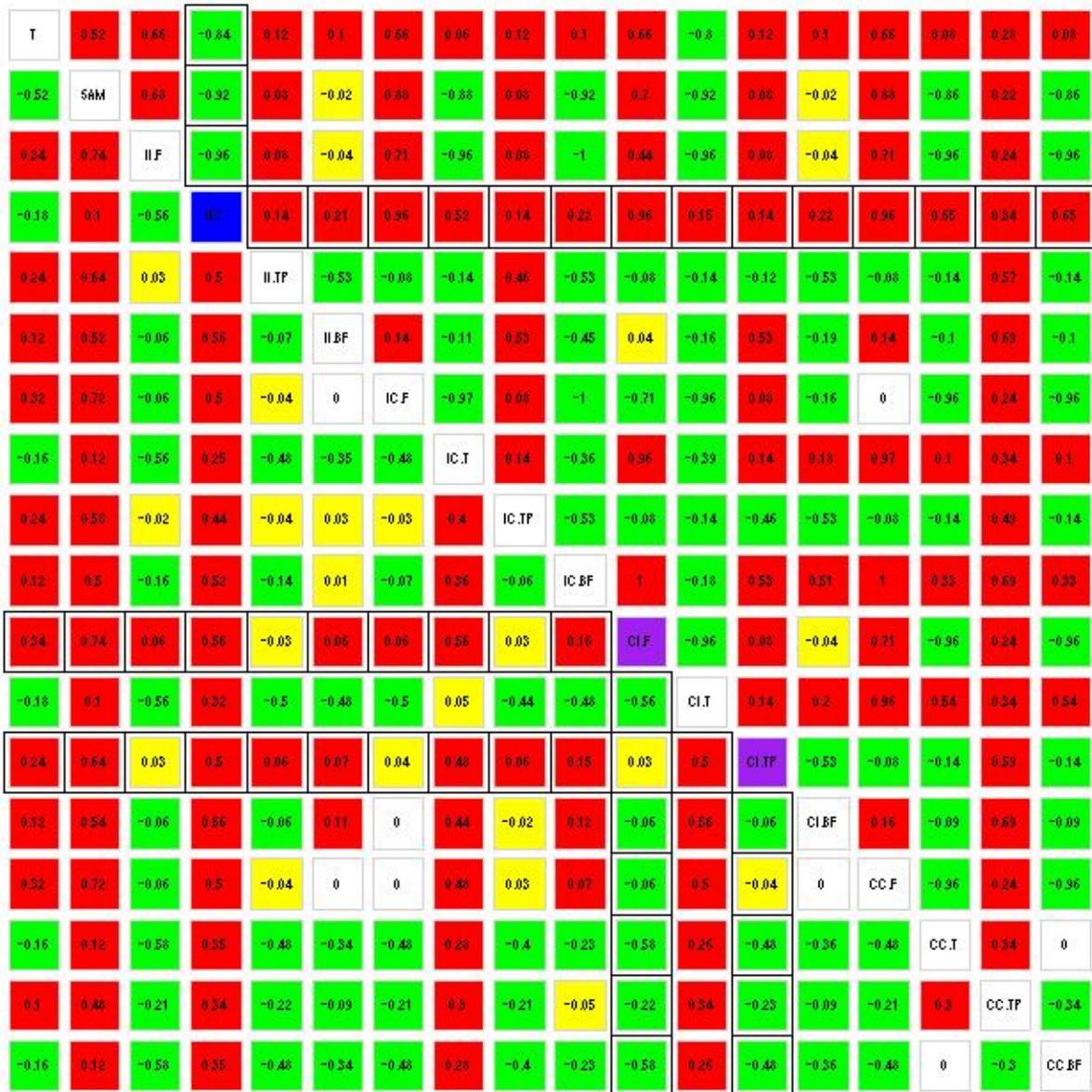


Figure 59: Heat map summary for simulations based on Independence model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

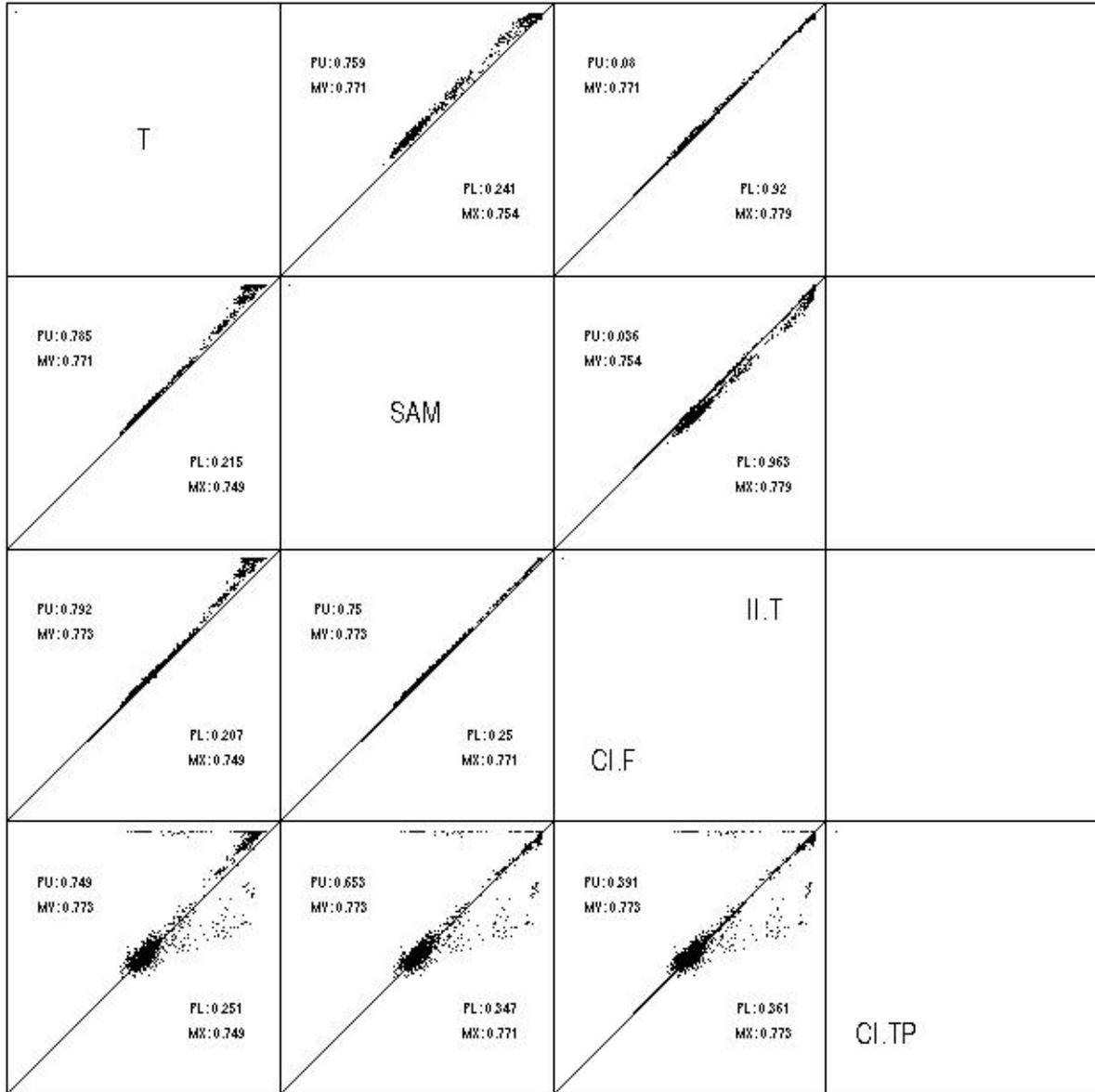


Figure 60: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

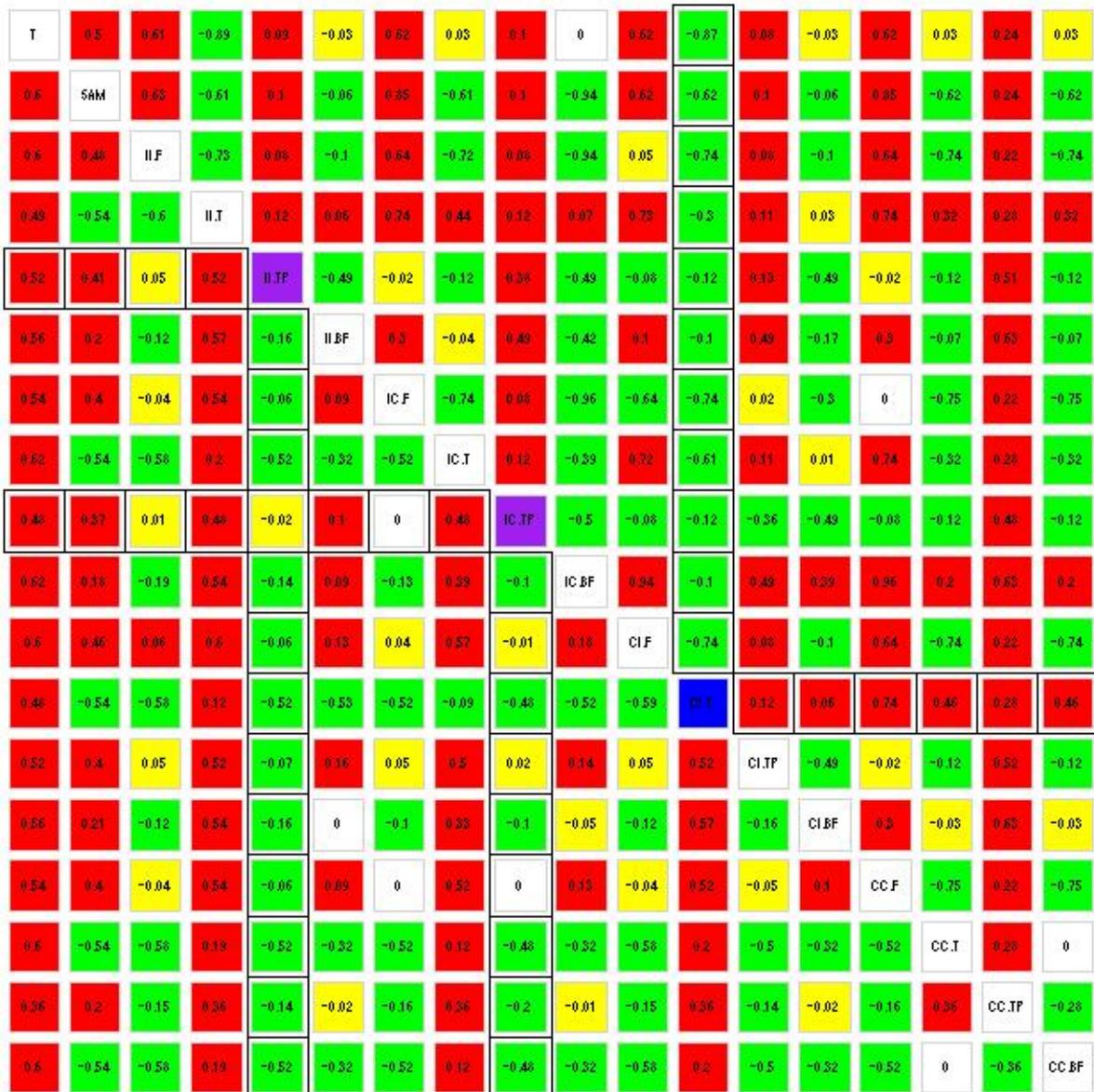


Figure 61: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

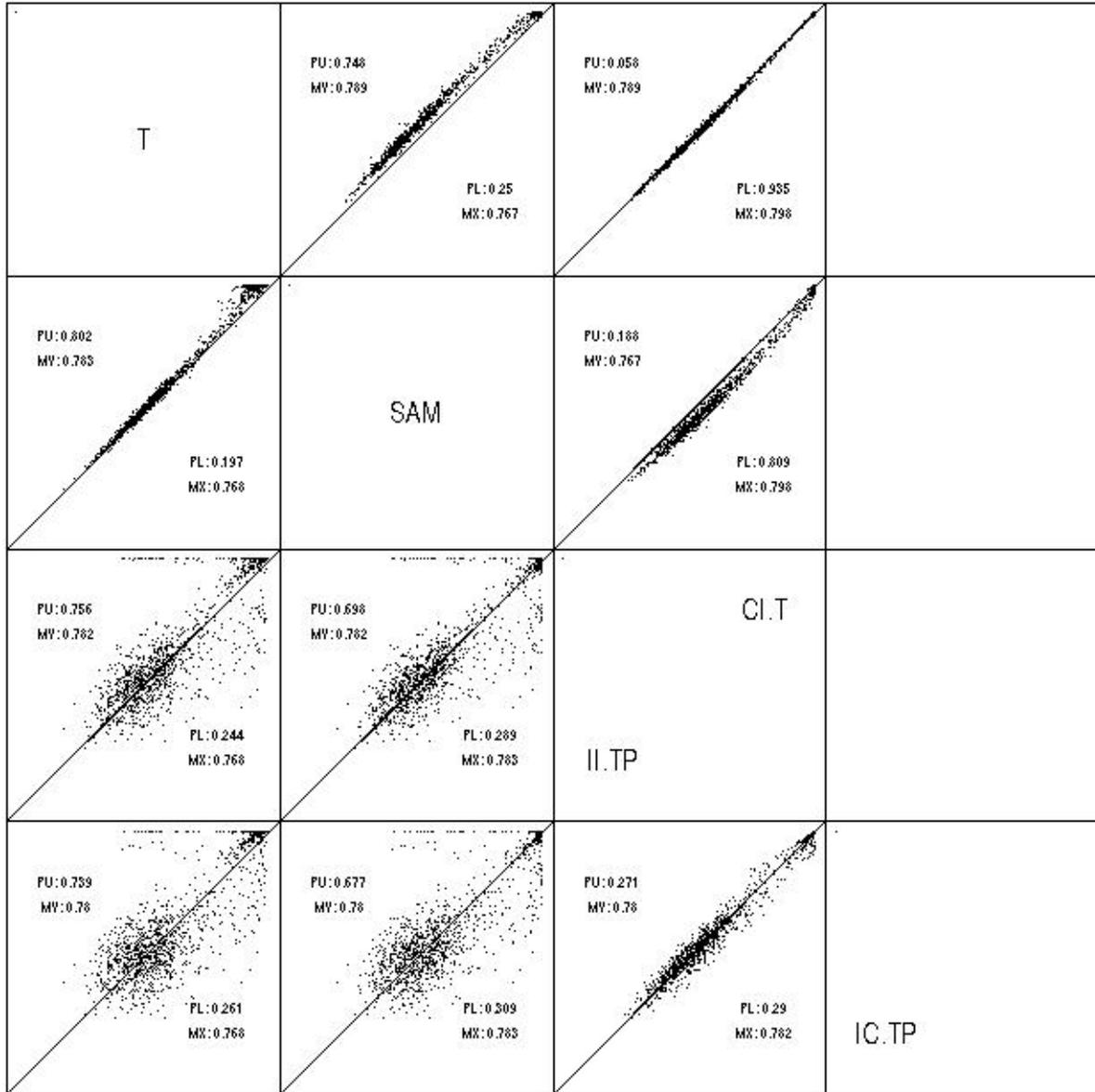


Figure 62: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

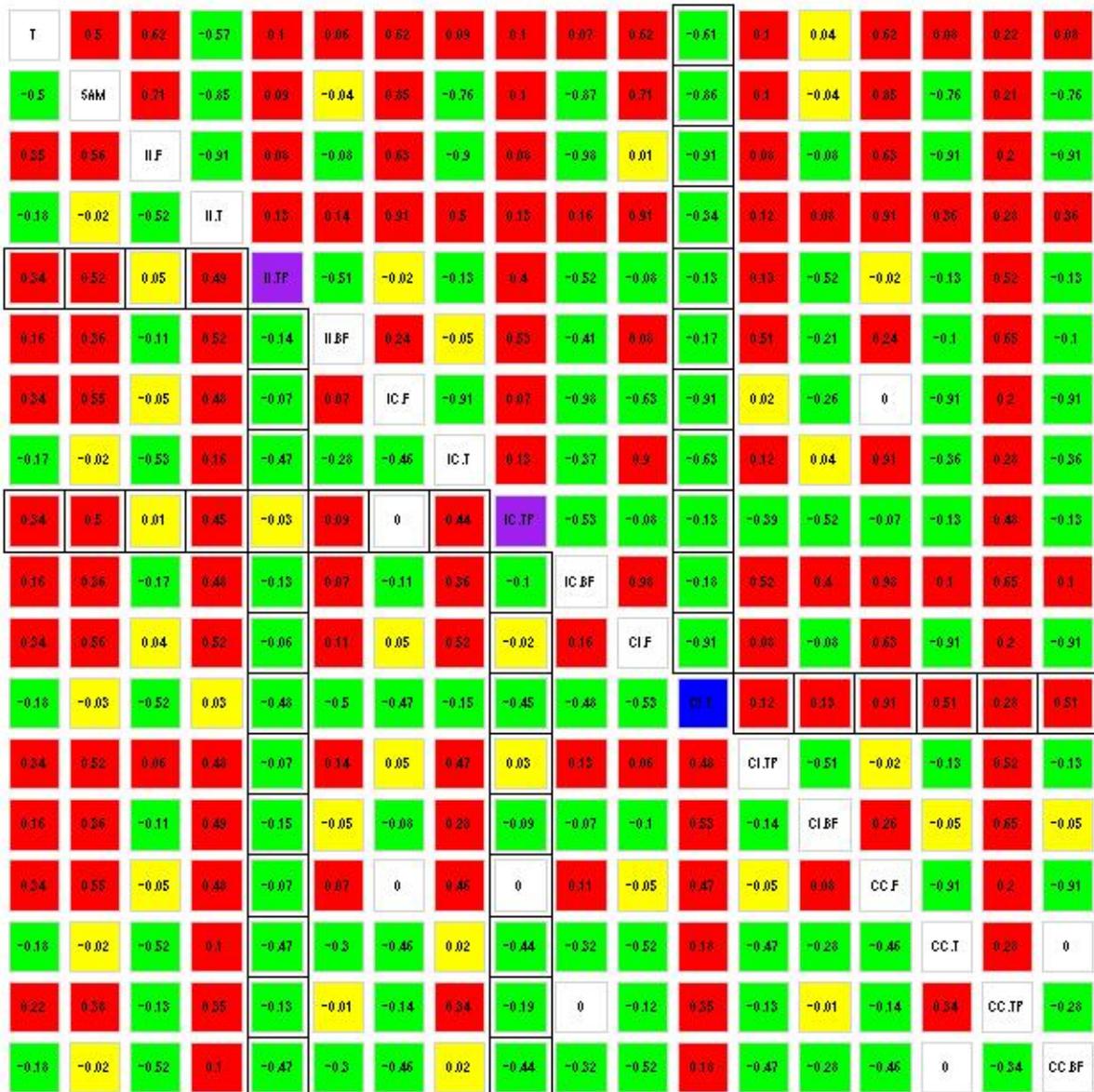


Figure 63: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

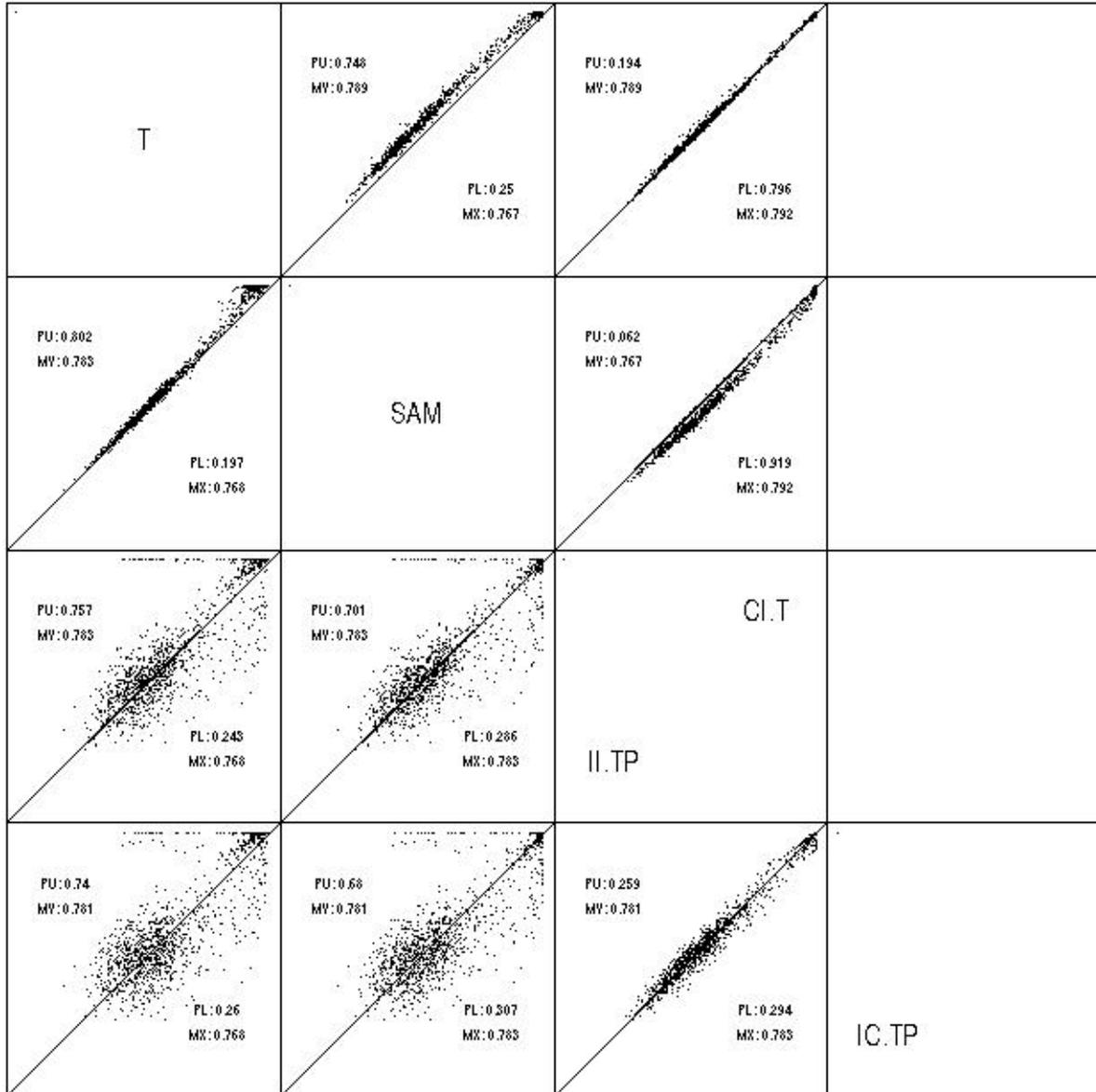


Figure 64: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

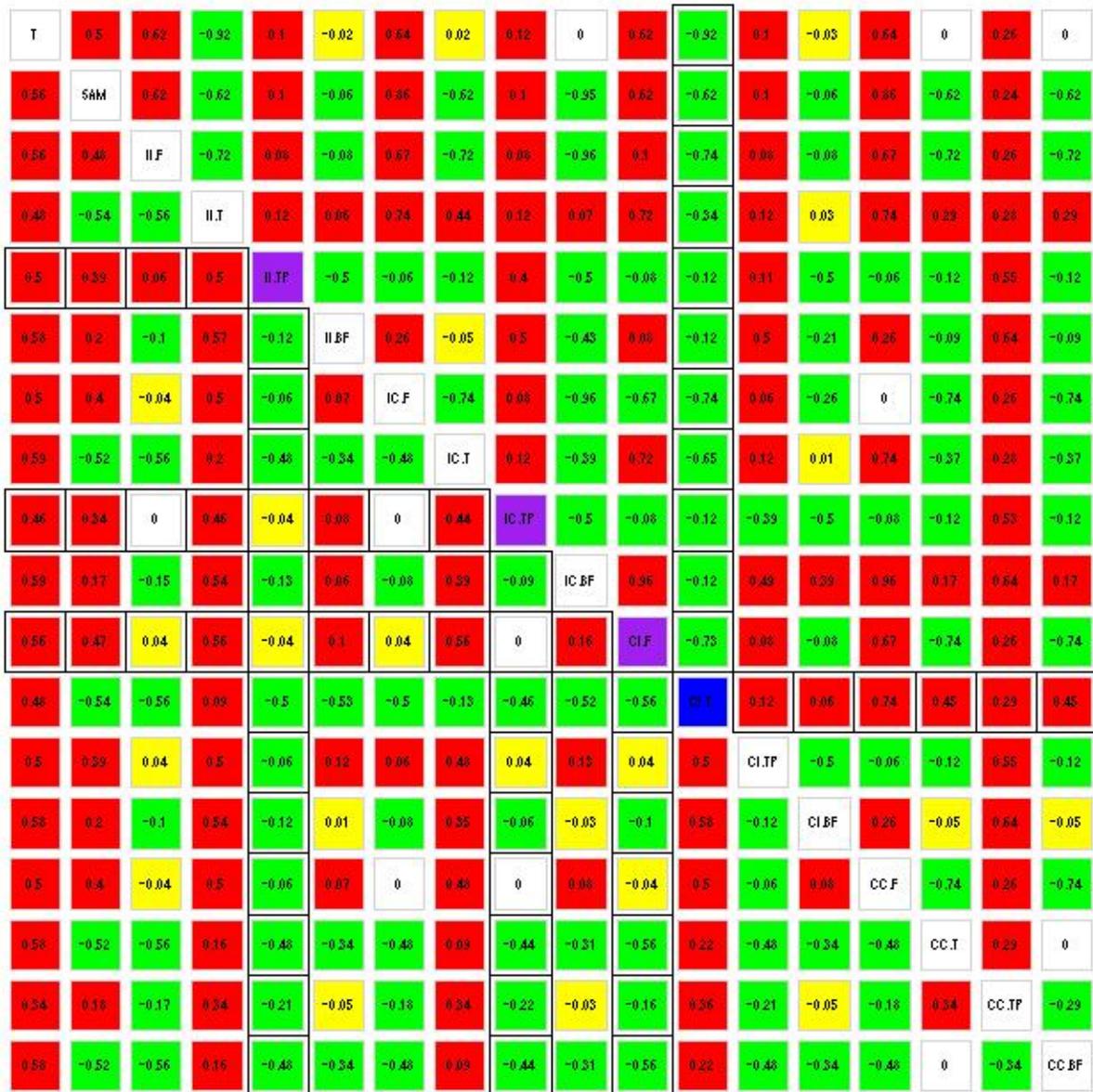


Figure 65: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

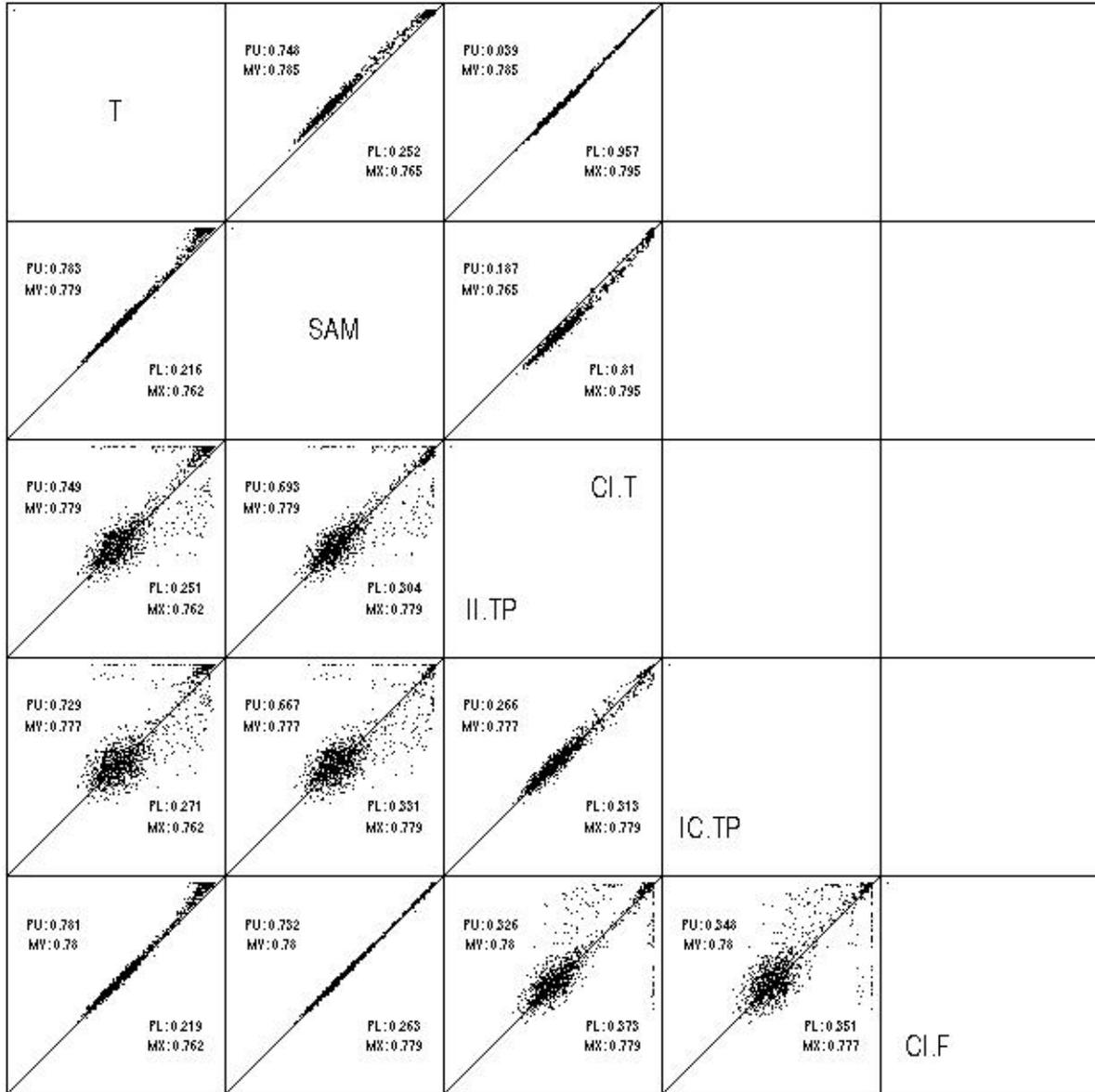


Figure 66: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

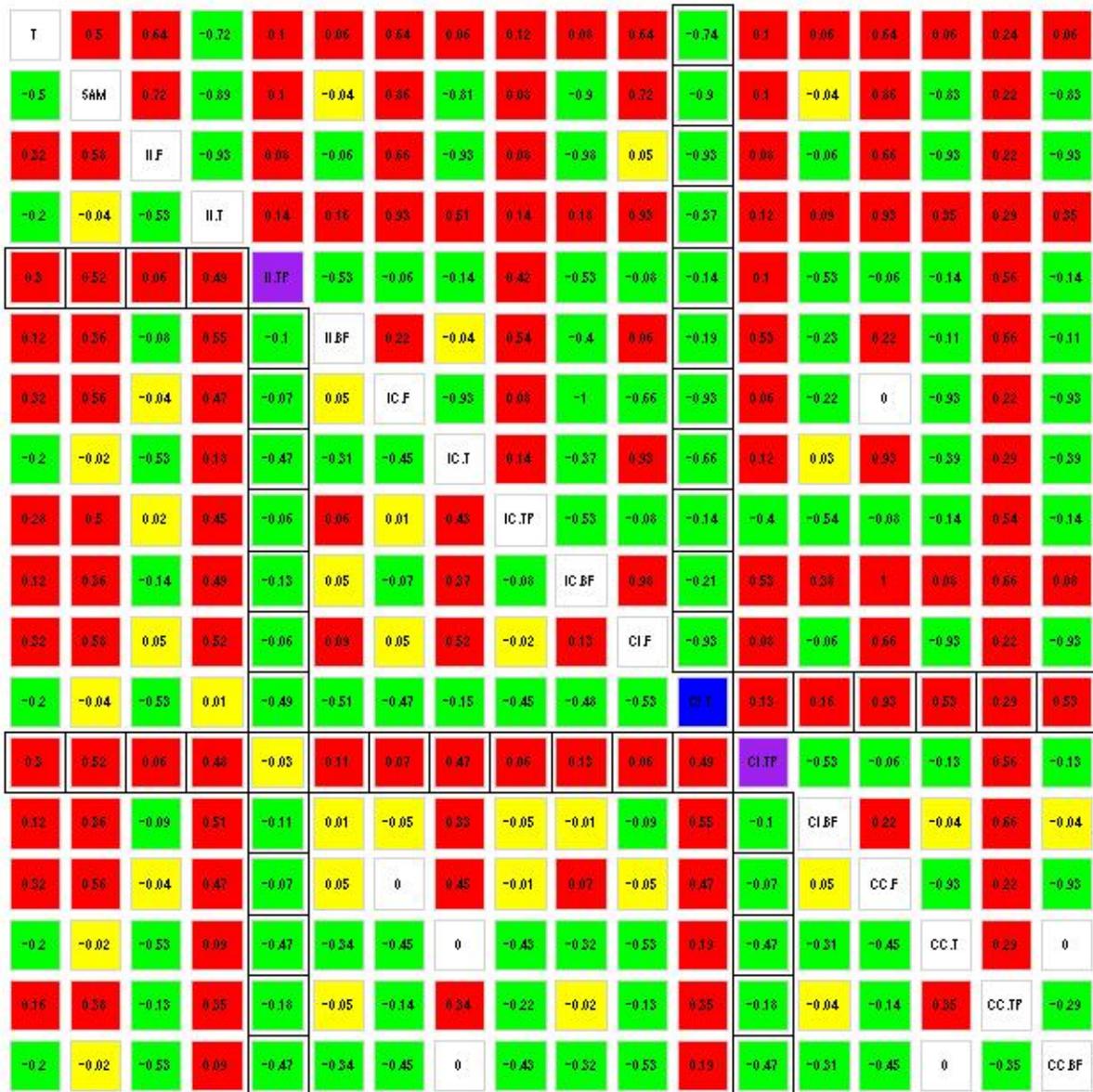


Figure 67: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

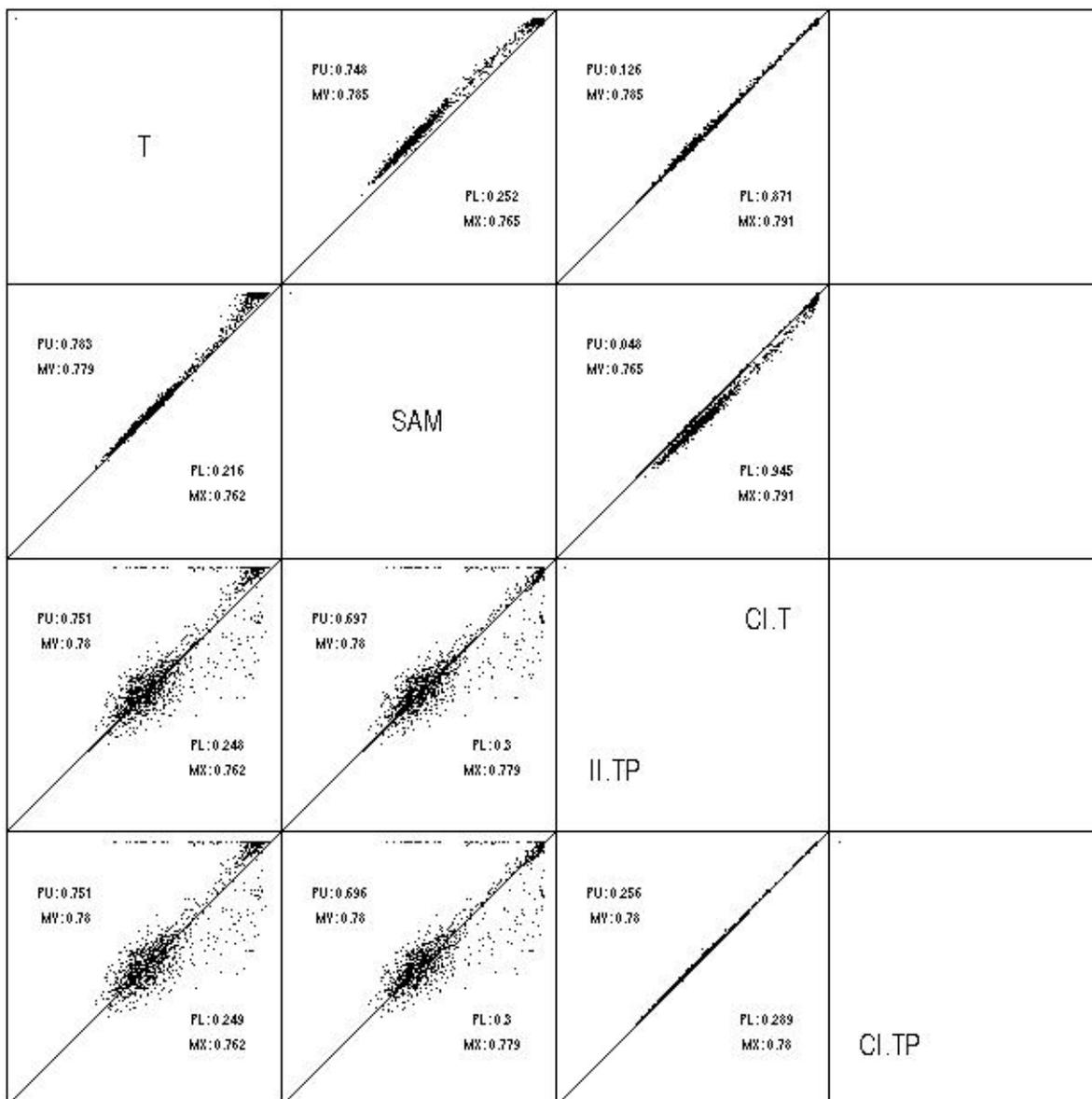


Figure 68: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

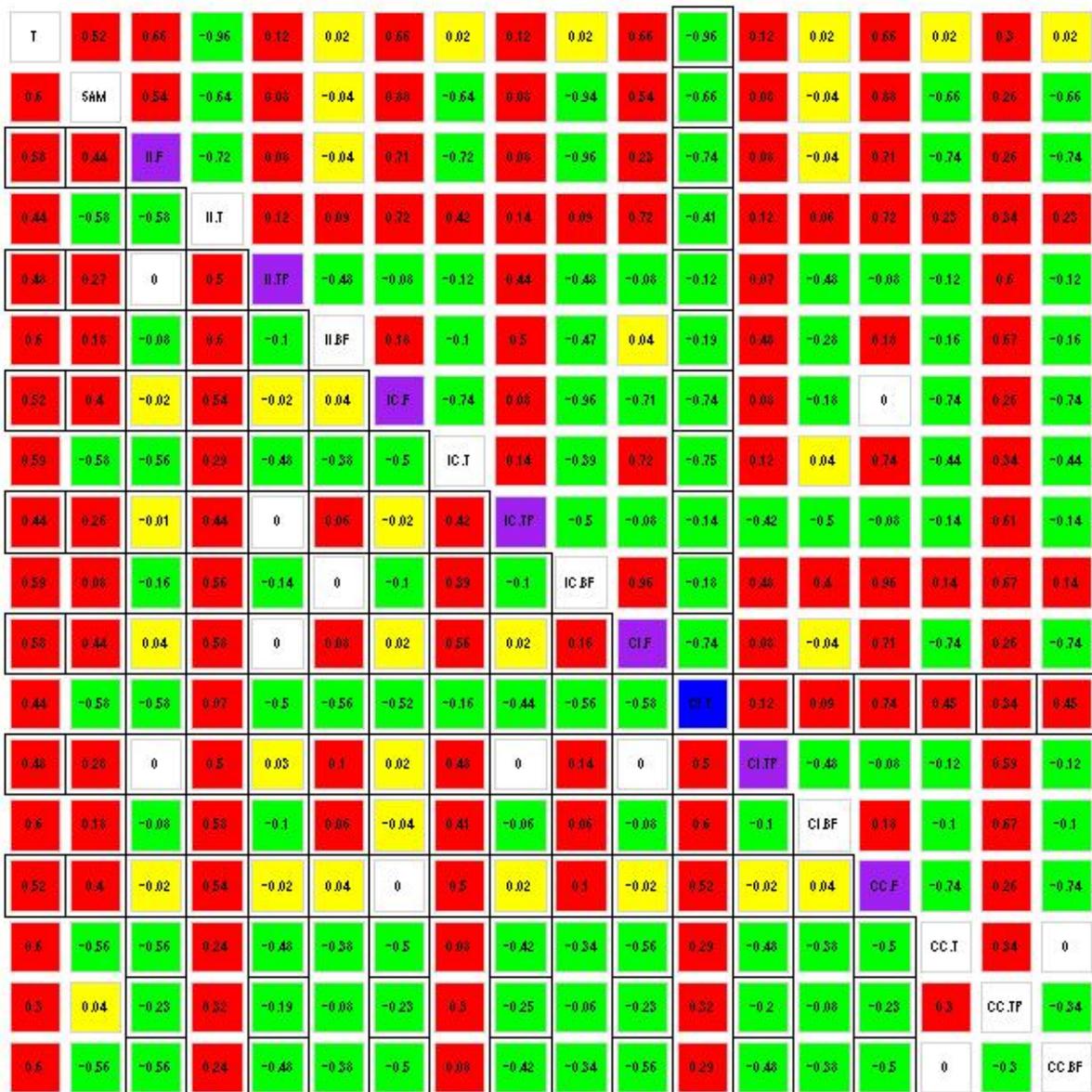


Figure 69: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

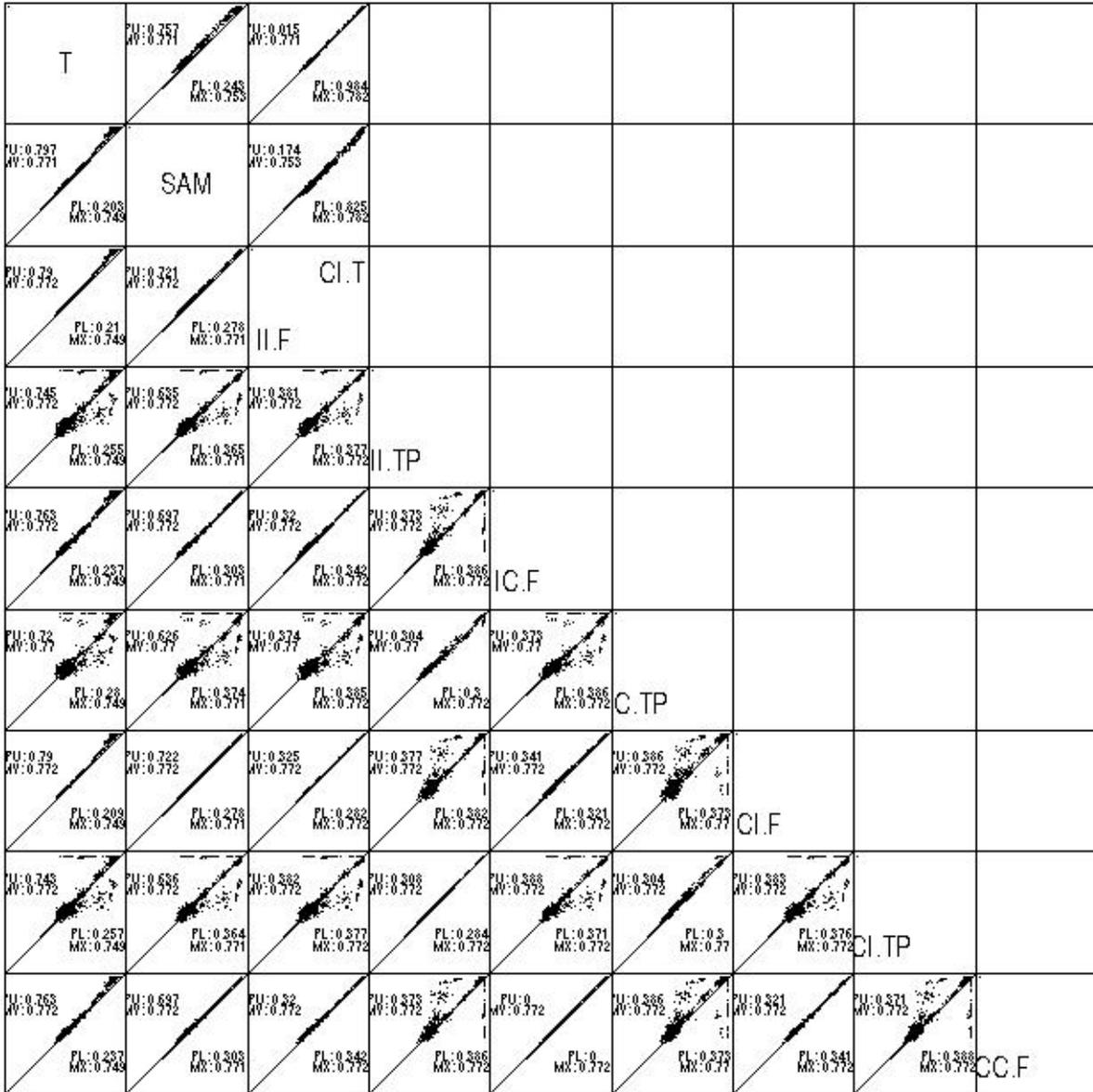


Figure 70: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

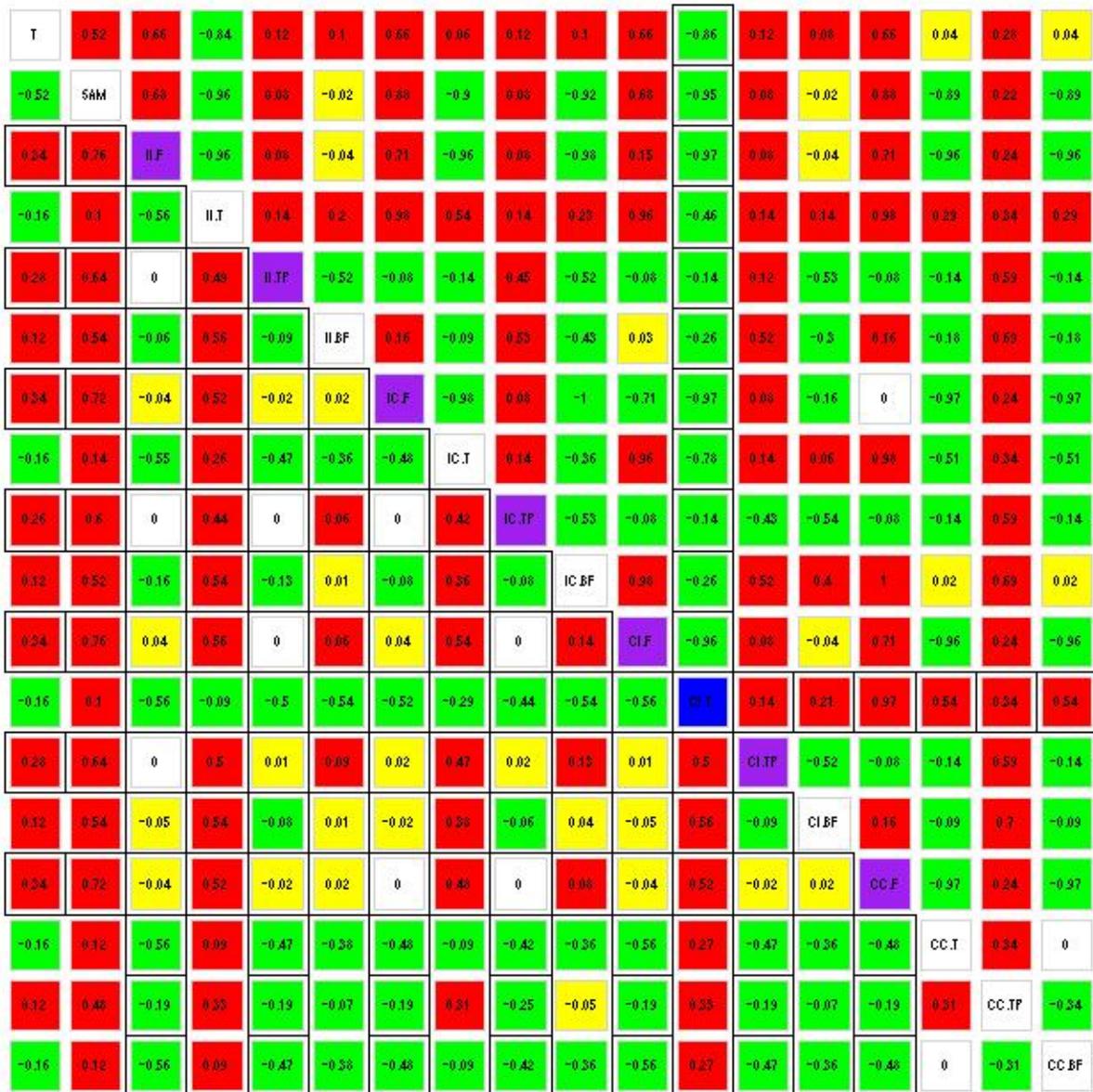


Figure 71: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

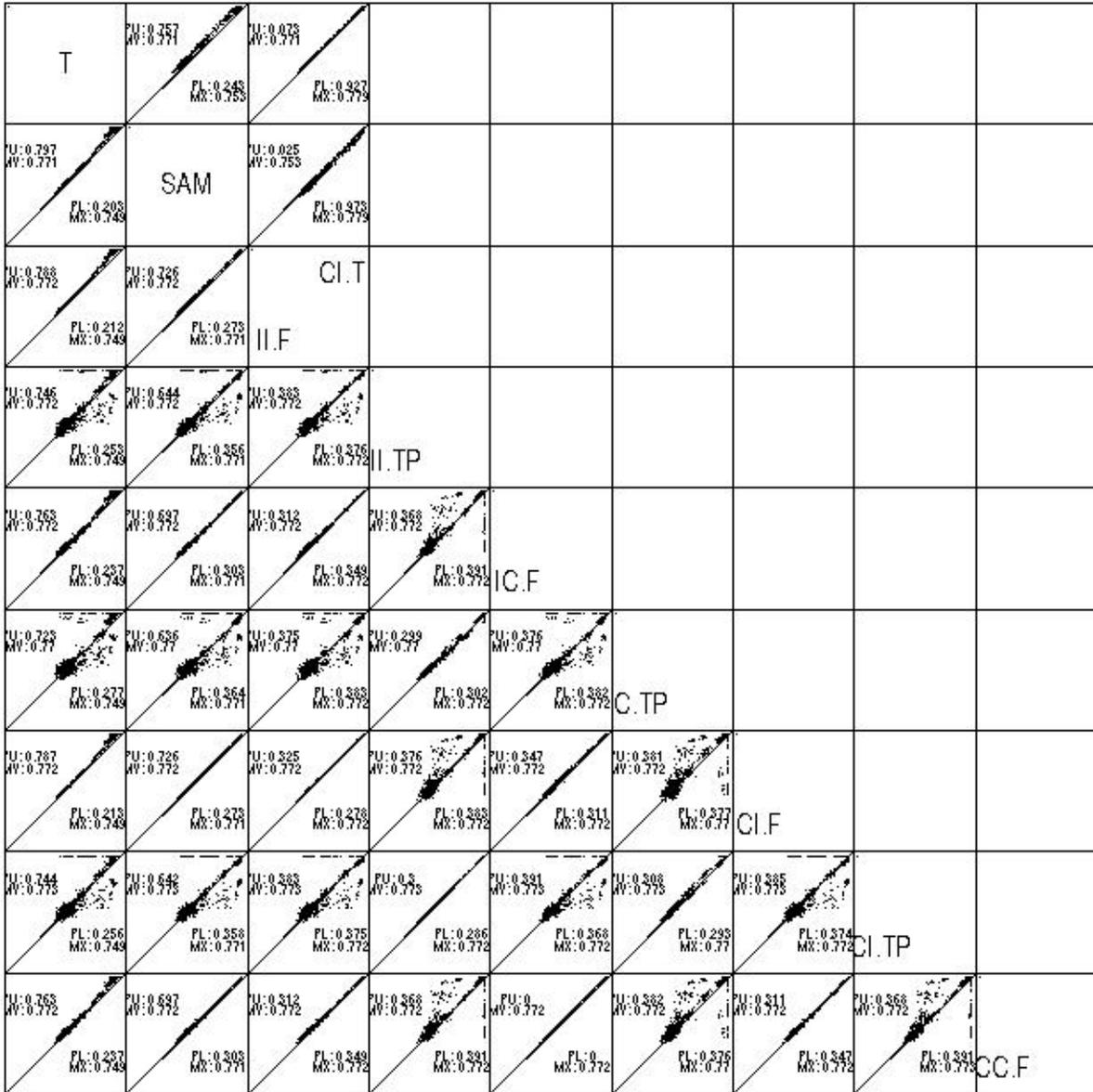


Figure 72: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

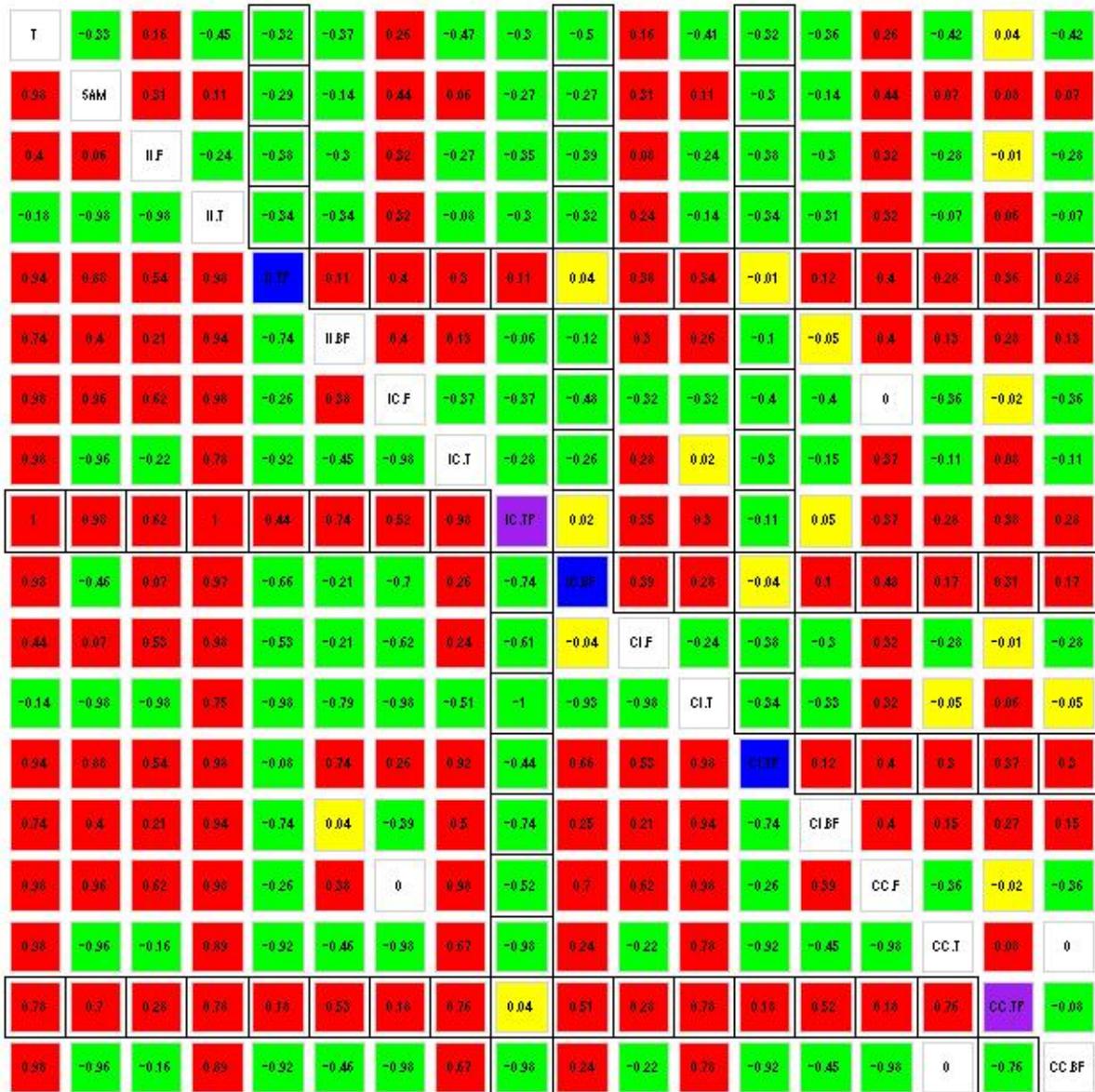


Figure 73: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

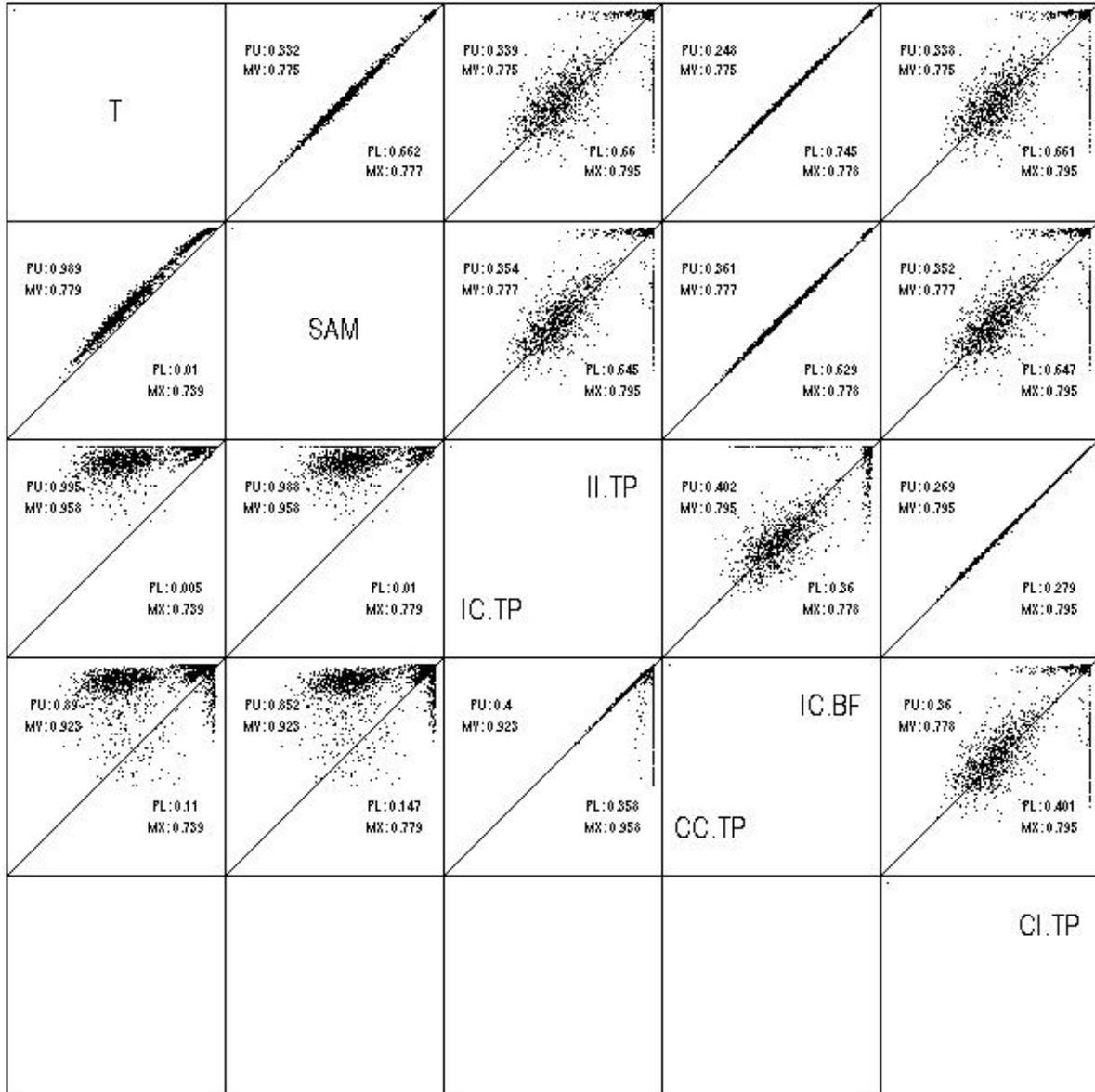


Figure 74: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

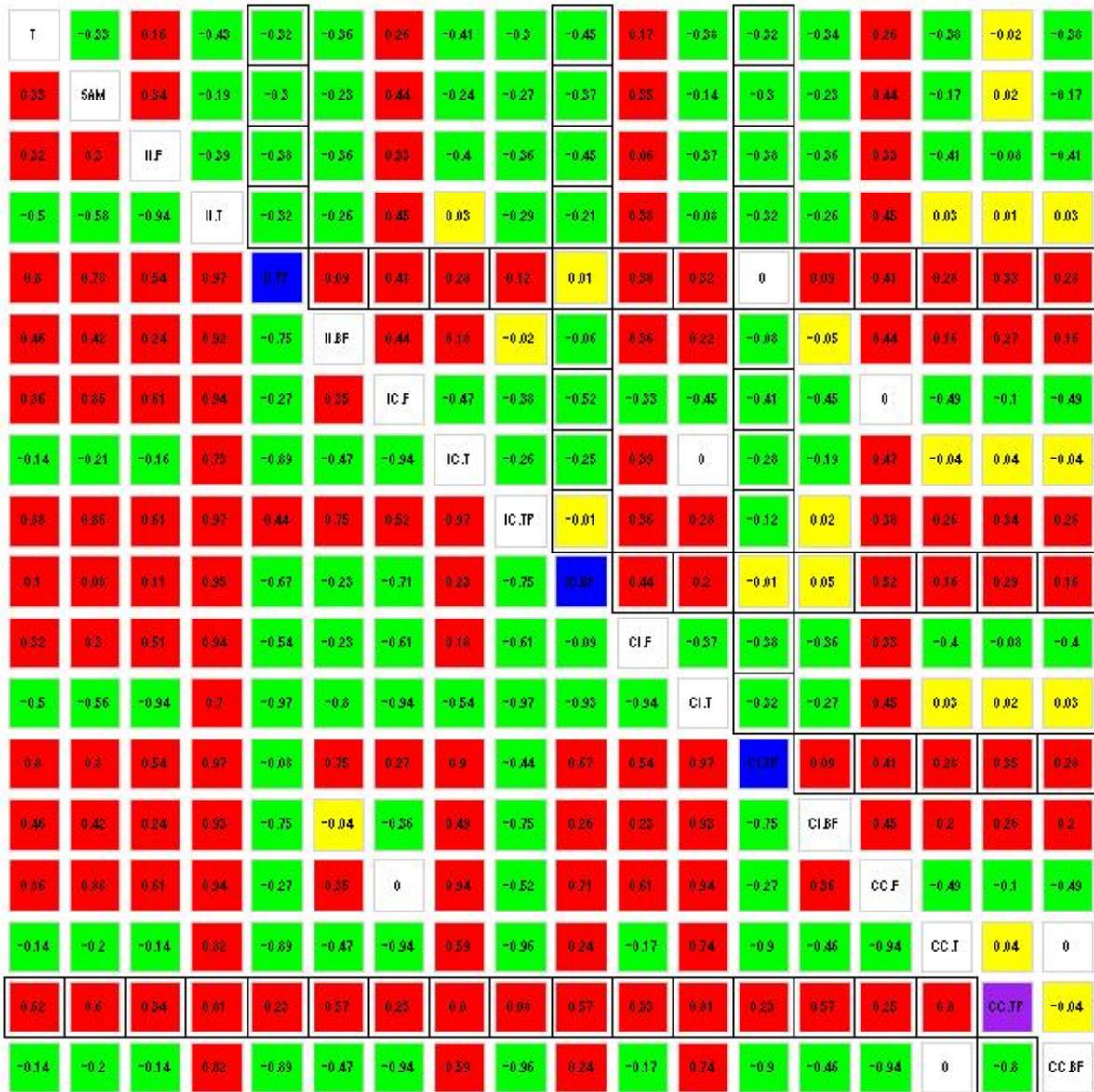


Figure 75: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

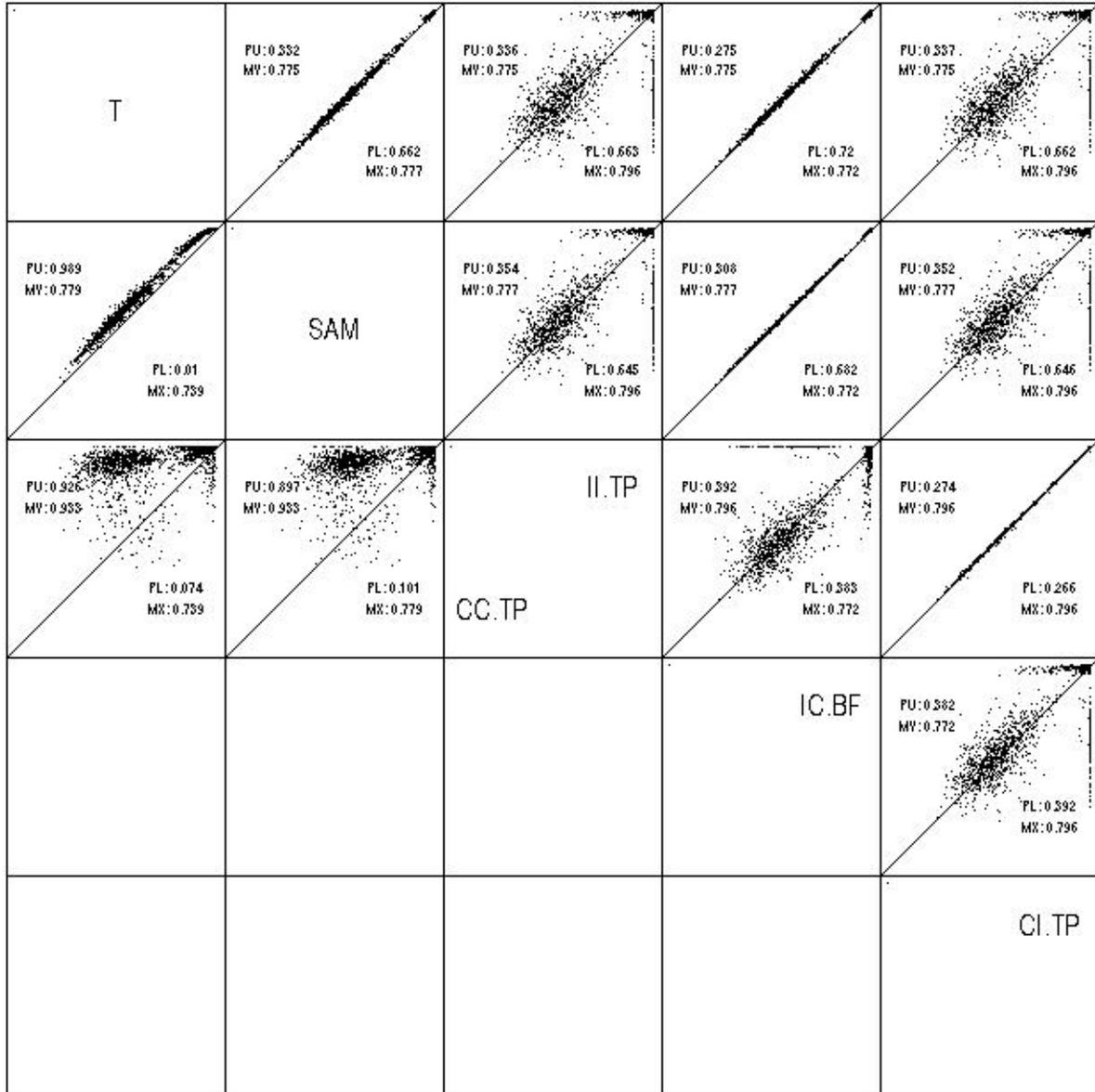


Figure 76: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.



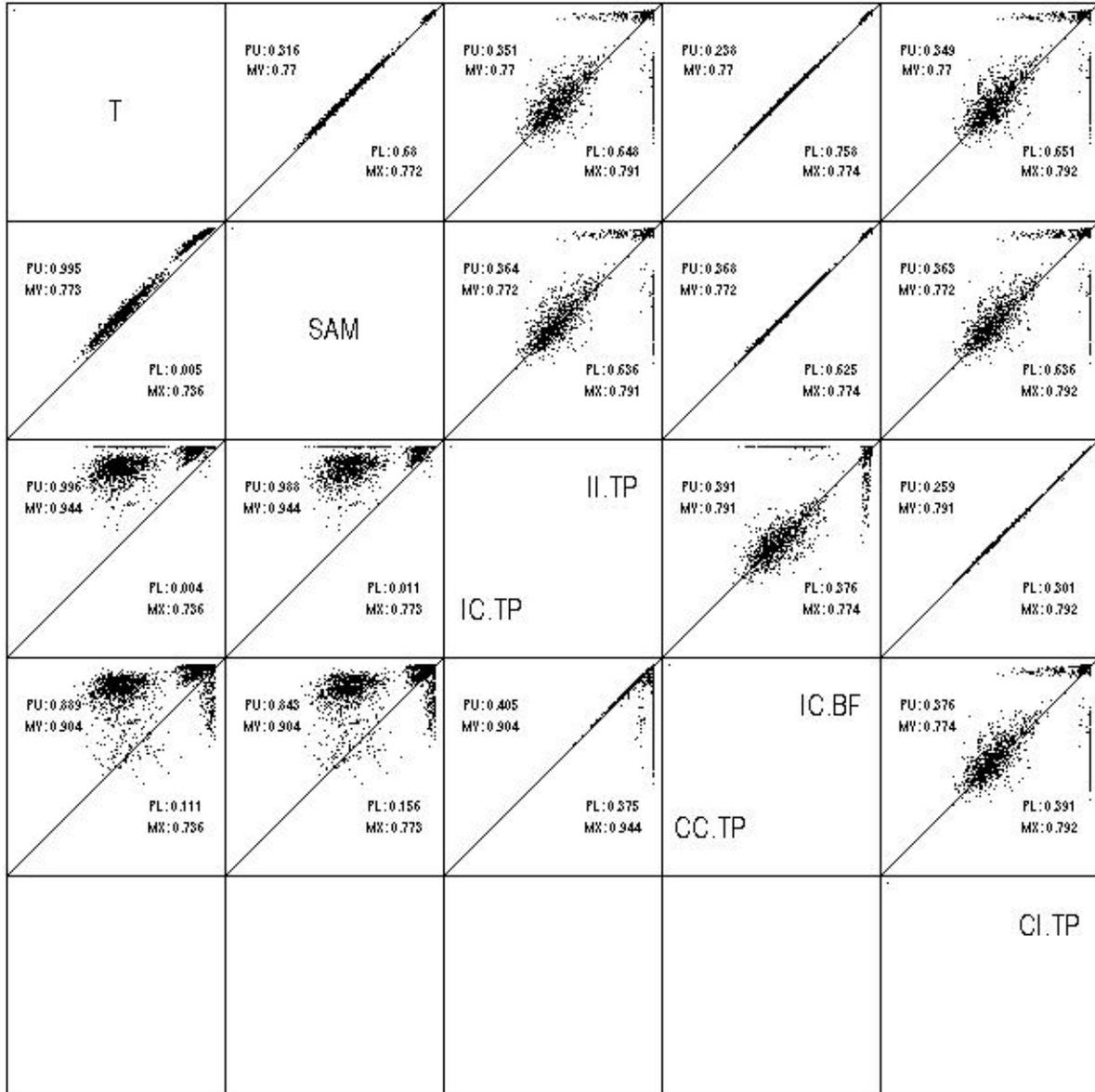


Figure 78: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

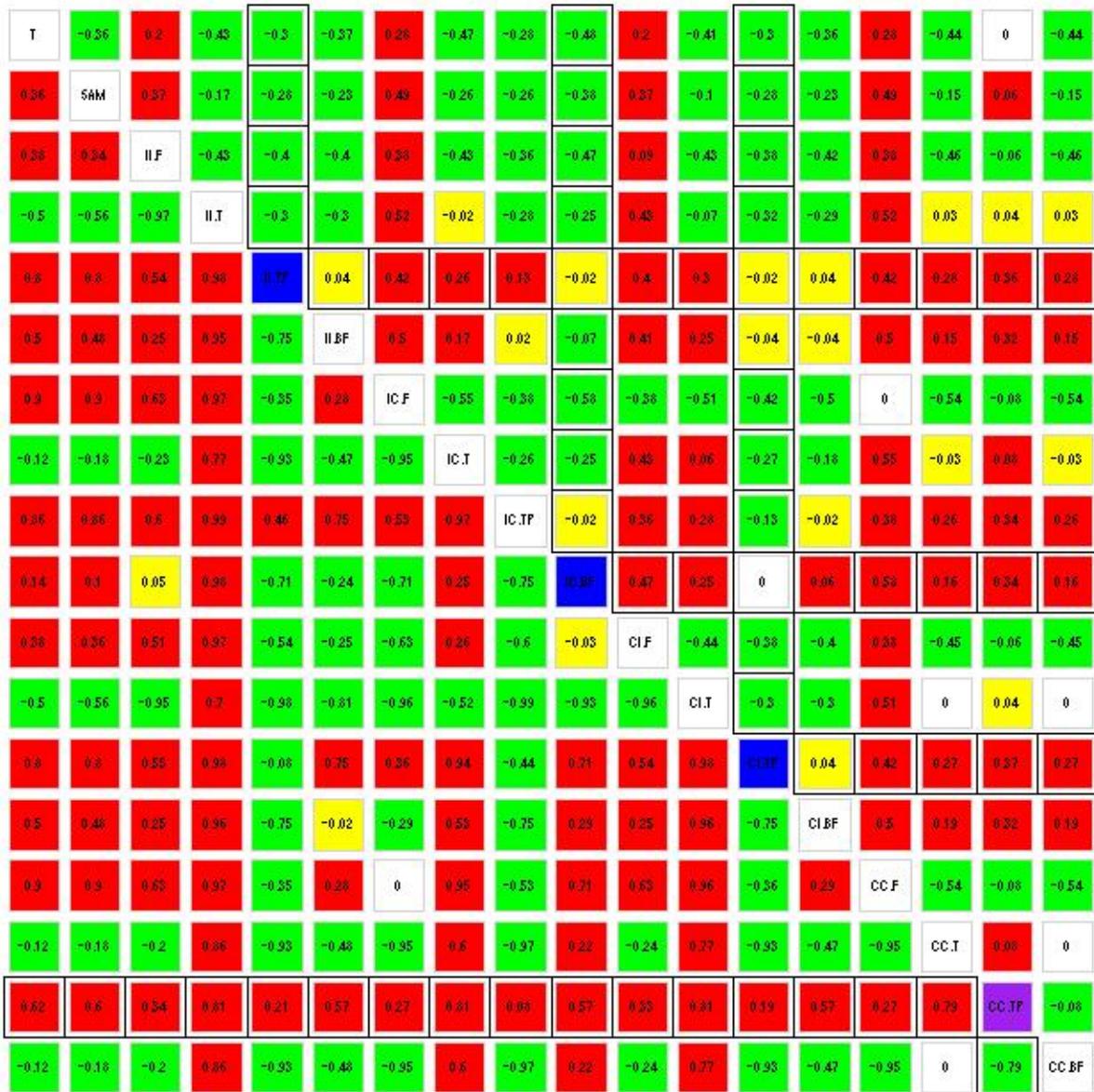


Figure 79: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

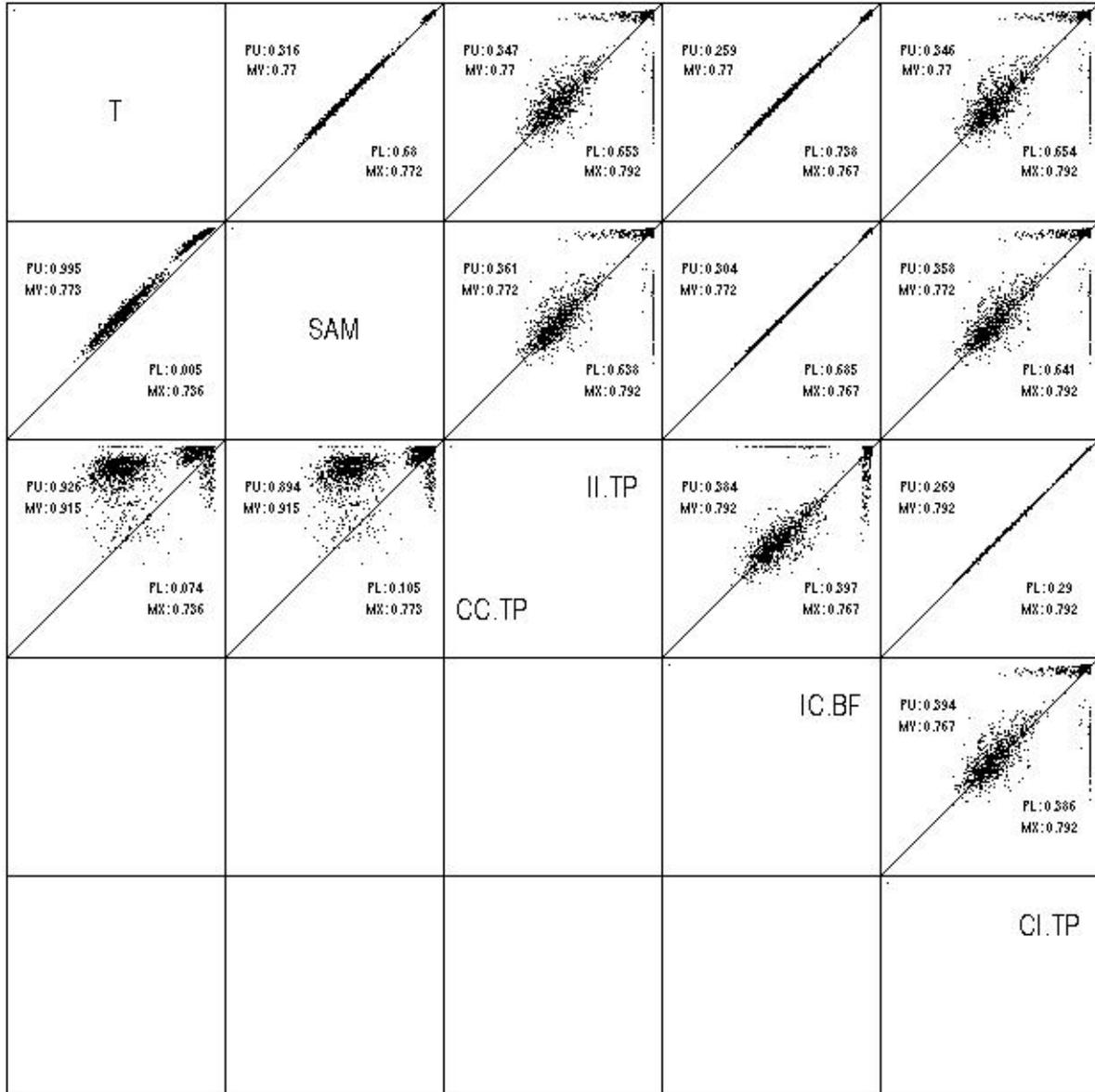


Figure 80: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

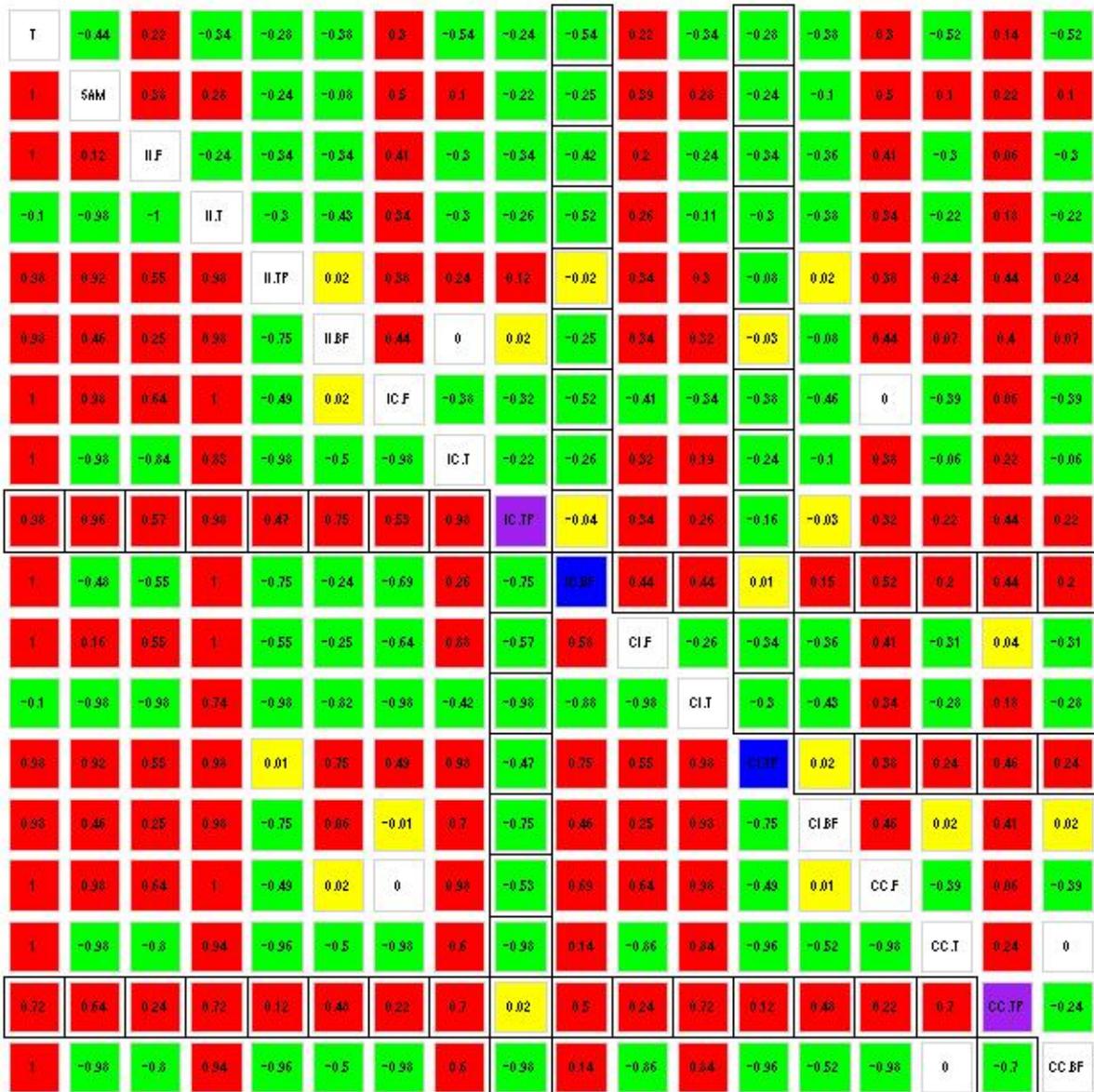


Figure 81: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

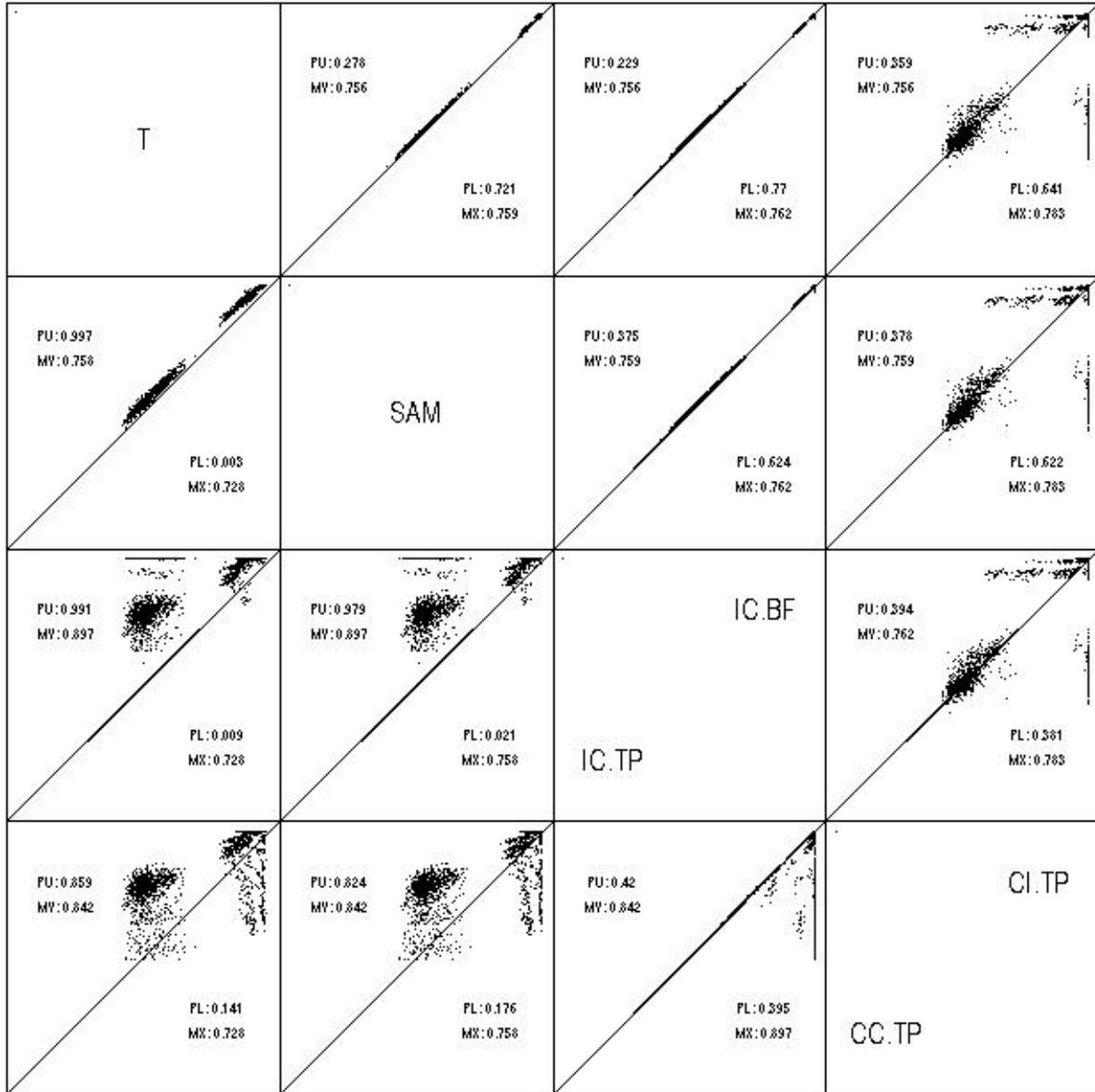


Figure 82: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

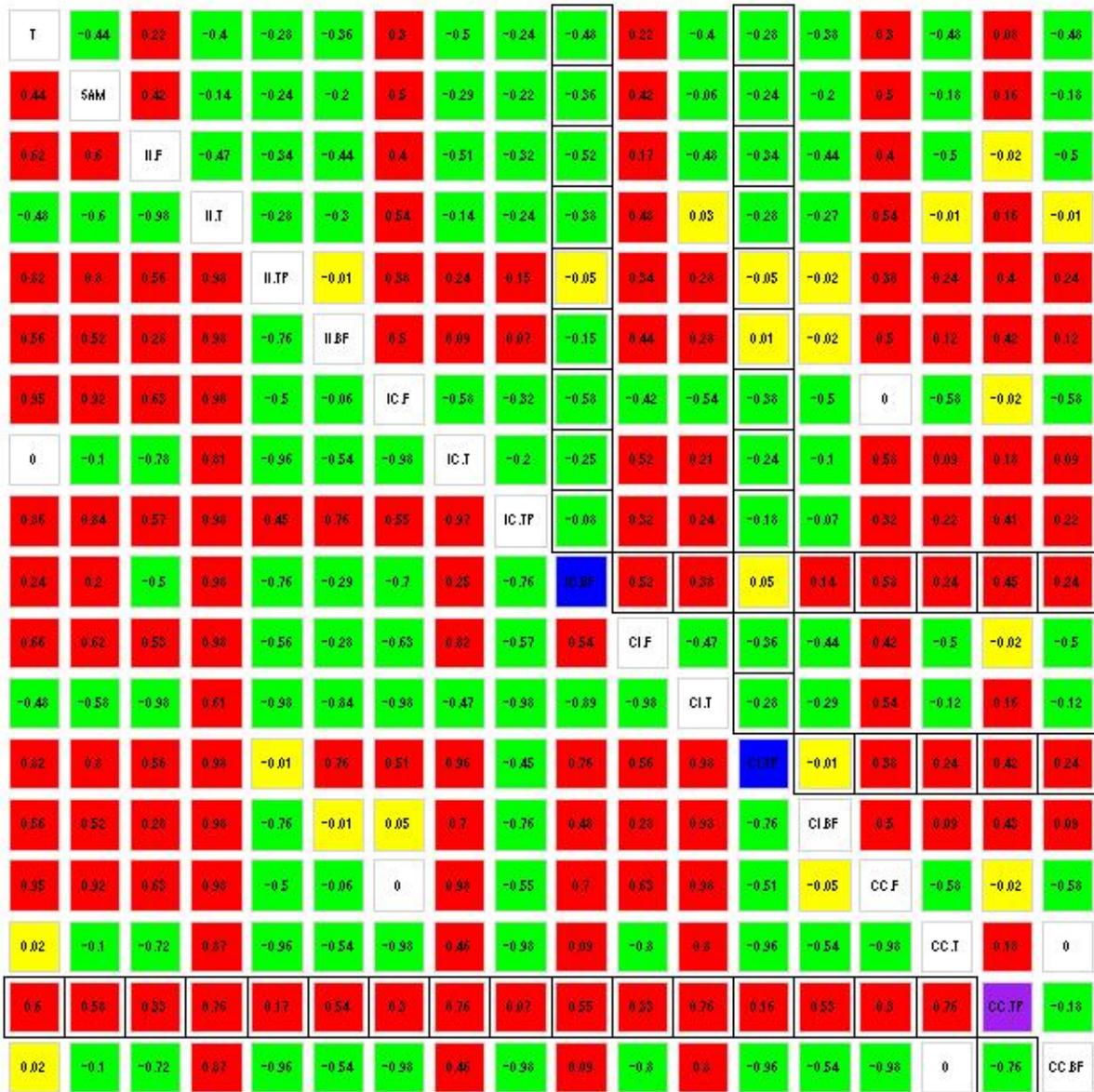


Figure 83: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

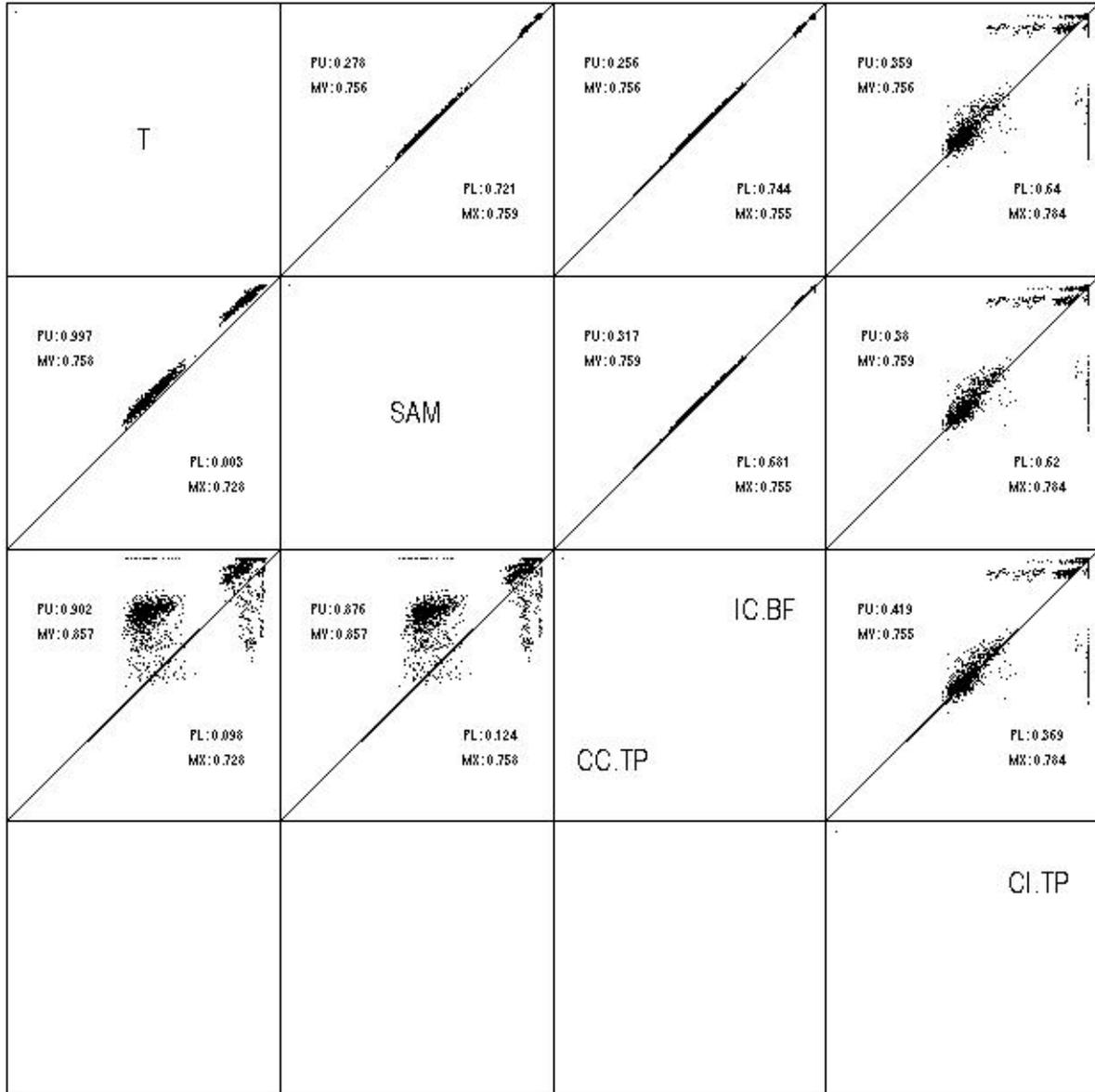


Figure 84: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

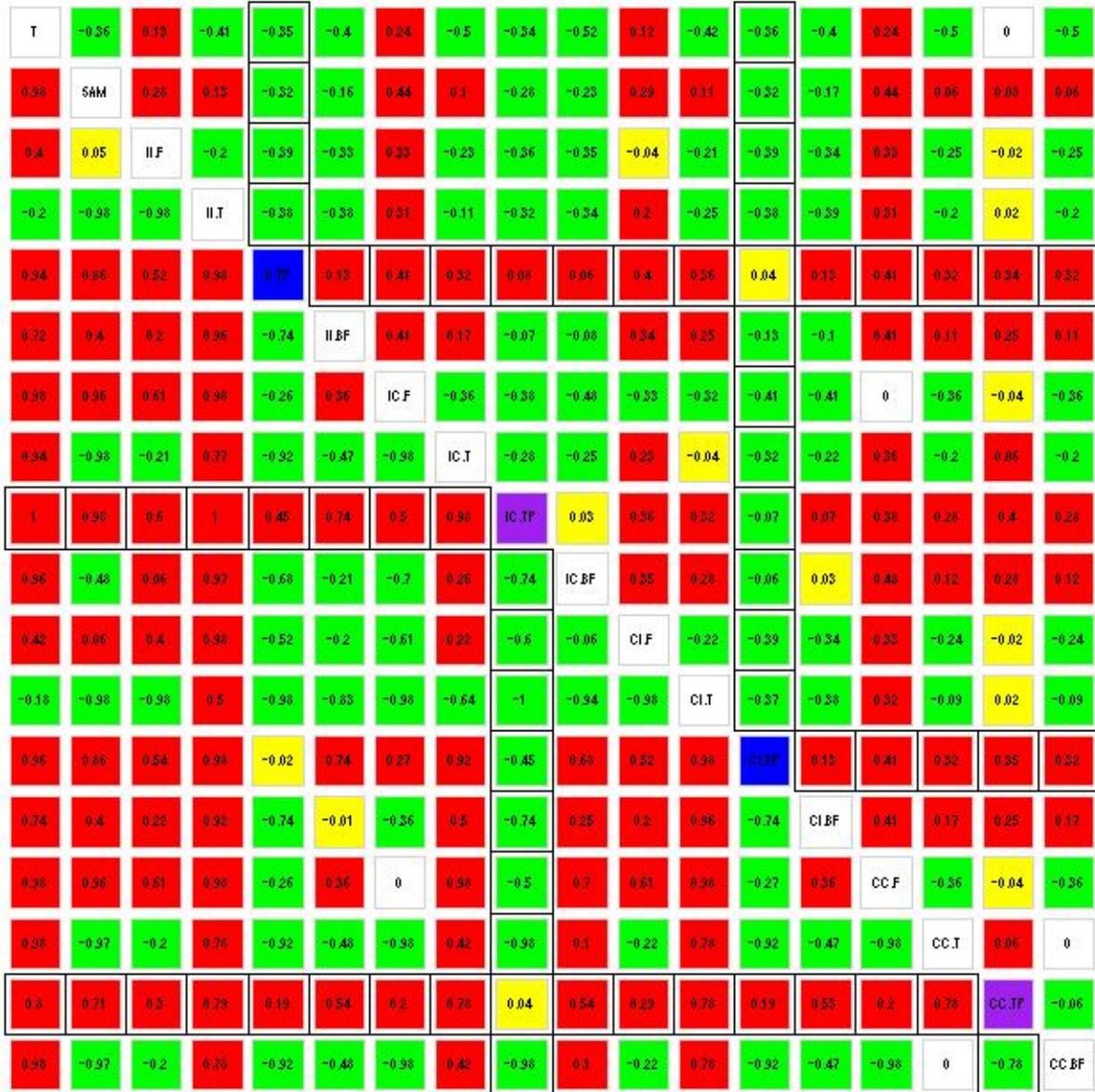


Figure 85: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

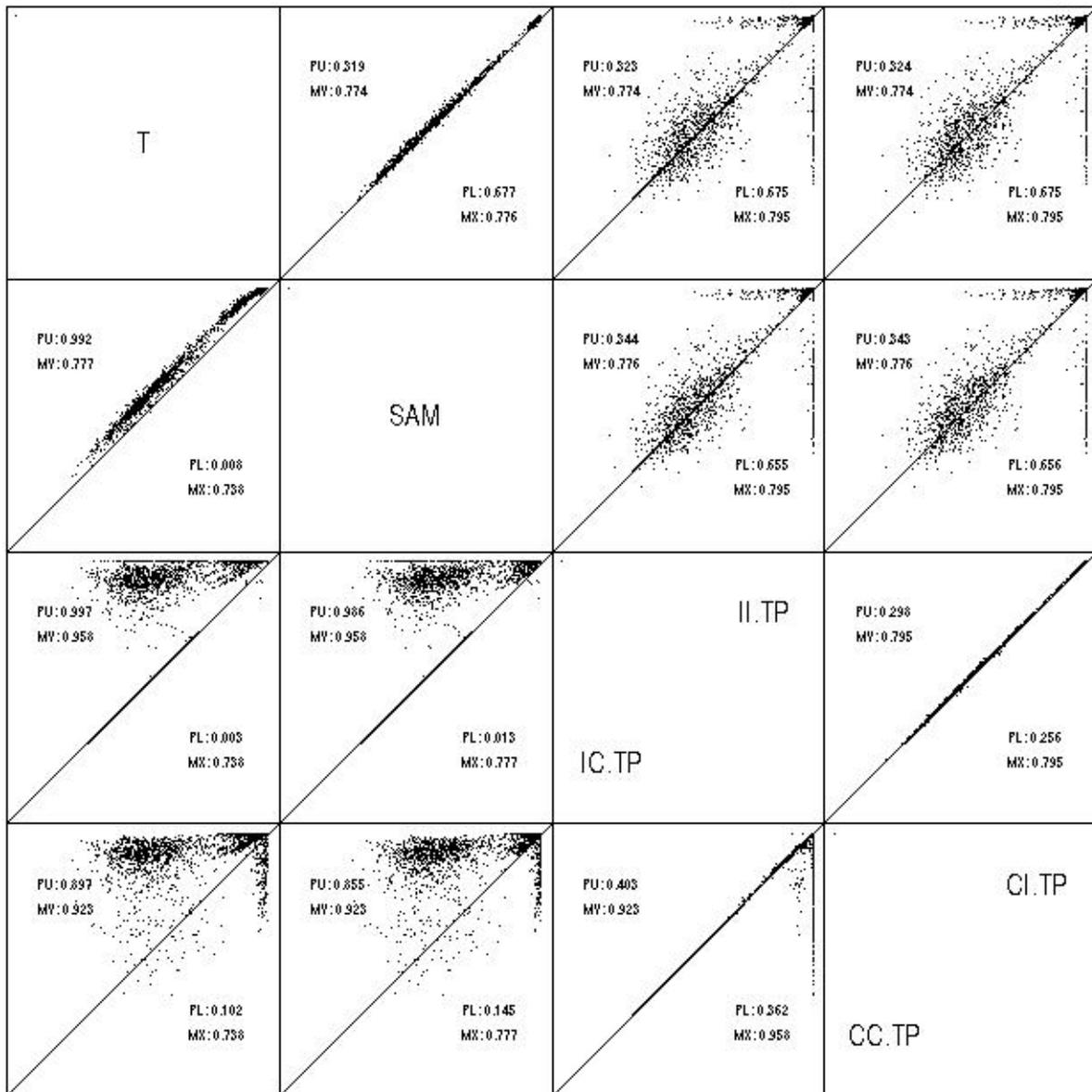


Figure 86: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

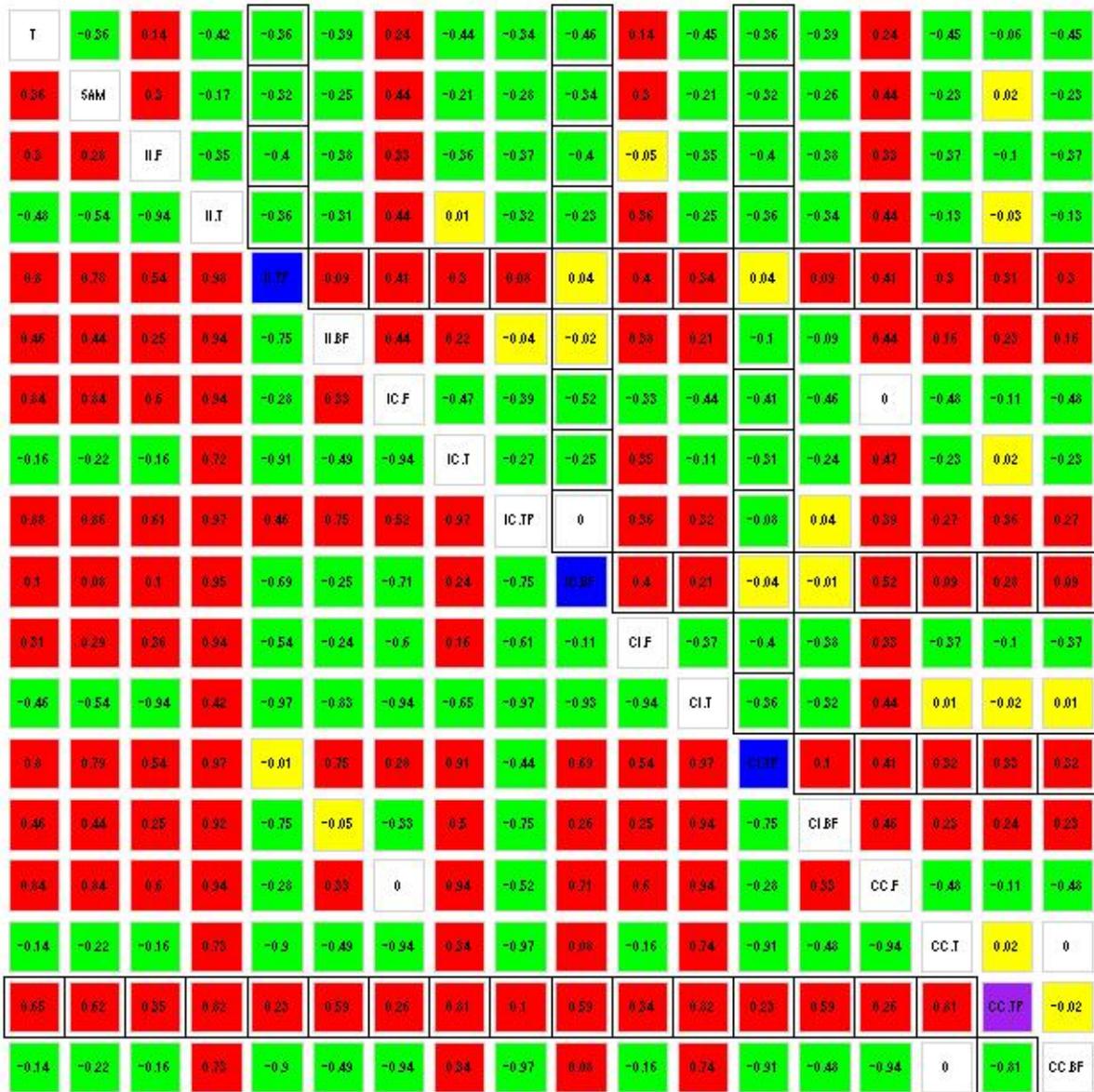


Figure 87: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

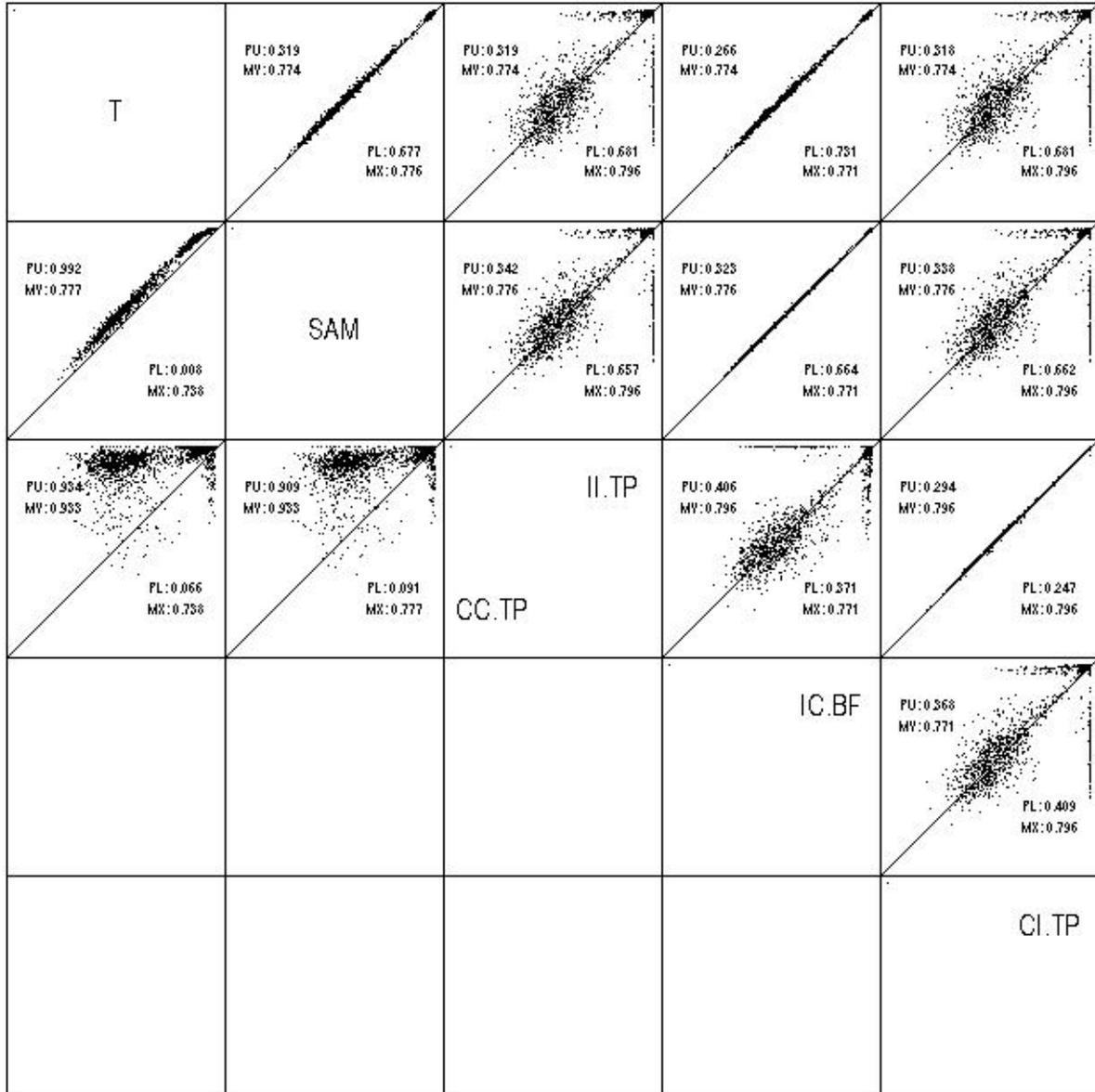


Figure 88: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

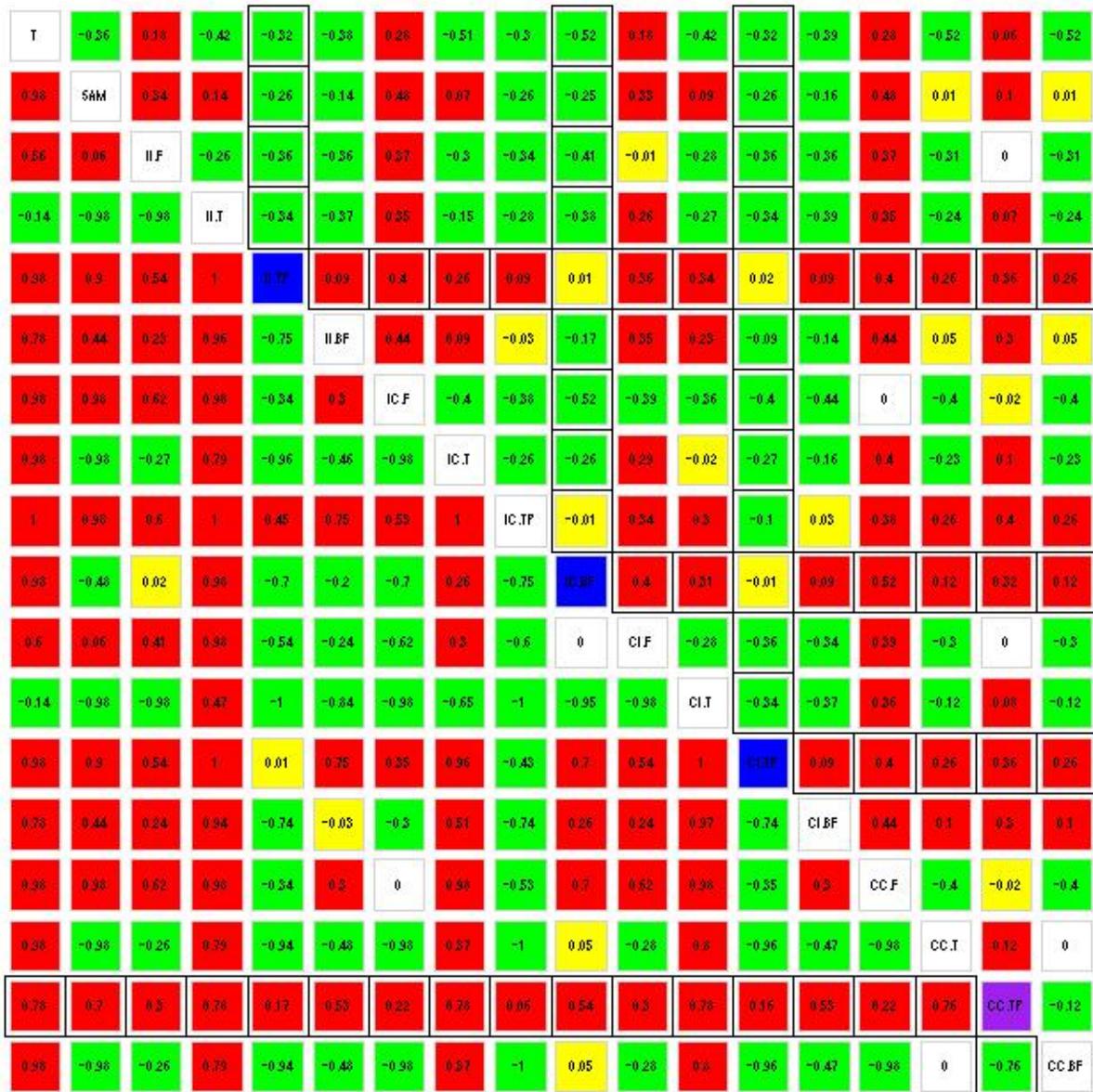


Figure 89: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 10 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

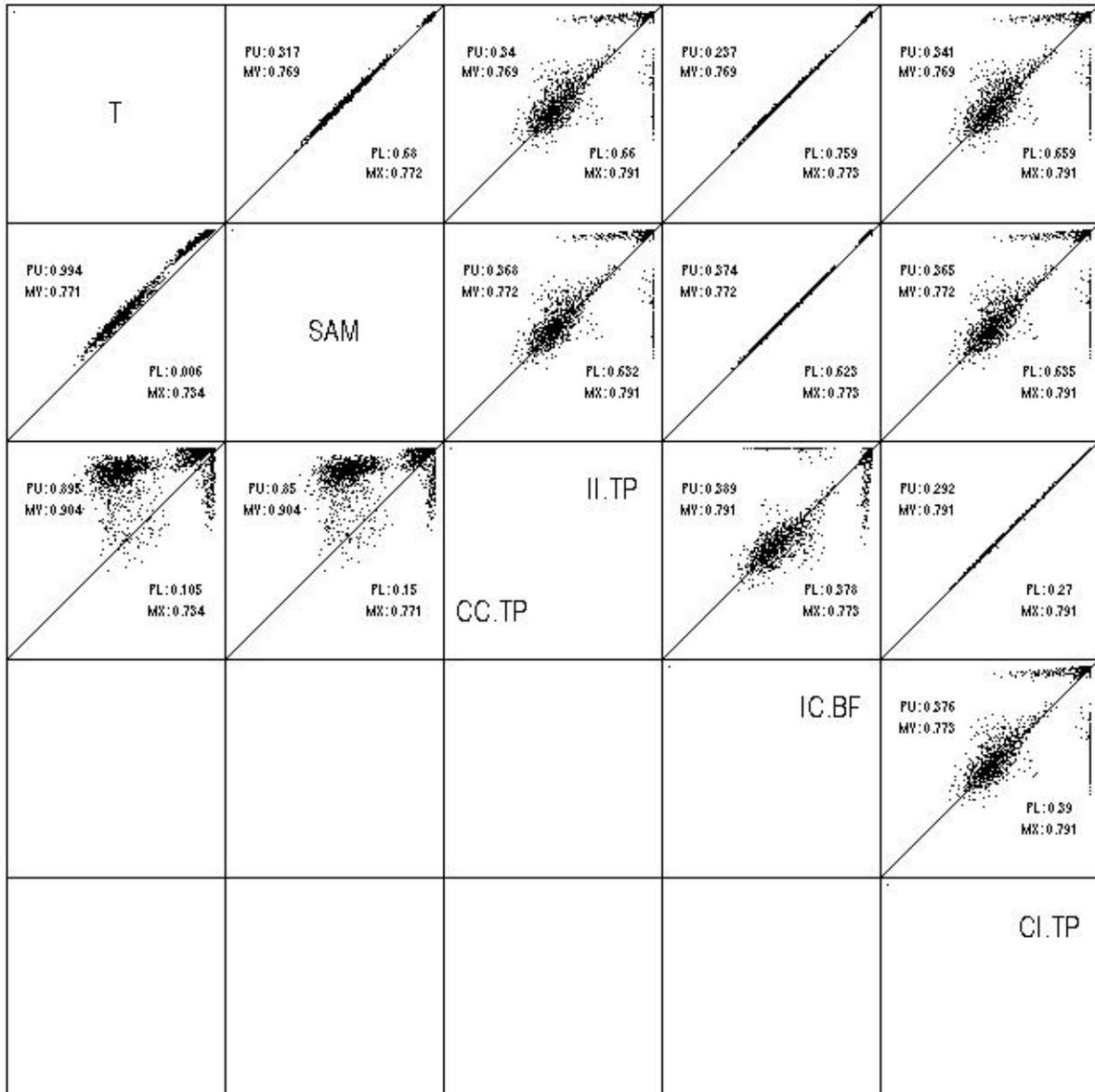


Figure 90: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

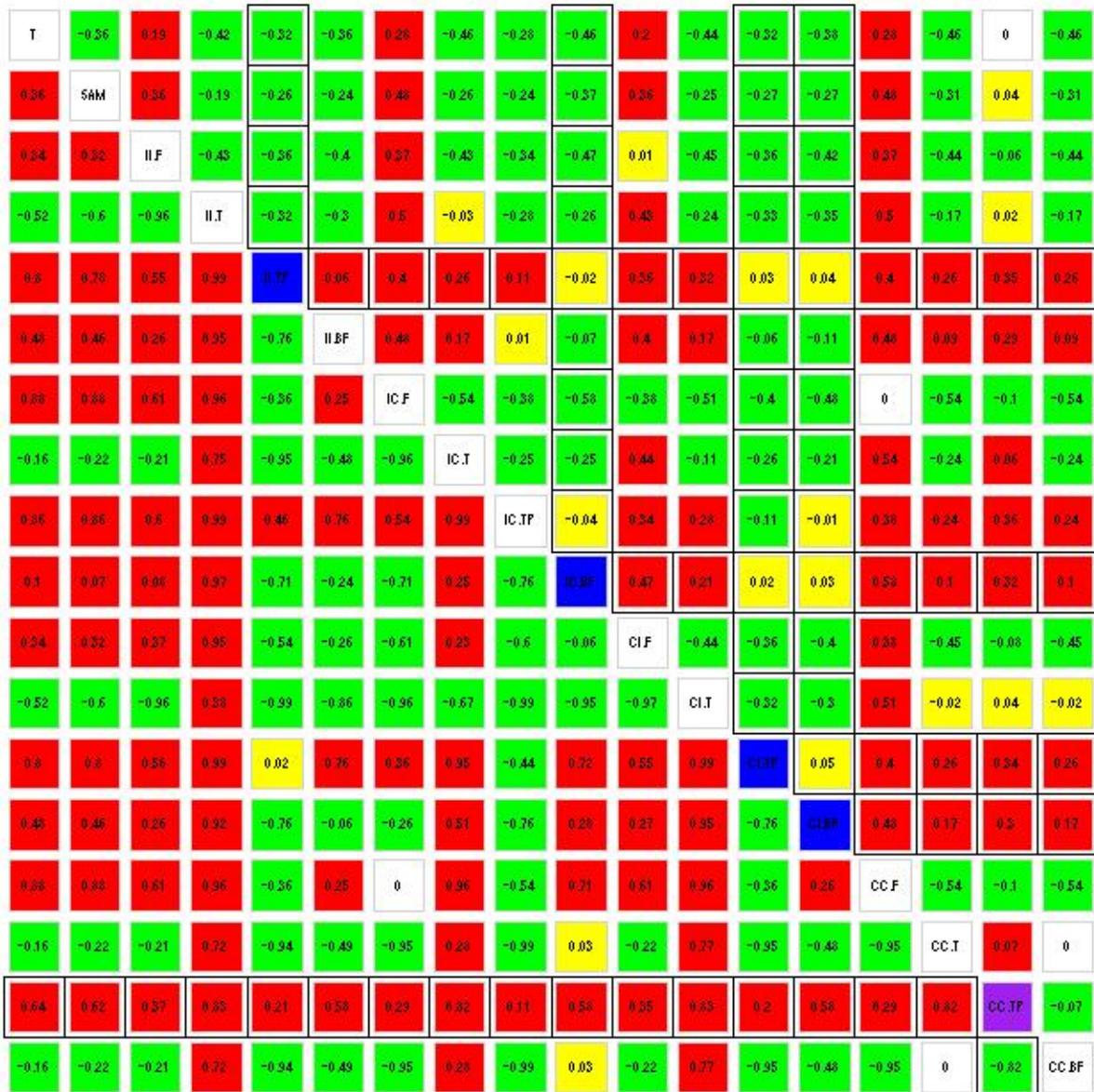


Figure 91: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

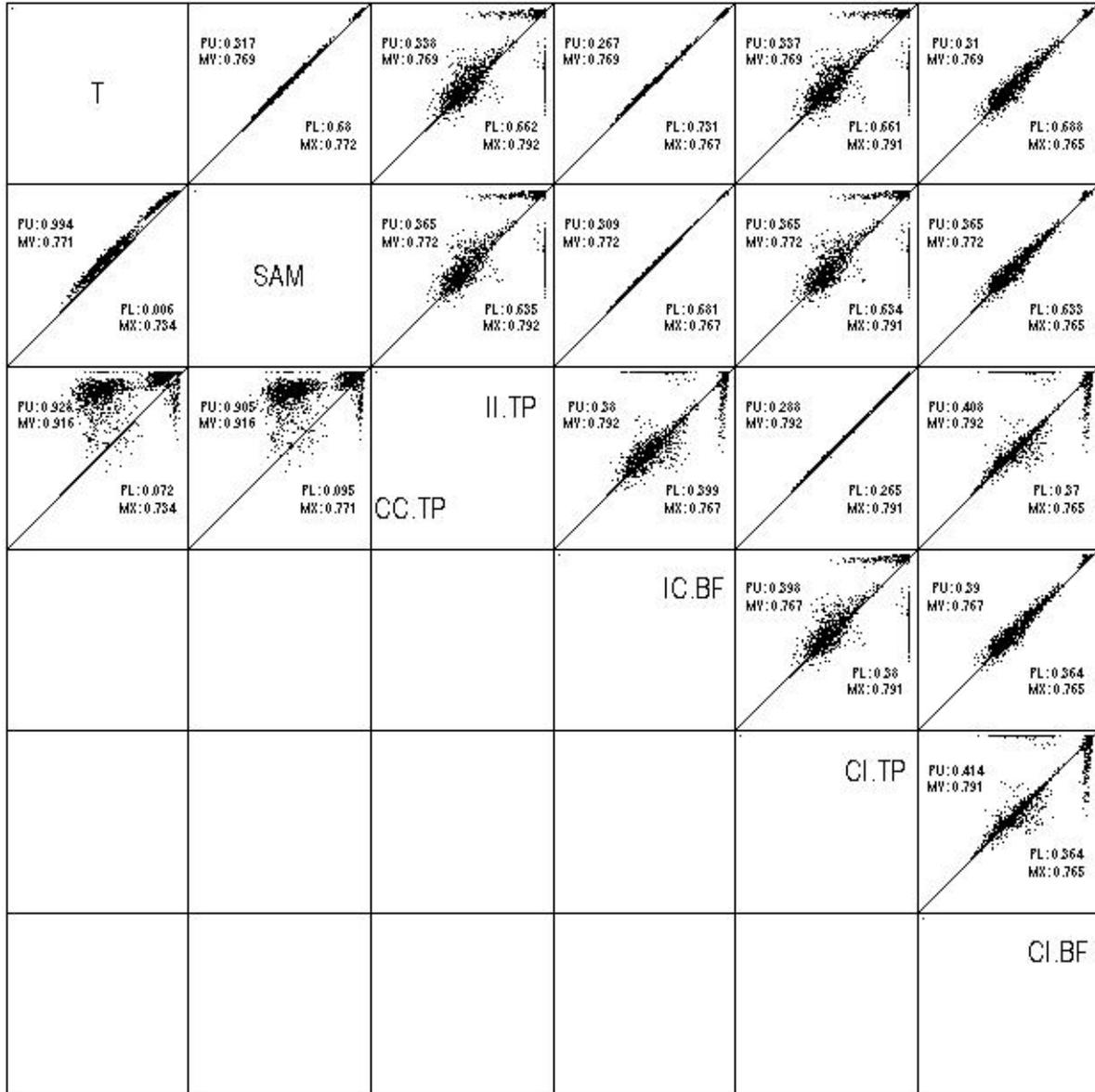


Figure 92: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

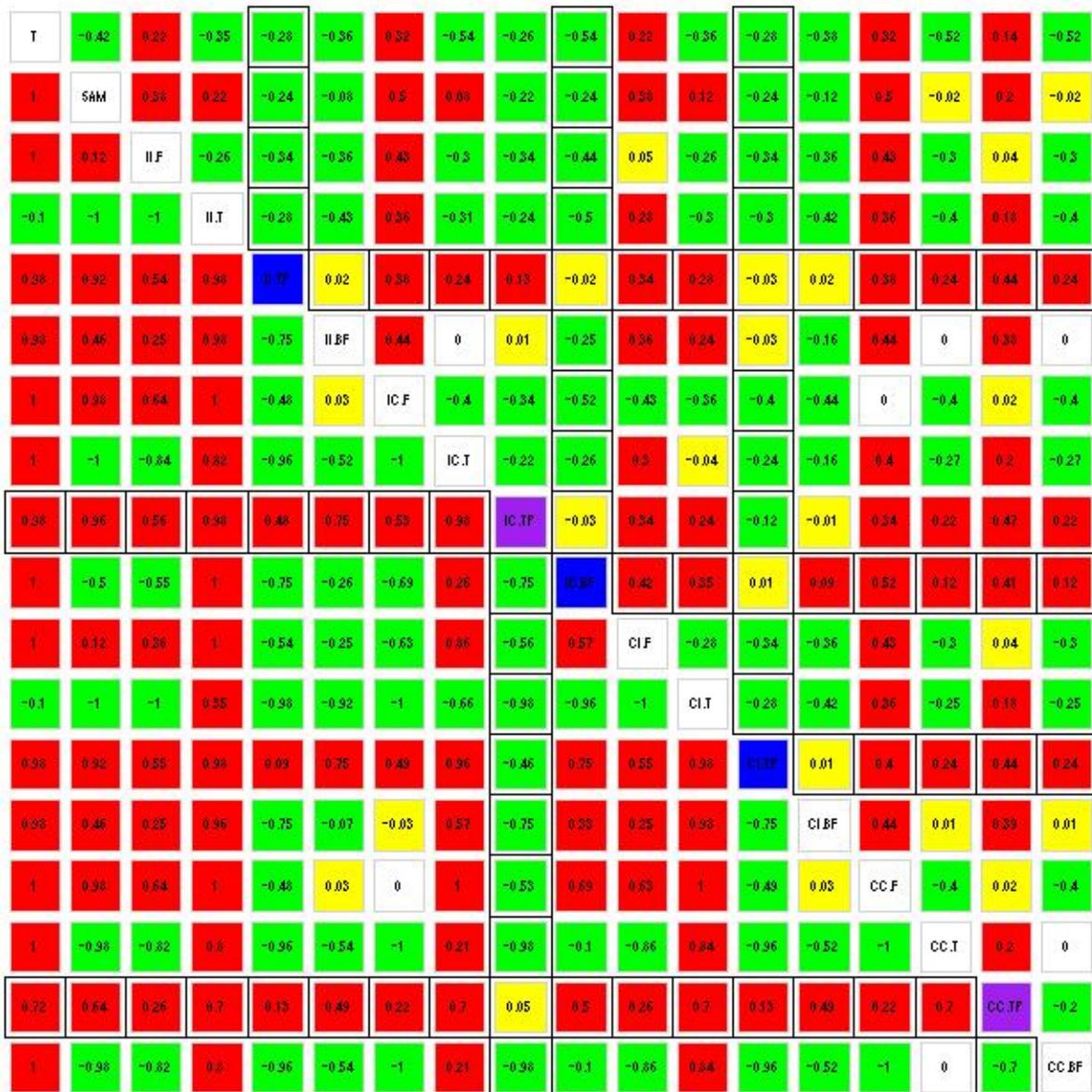


Figure 93: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

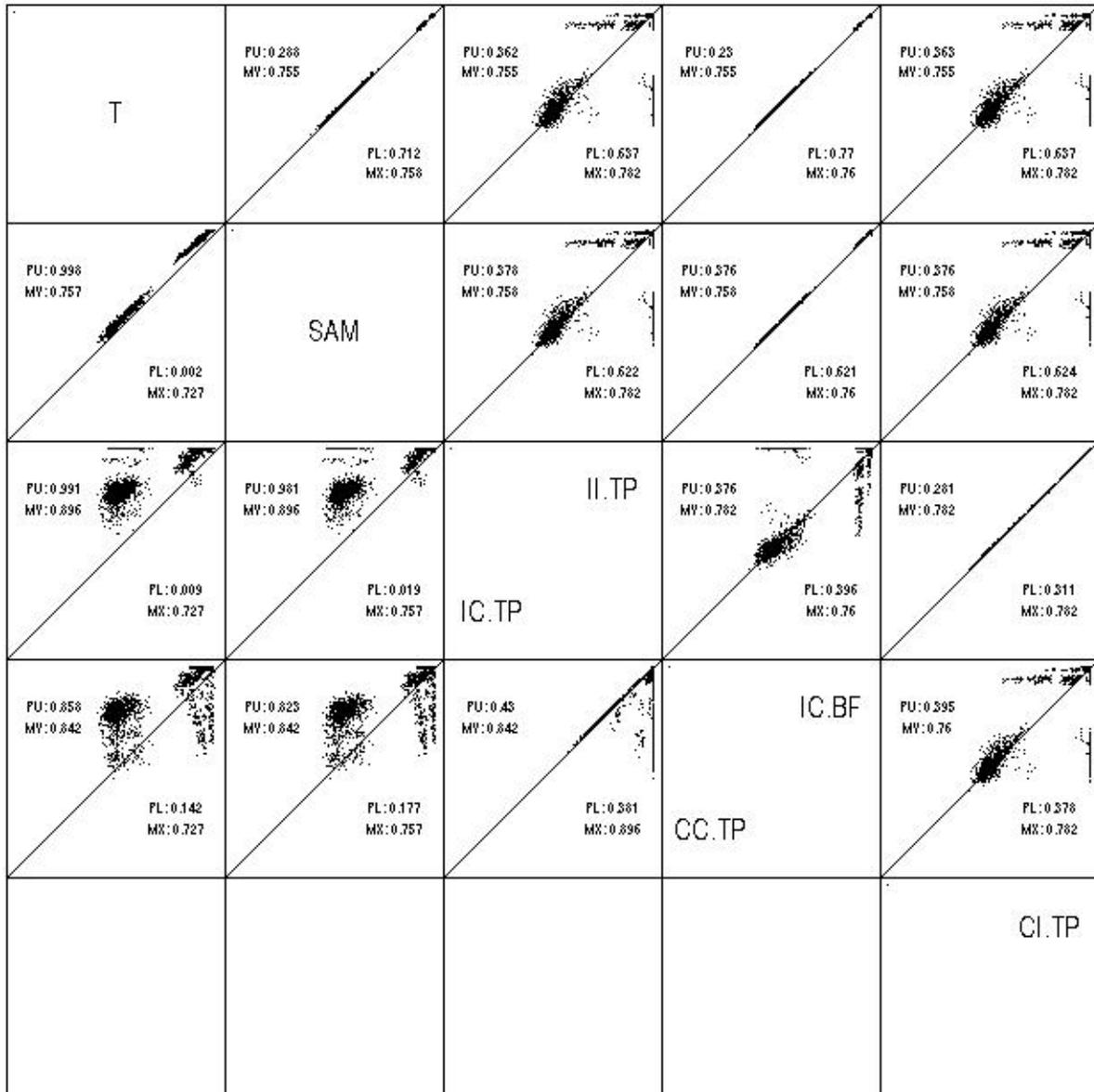


Figure 94: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

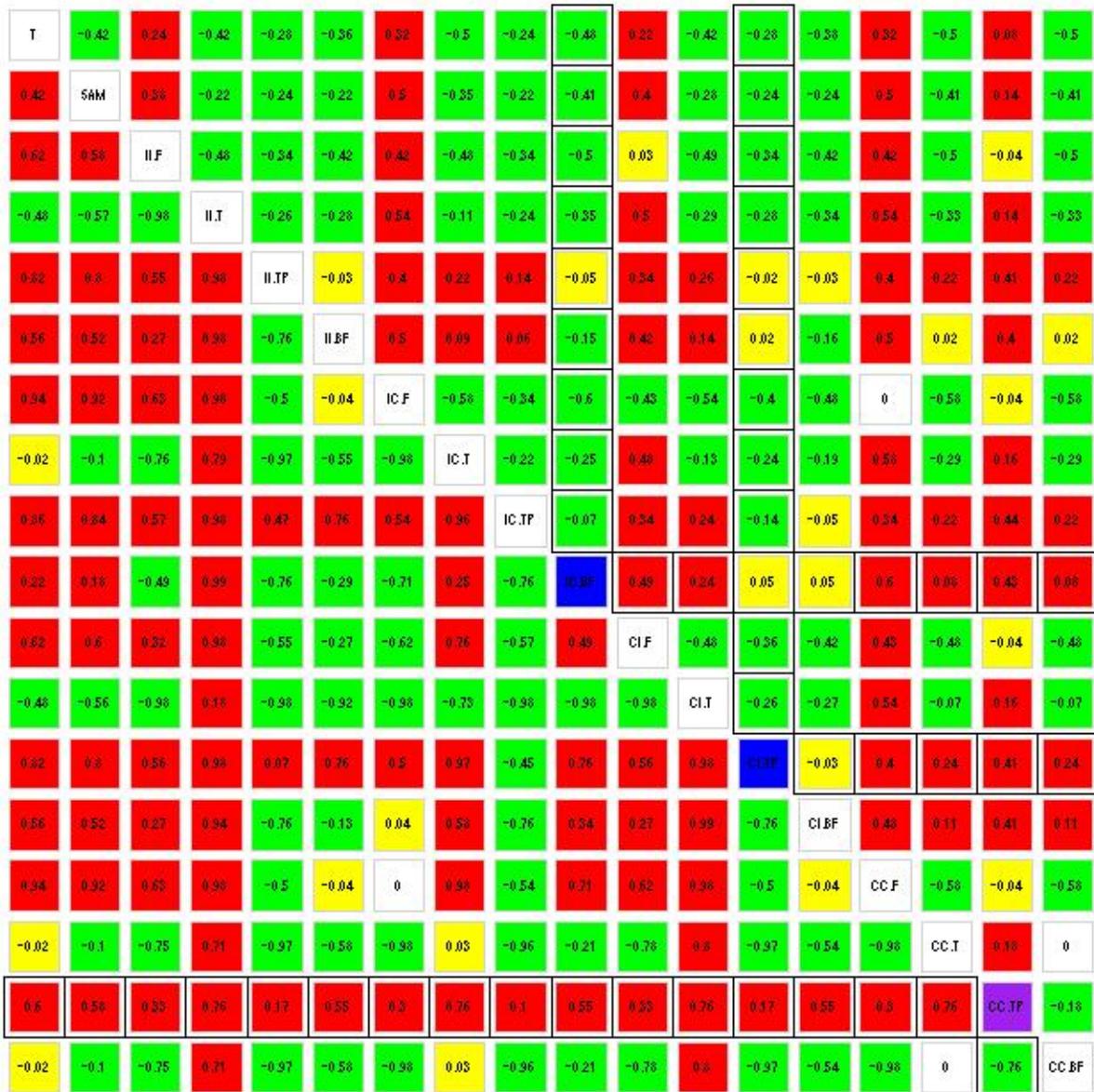


Figure 95: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

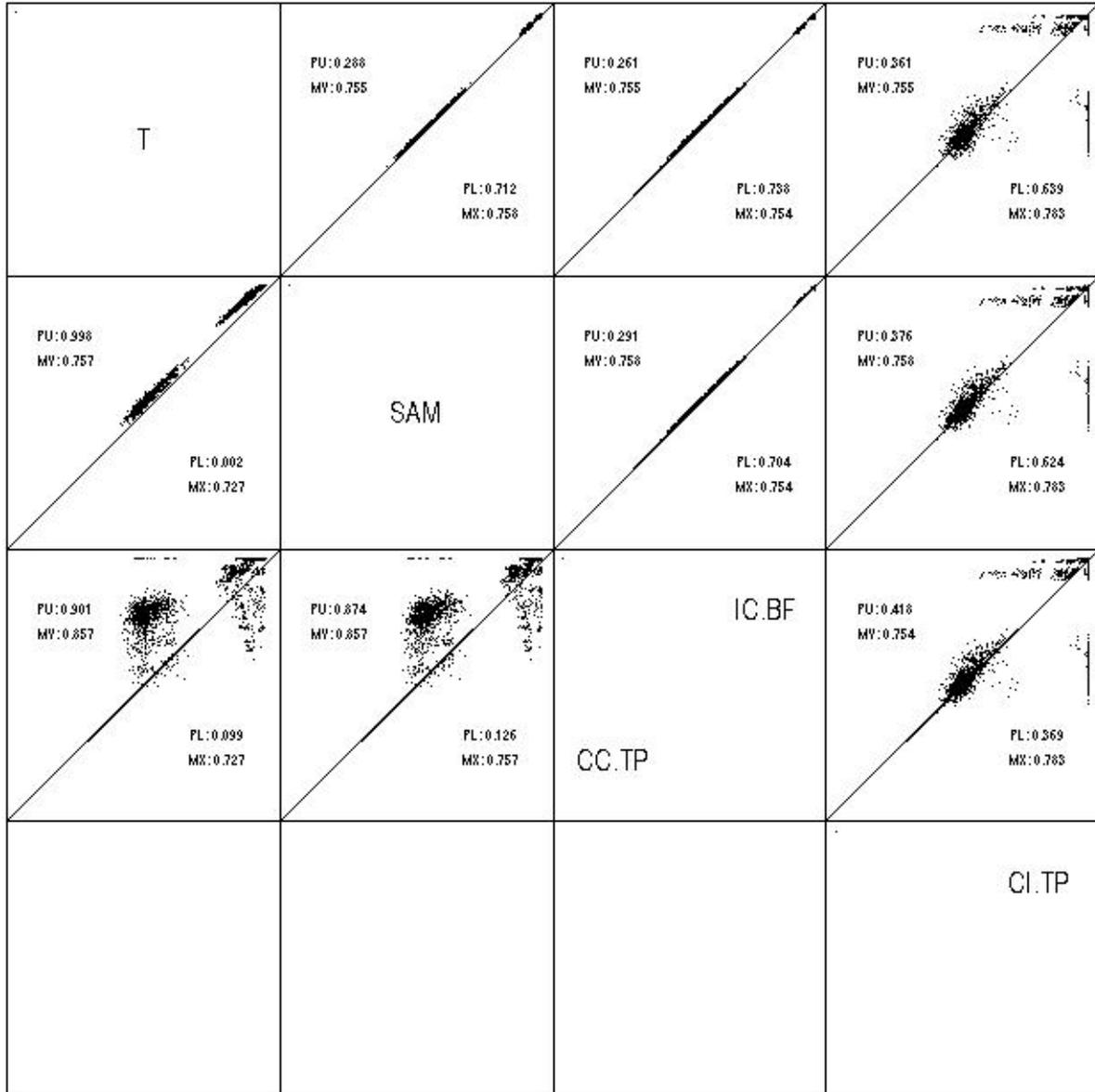


Figure 96: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

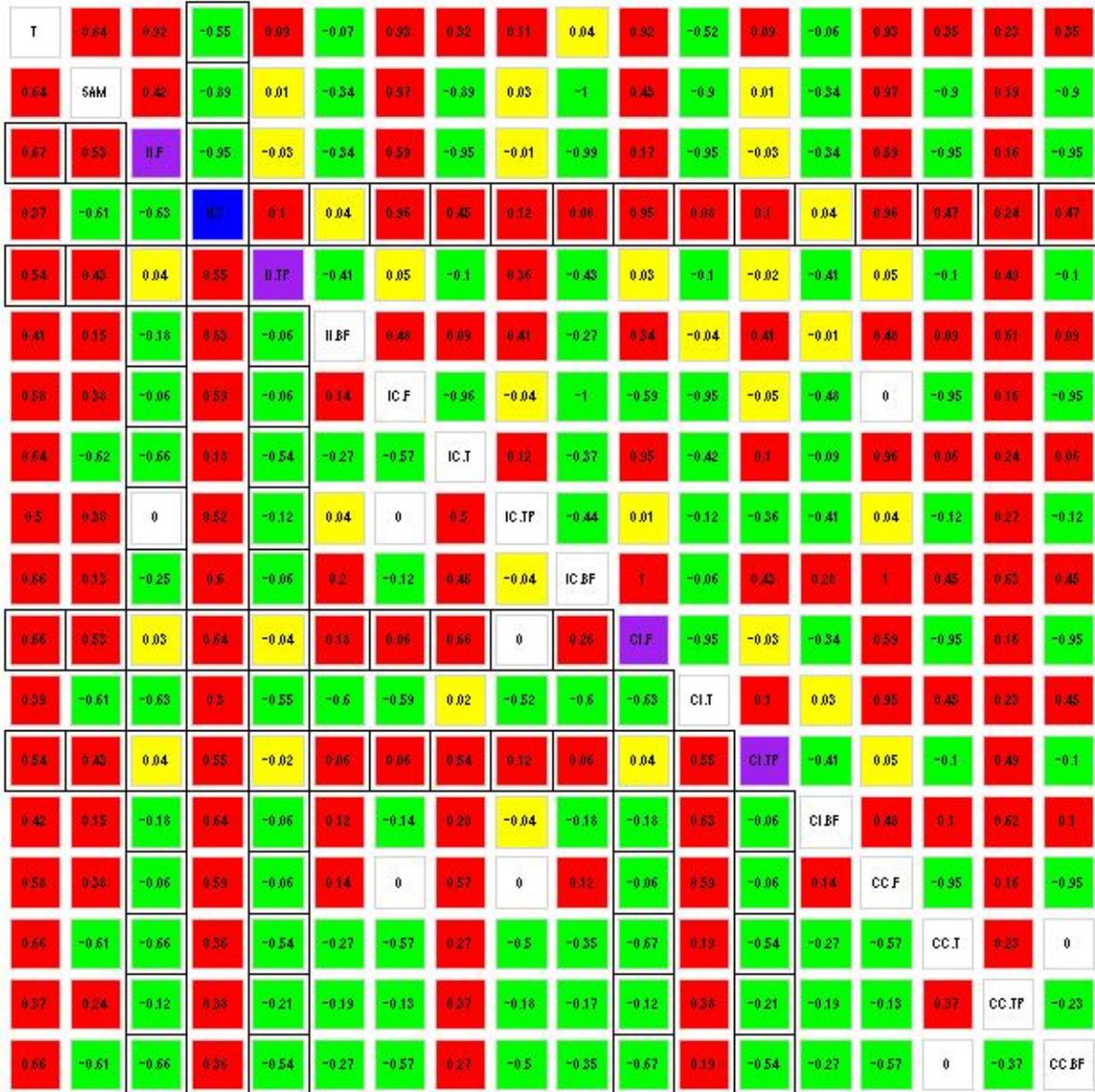


Figure 97: Heat map summary for simulations based on Independence model with number of replicates equal to 100 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

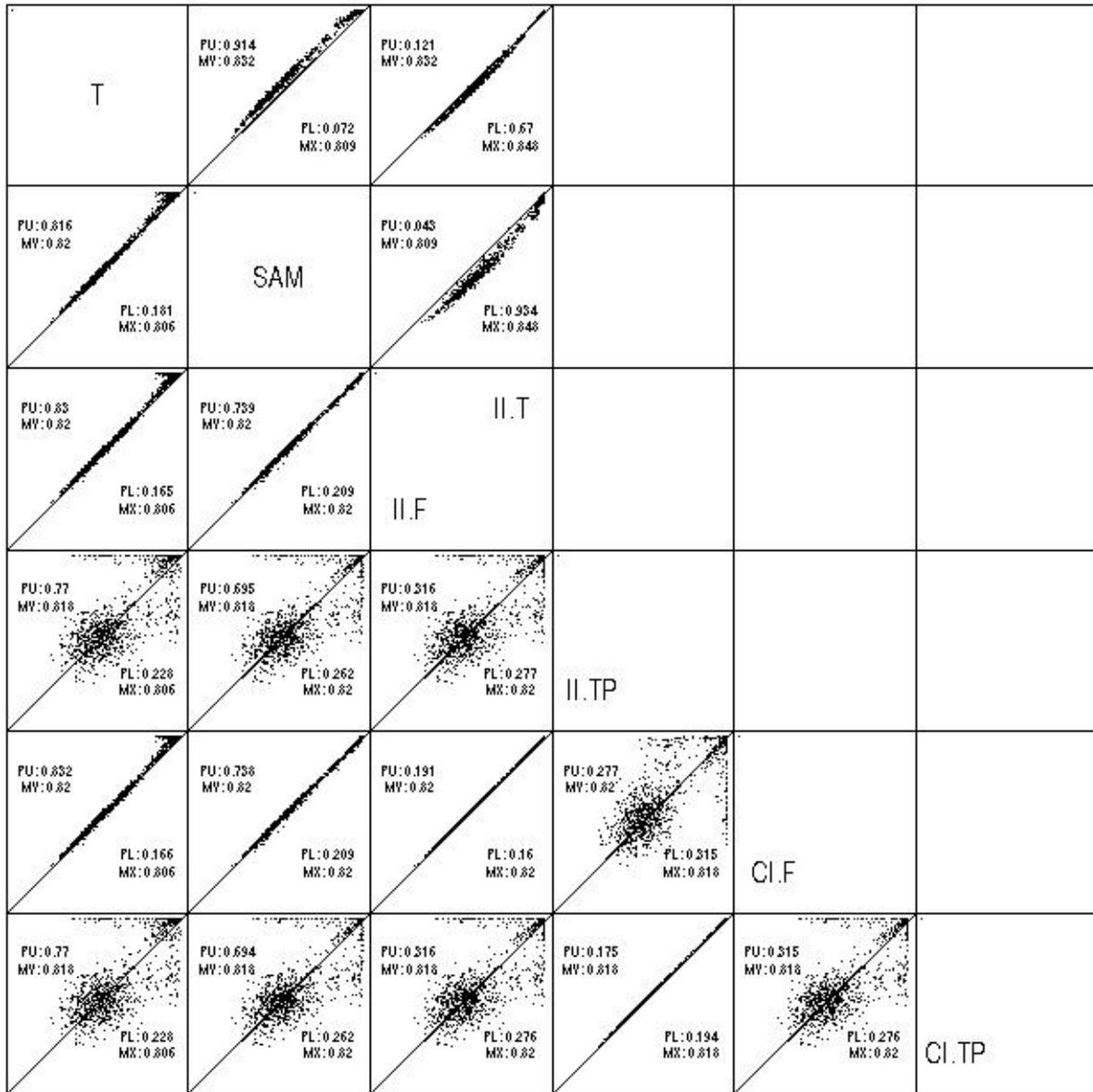


Figure 98: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 100 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

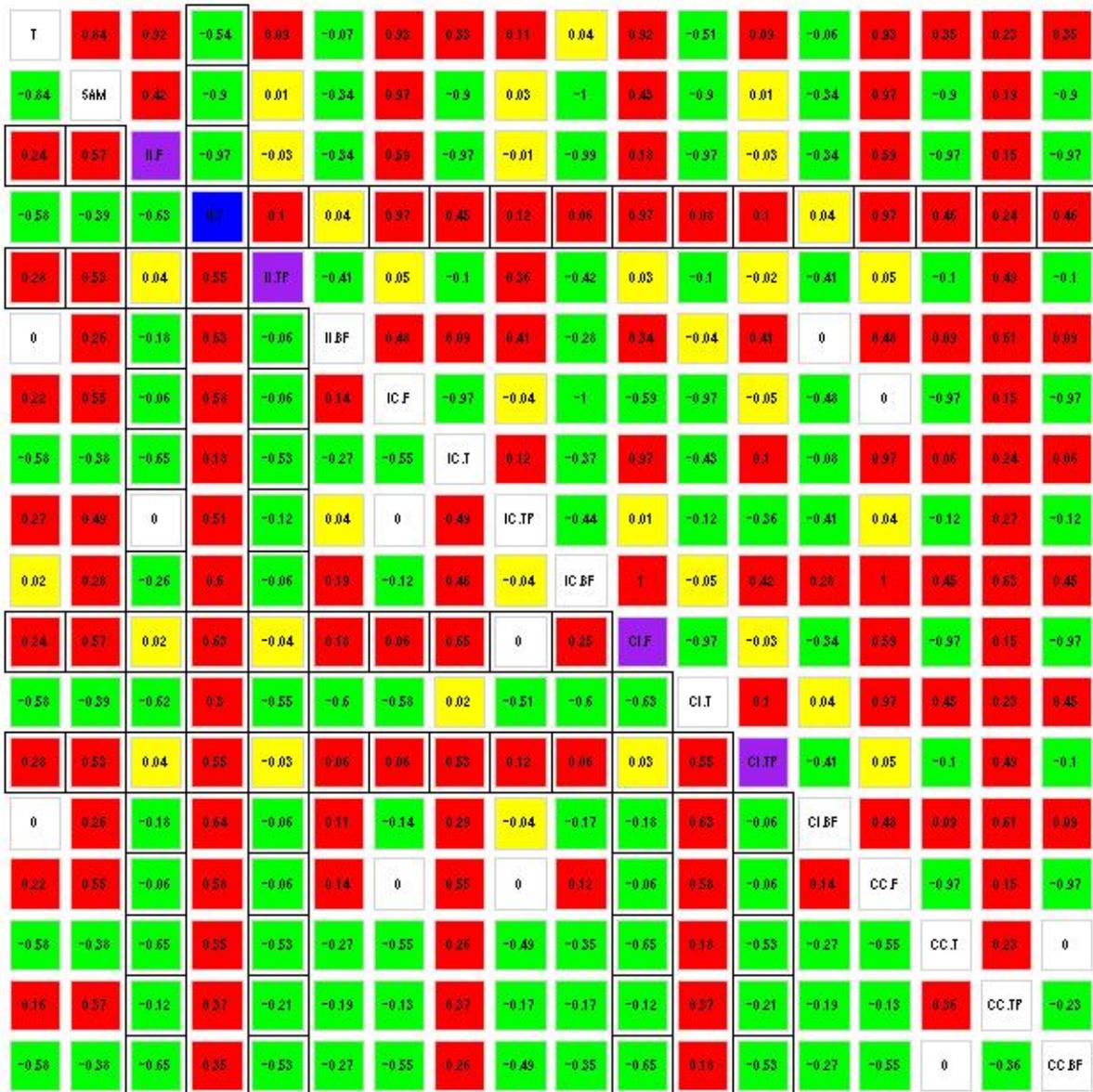


Figure 99: Heat map summary for simulations based on Independence model with number of replicates equal to 100 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

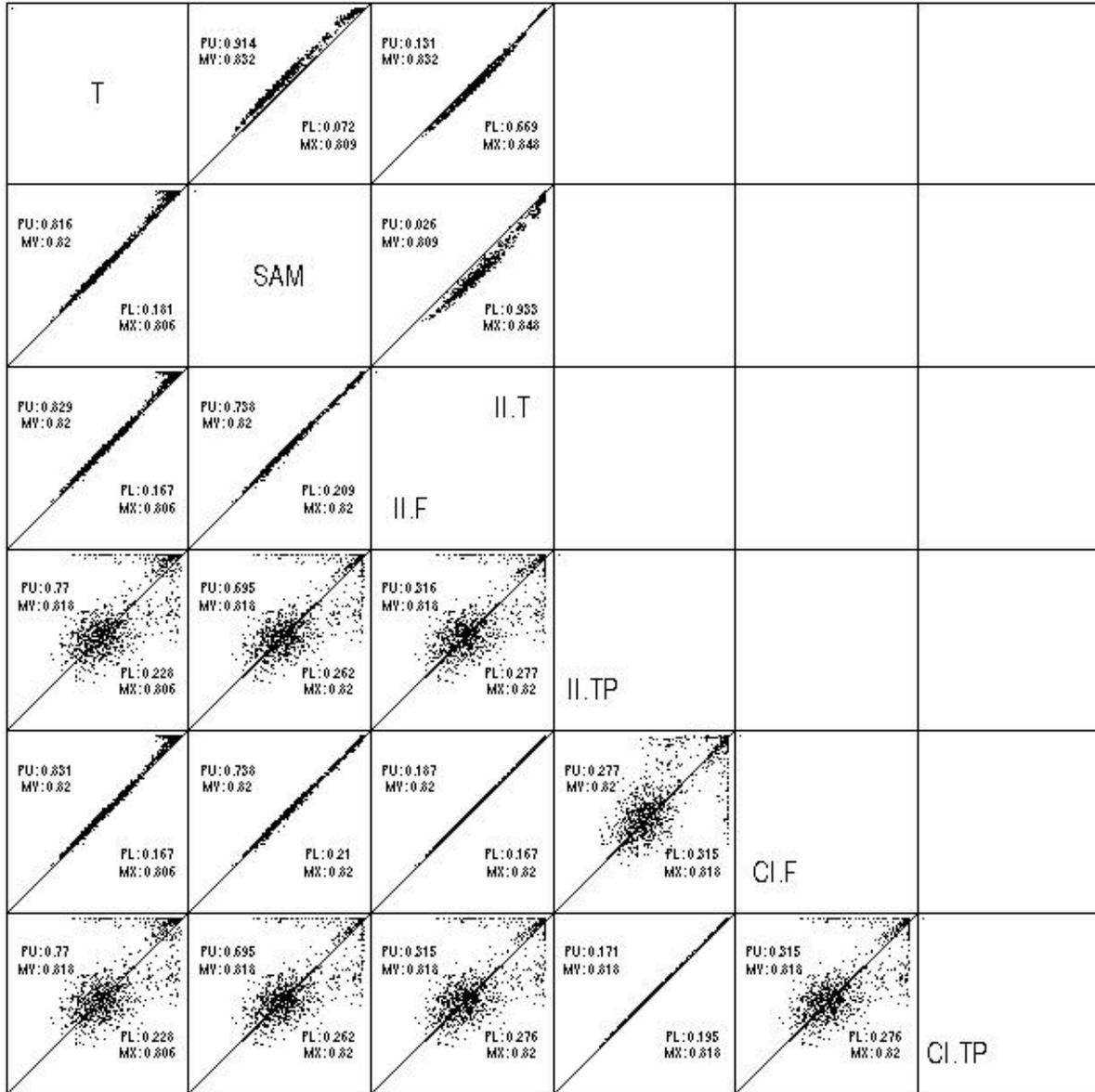


Figure 100: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 100 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

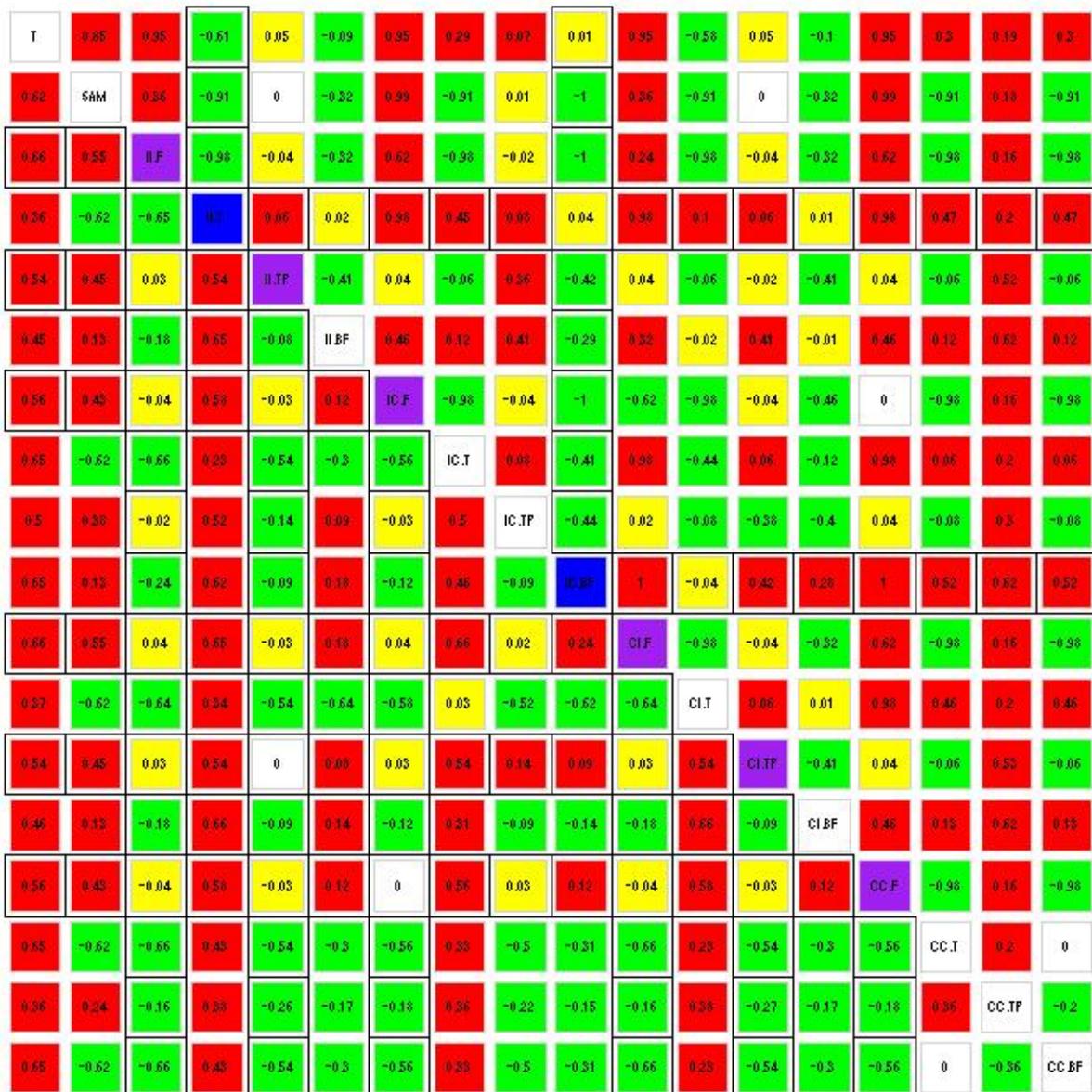


Figure 101: Heat map summary for simulations based on Independence model with number of replicates equal to 100 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

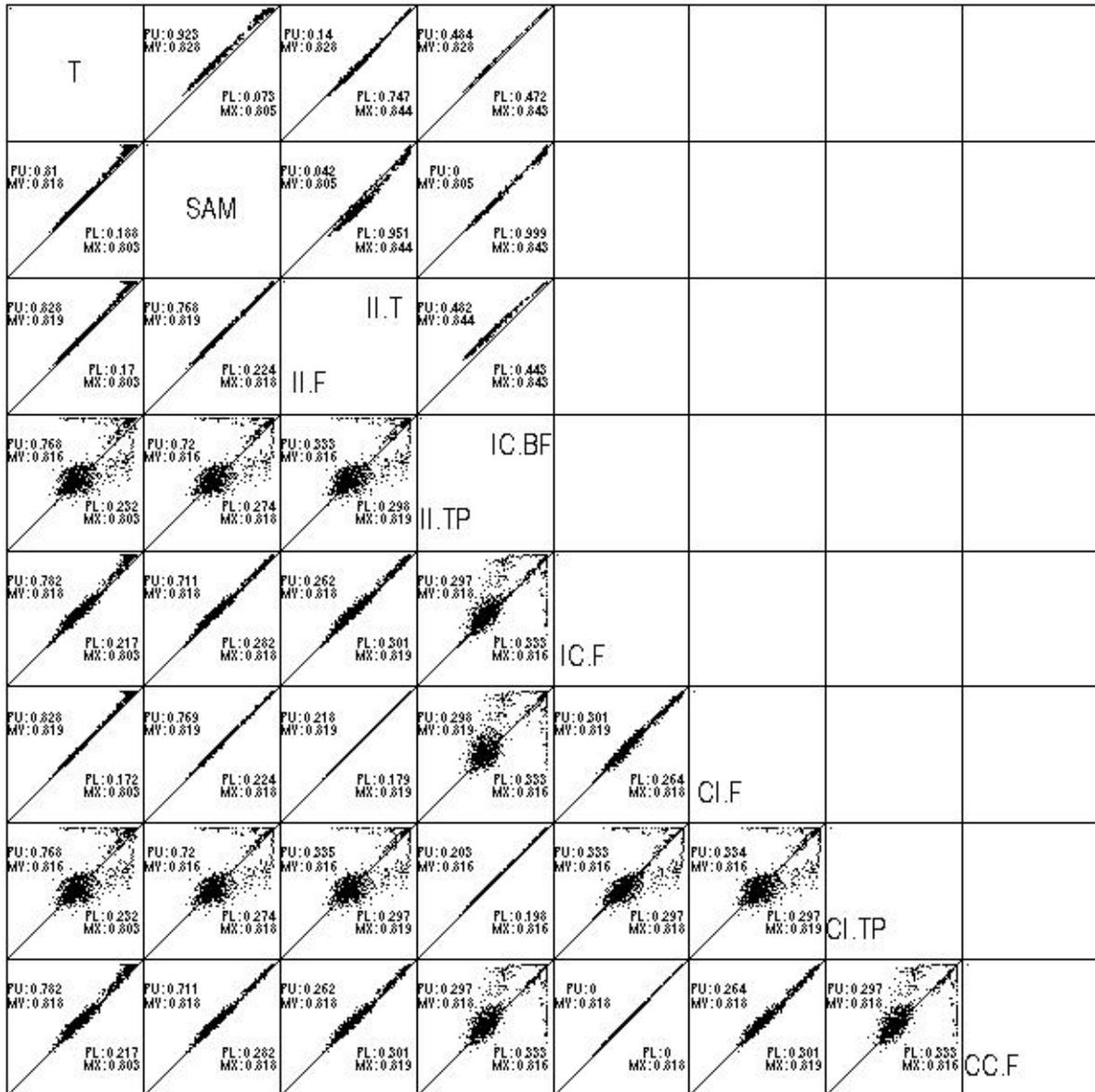


Figure 102: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 100 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

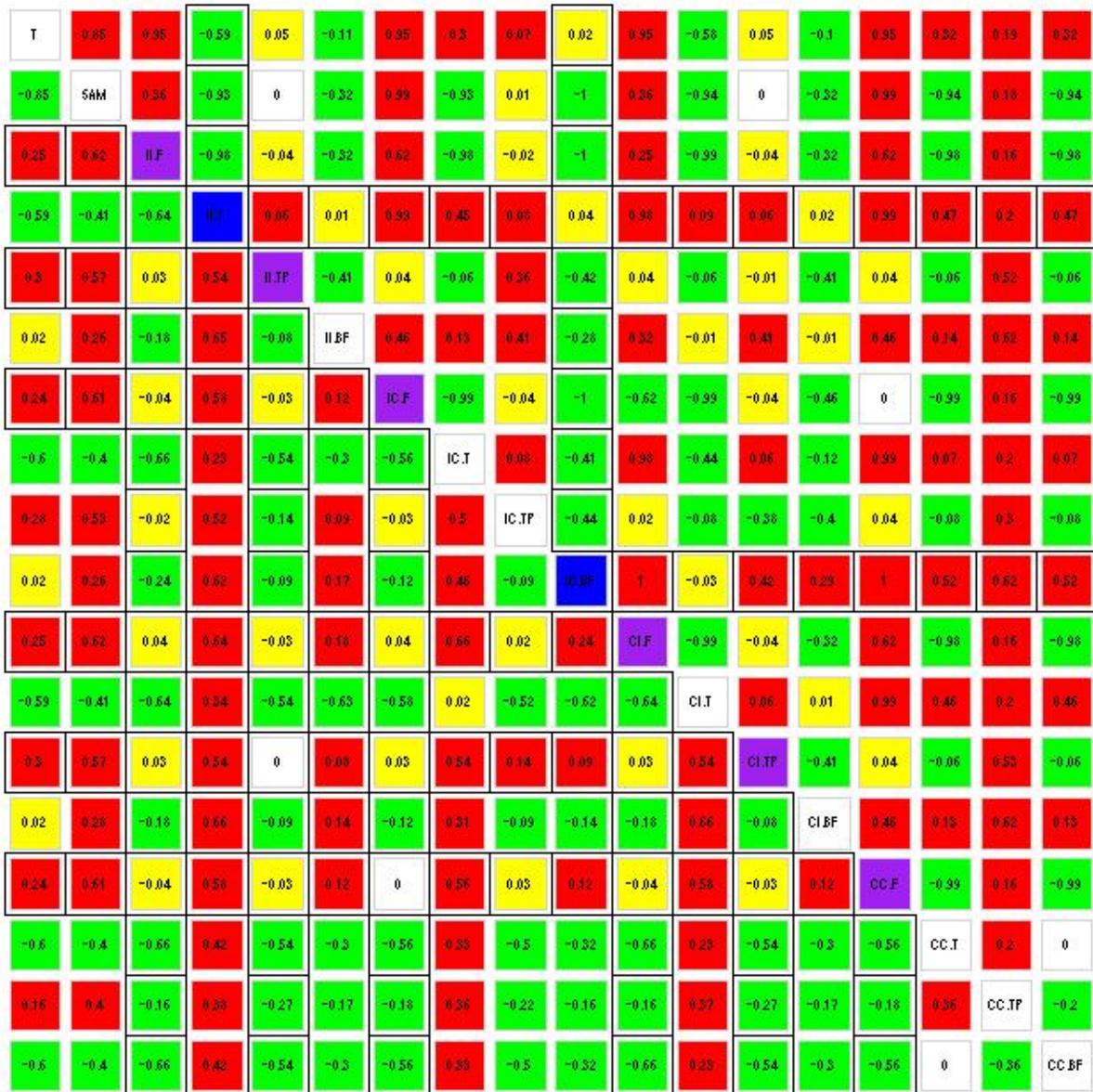


Figure 103: Heat map summary for simulations based on Independence model with number of replicates equal to 100 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

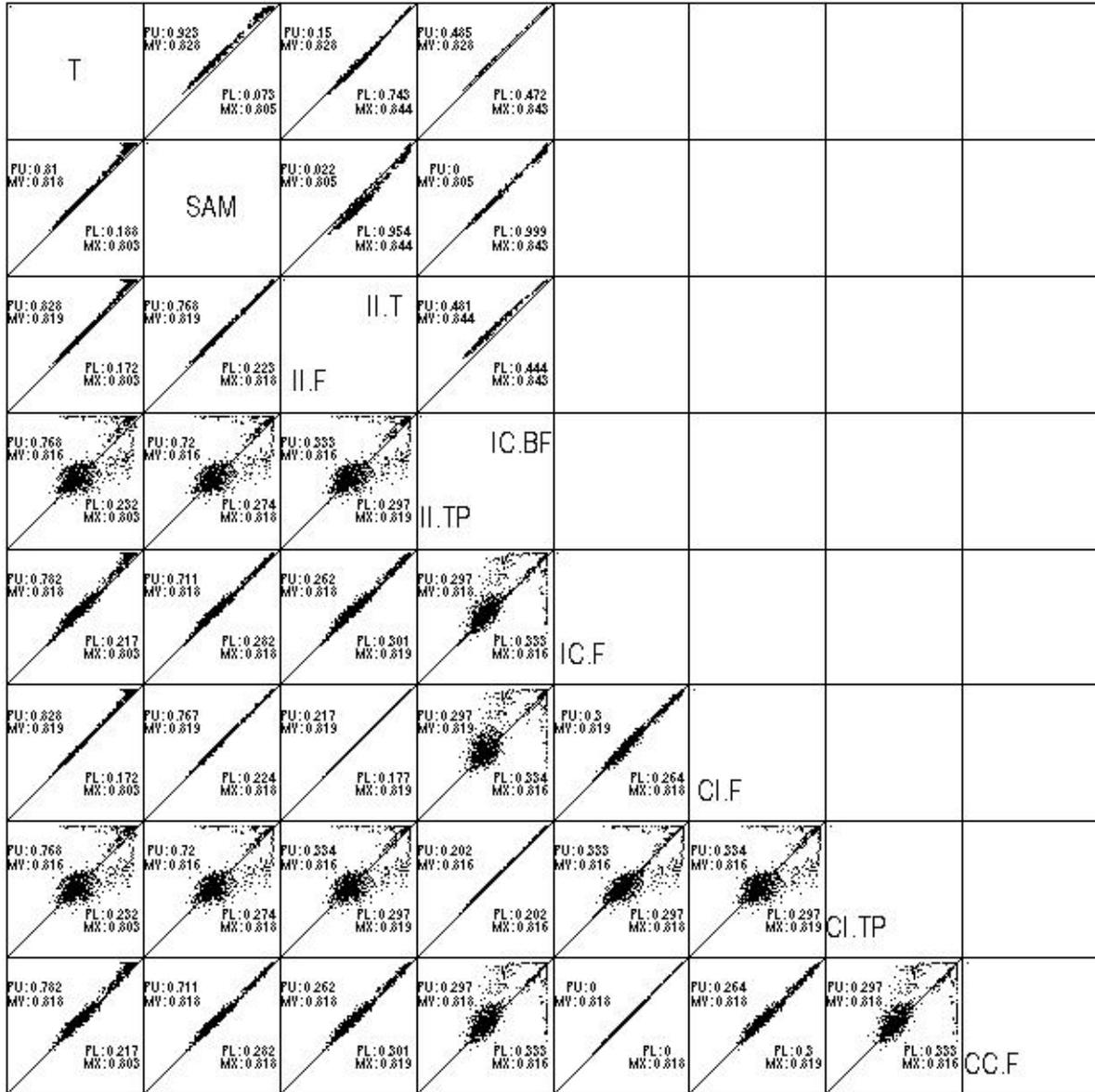


Figure 104: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 100 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

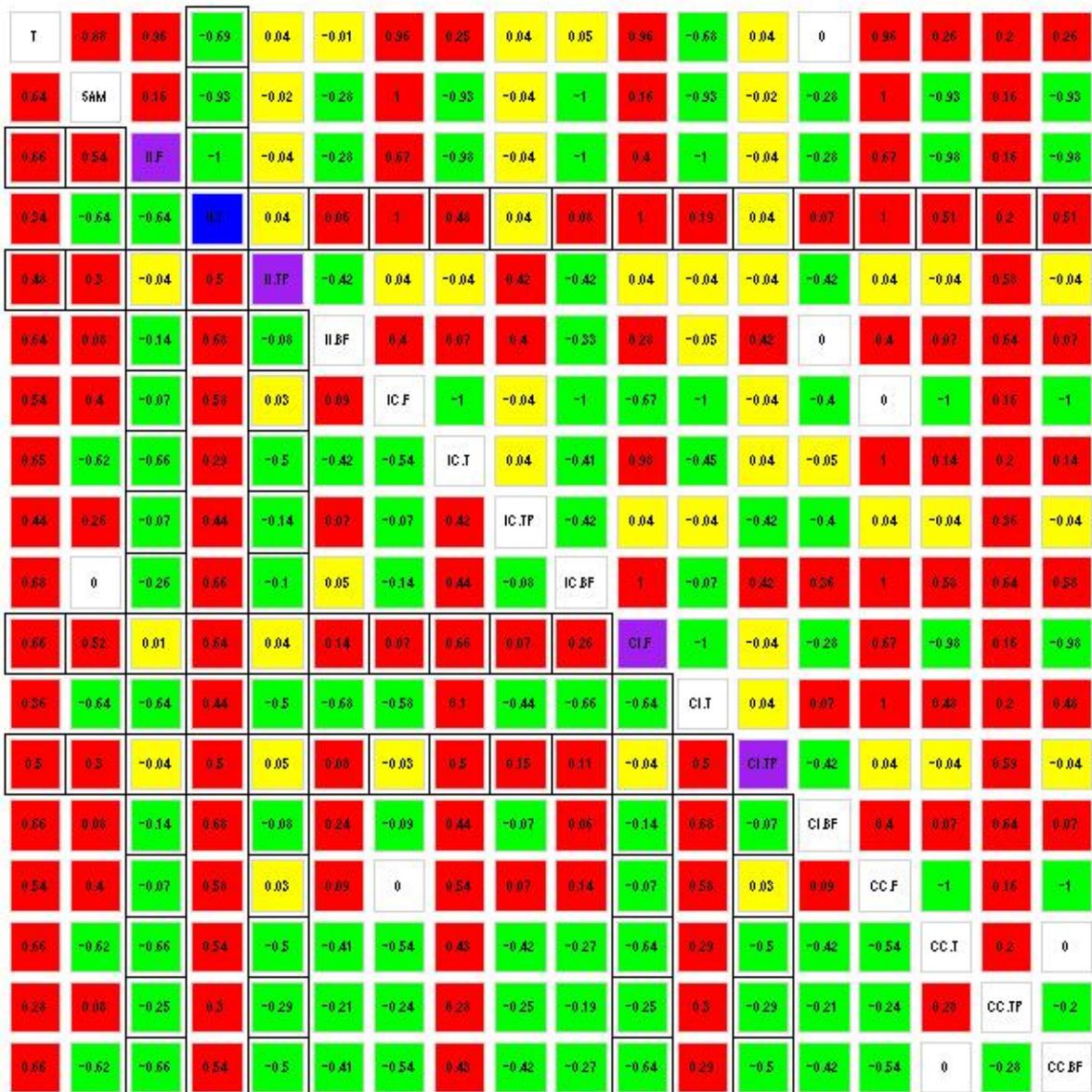


Figure 105: Heat map summary for simulations based on Independence model with number of replicates equal to 100 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

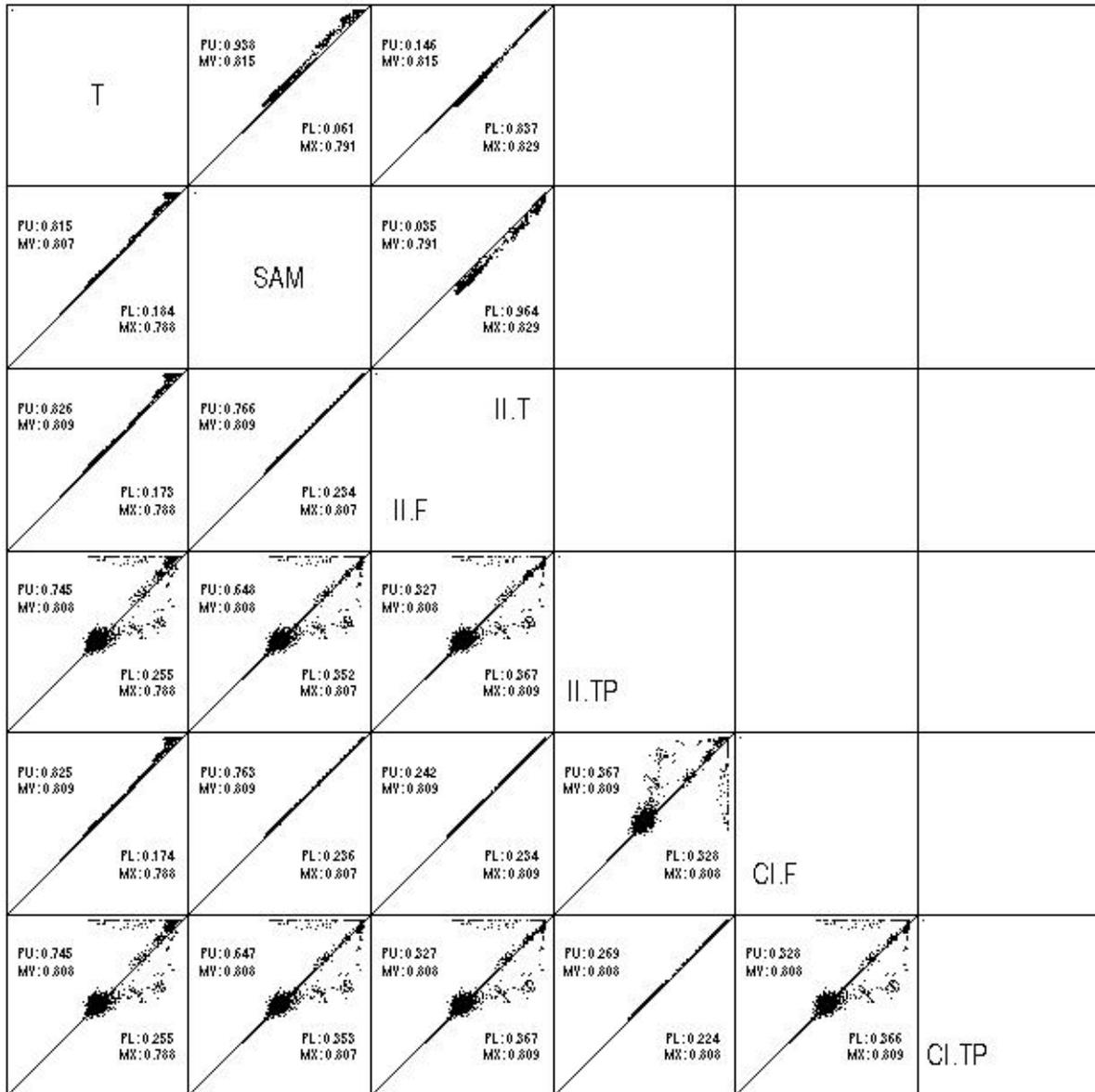


Figure 106: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

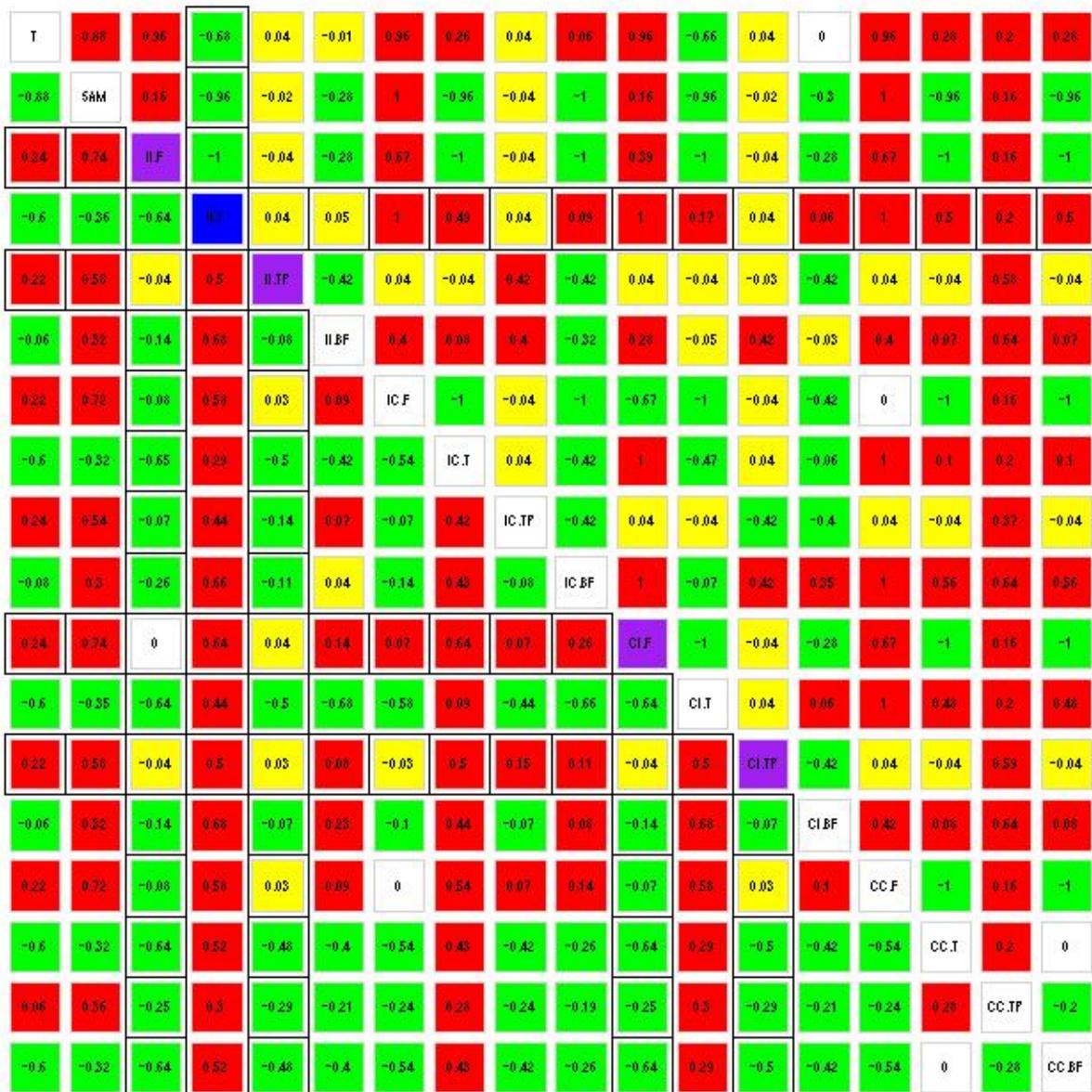


Figure 107: Heat map summary for simulations based on Independence model with number of replicates equal to 100 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

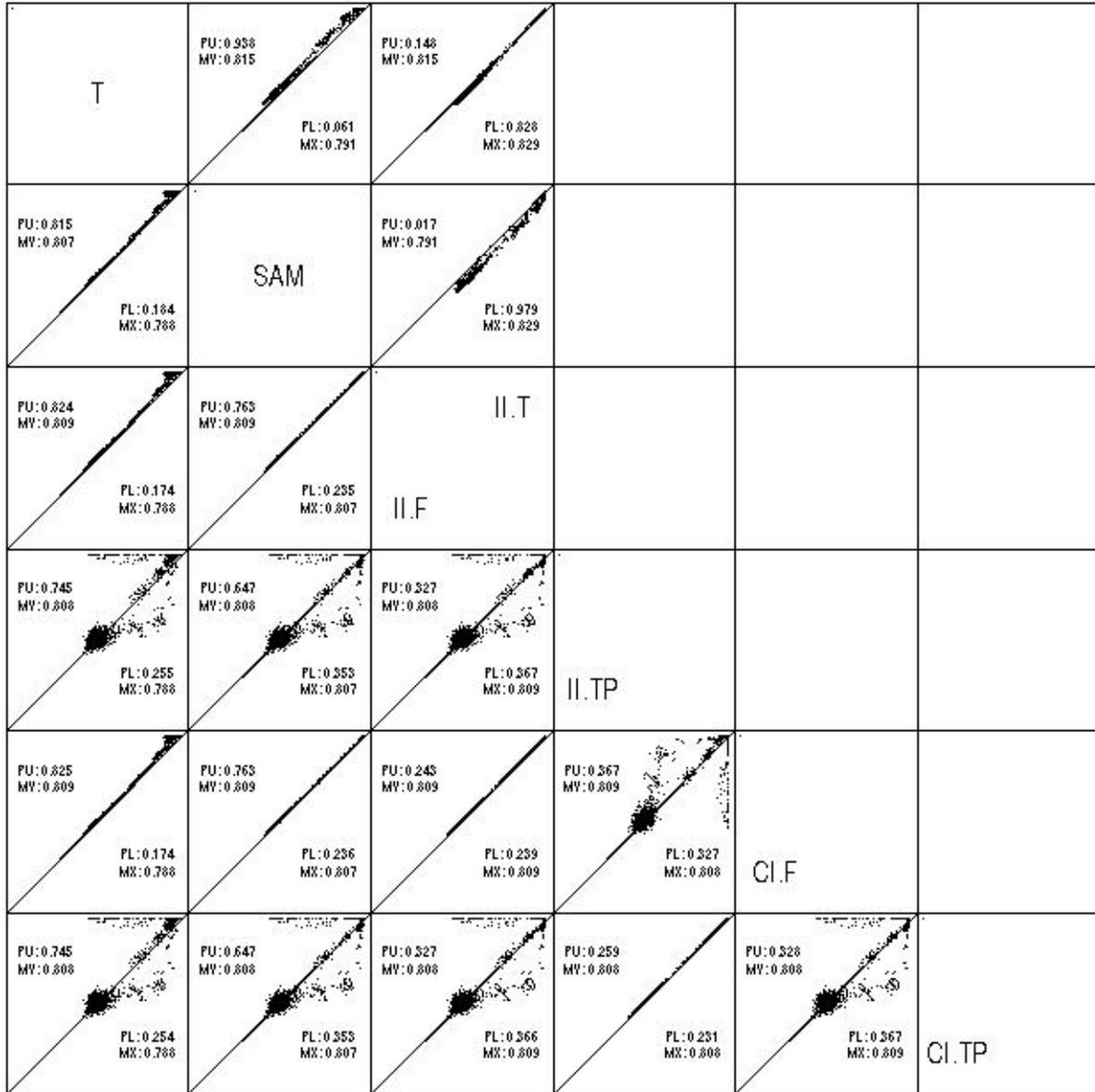


Figure 108: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

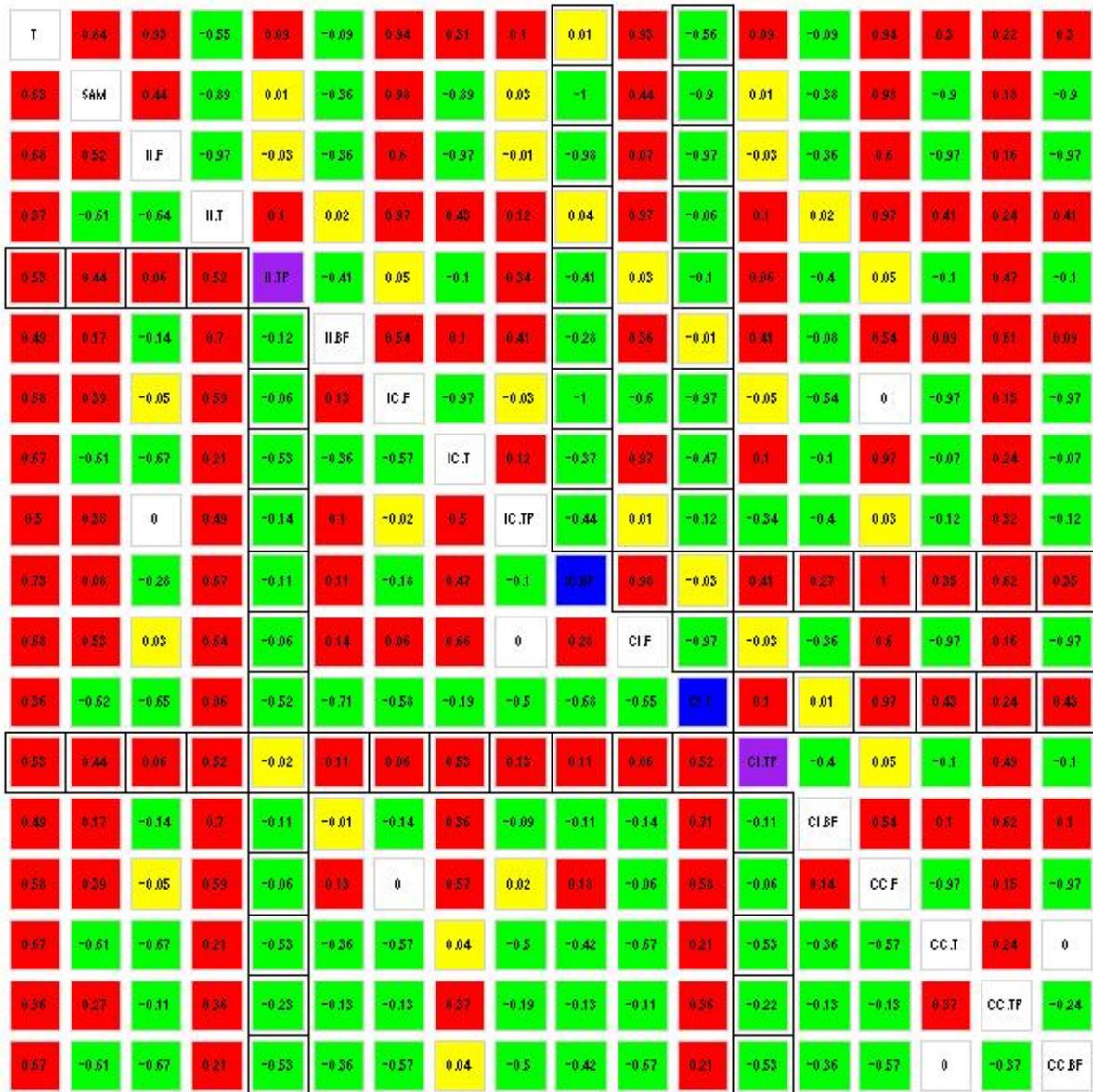


Figure 109: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 100 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

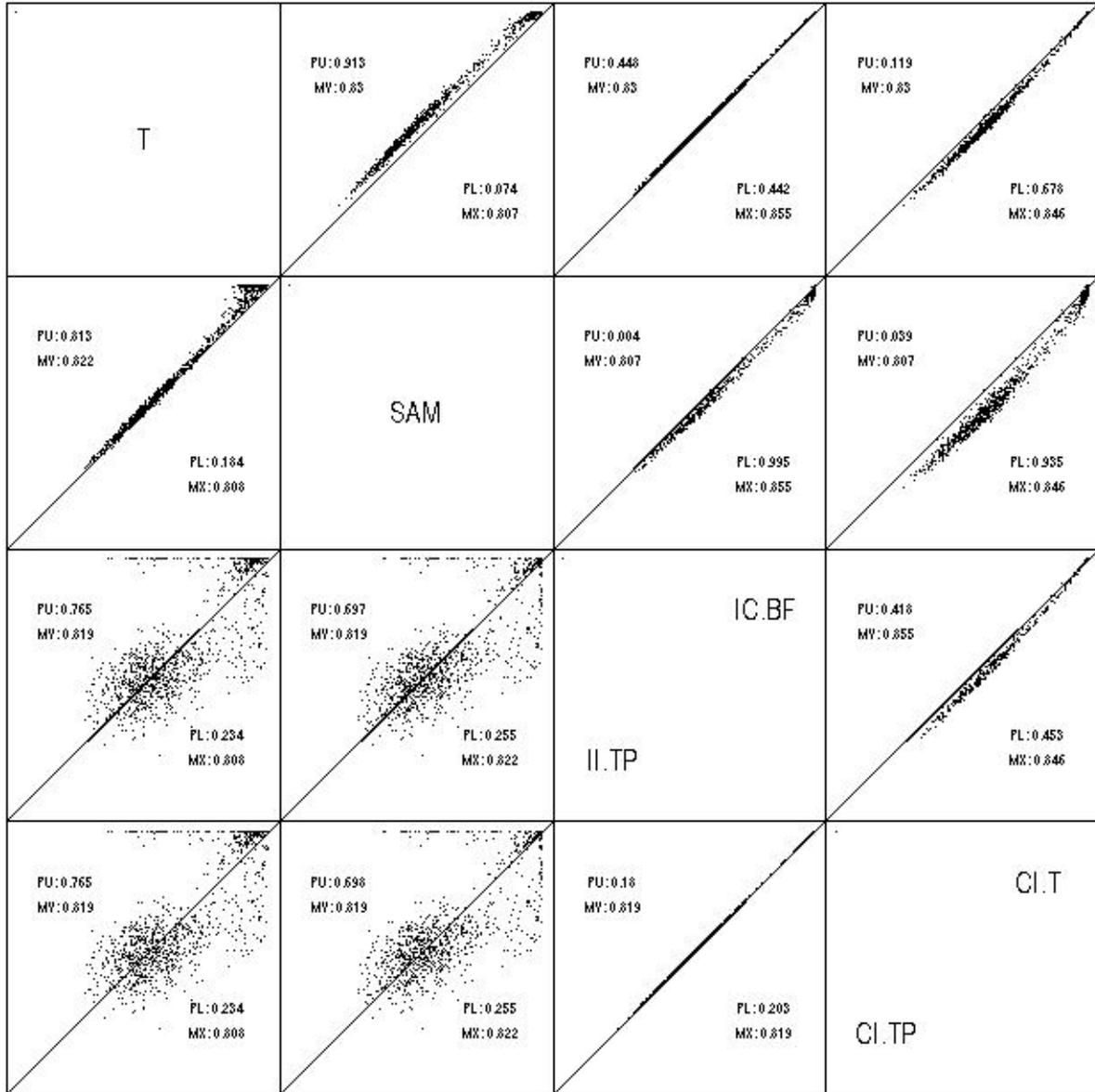


Figure 110: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

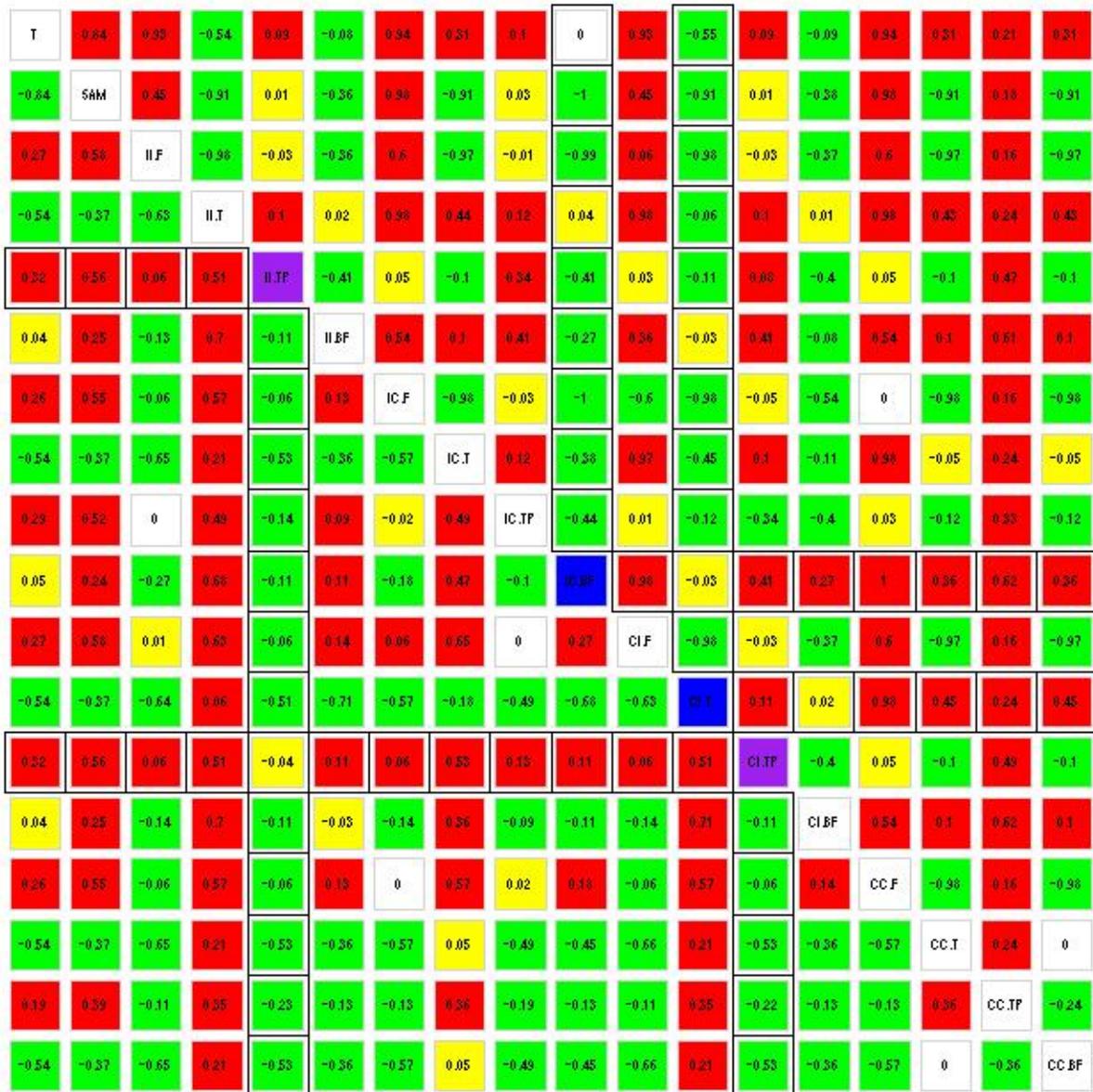


Figure 111: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 100 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

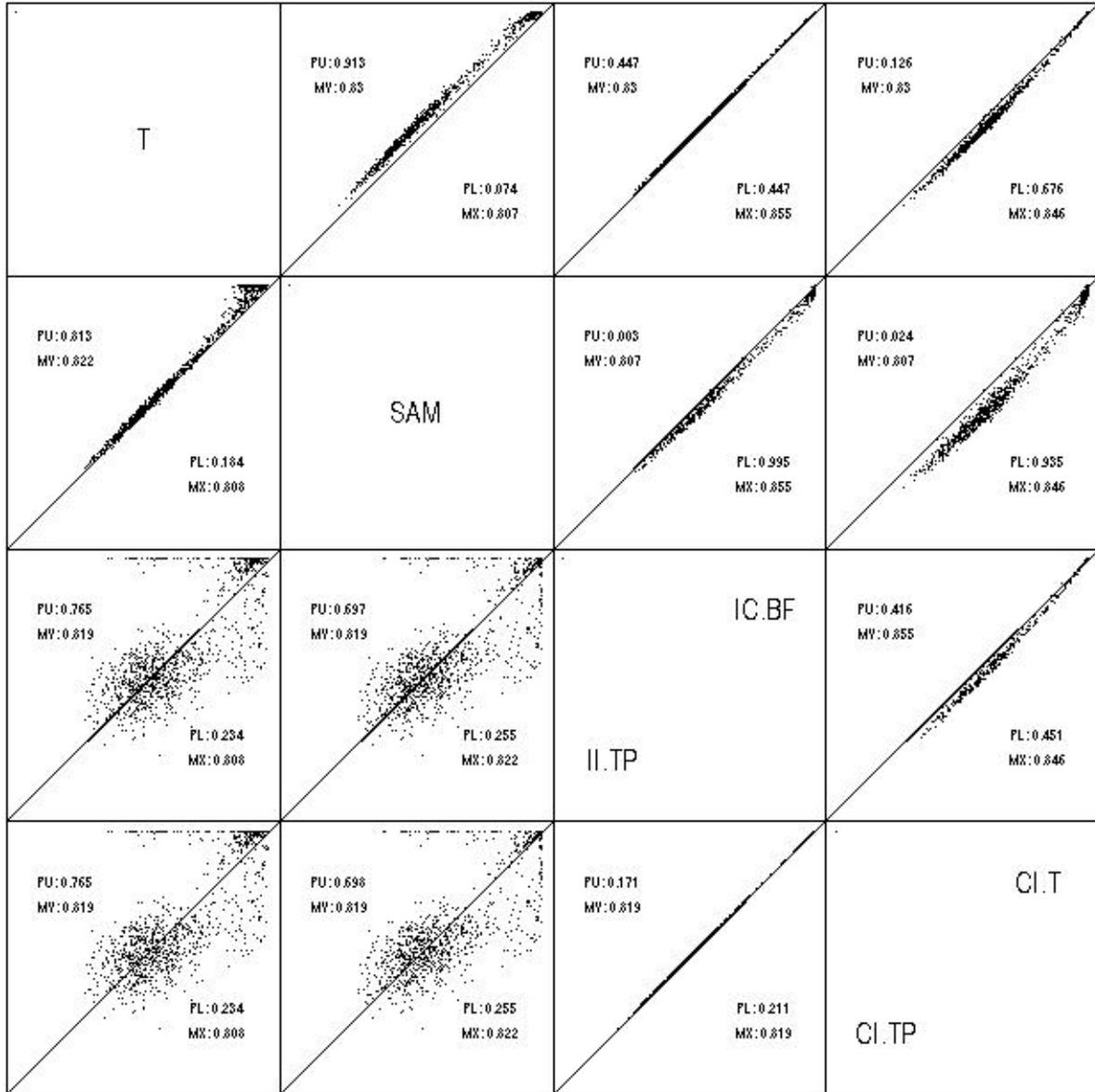


Figure 112: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

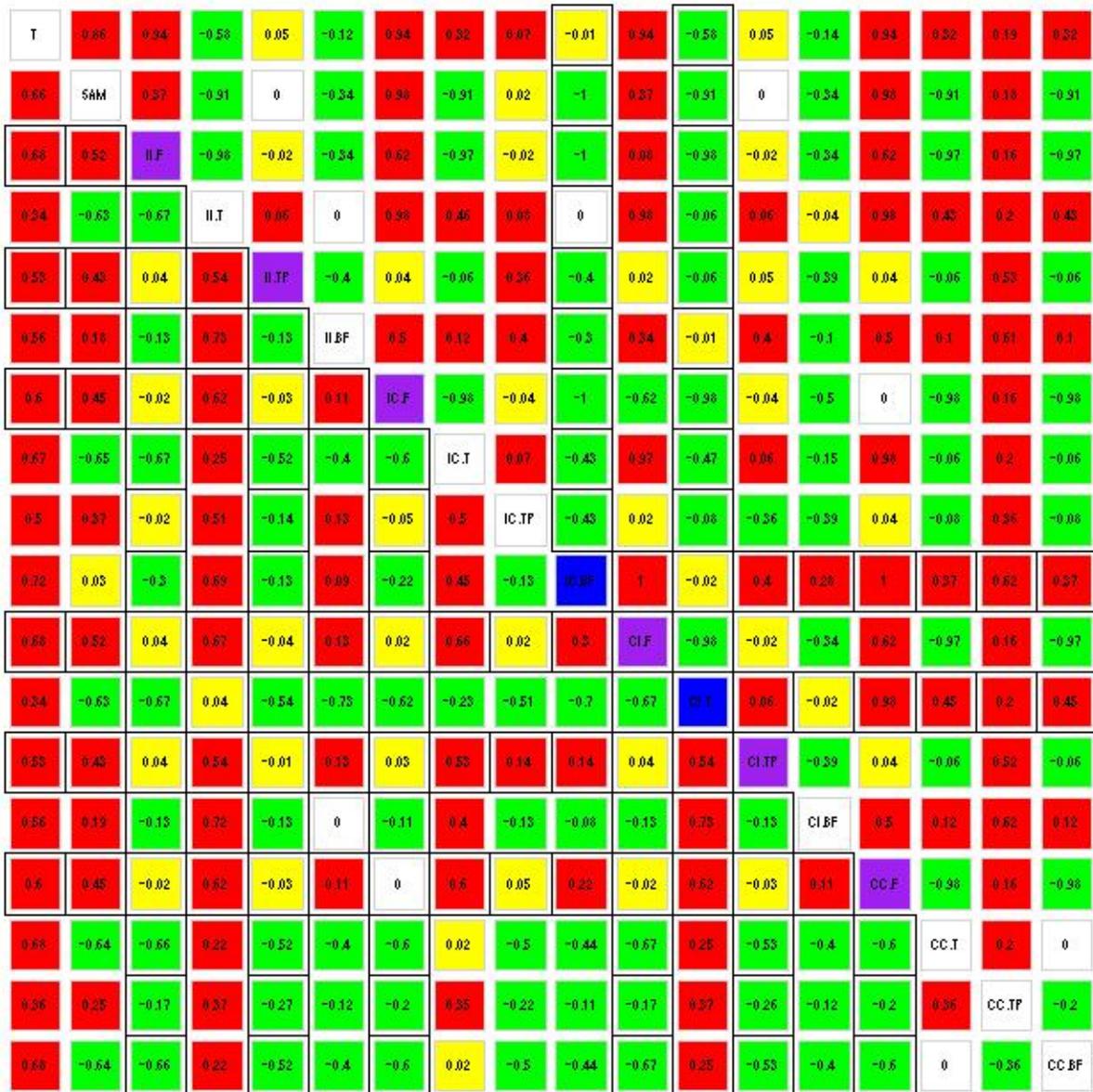


Figure 113: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 100 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

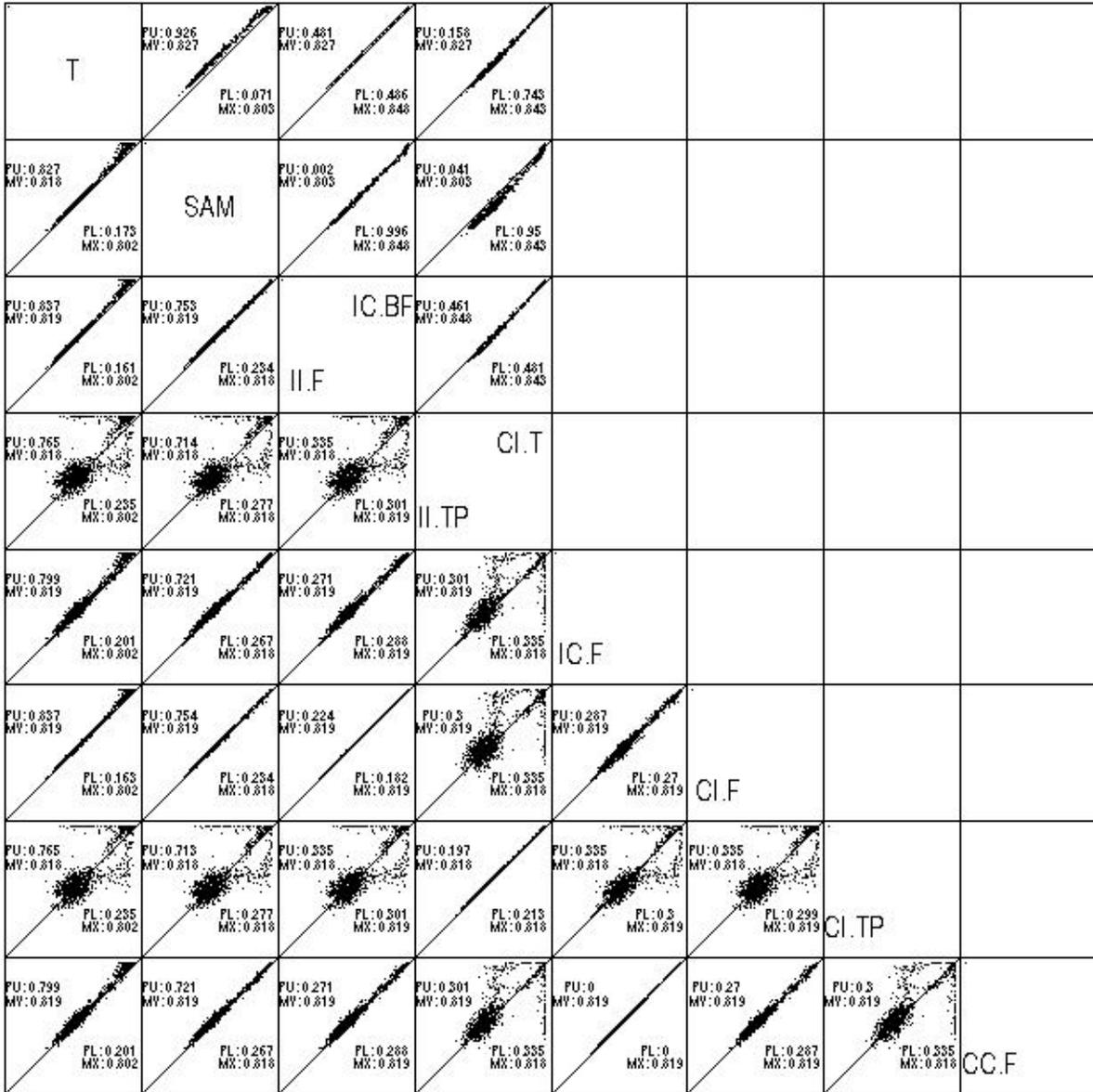


Figure 114: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

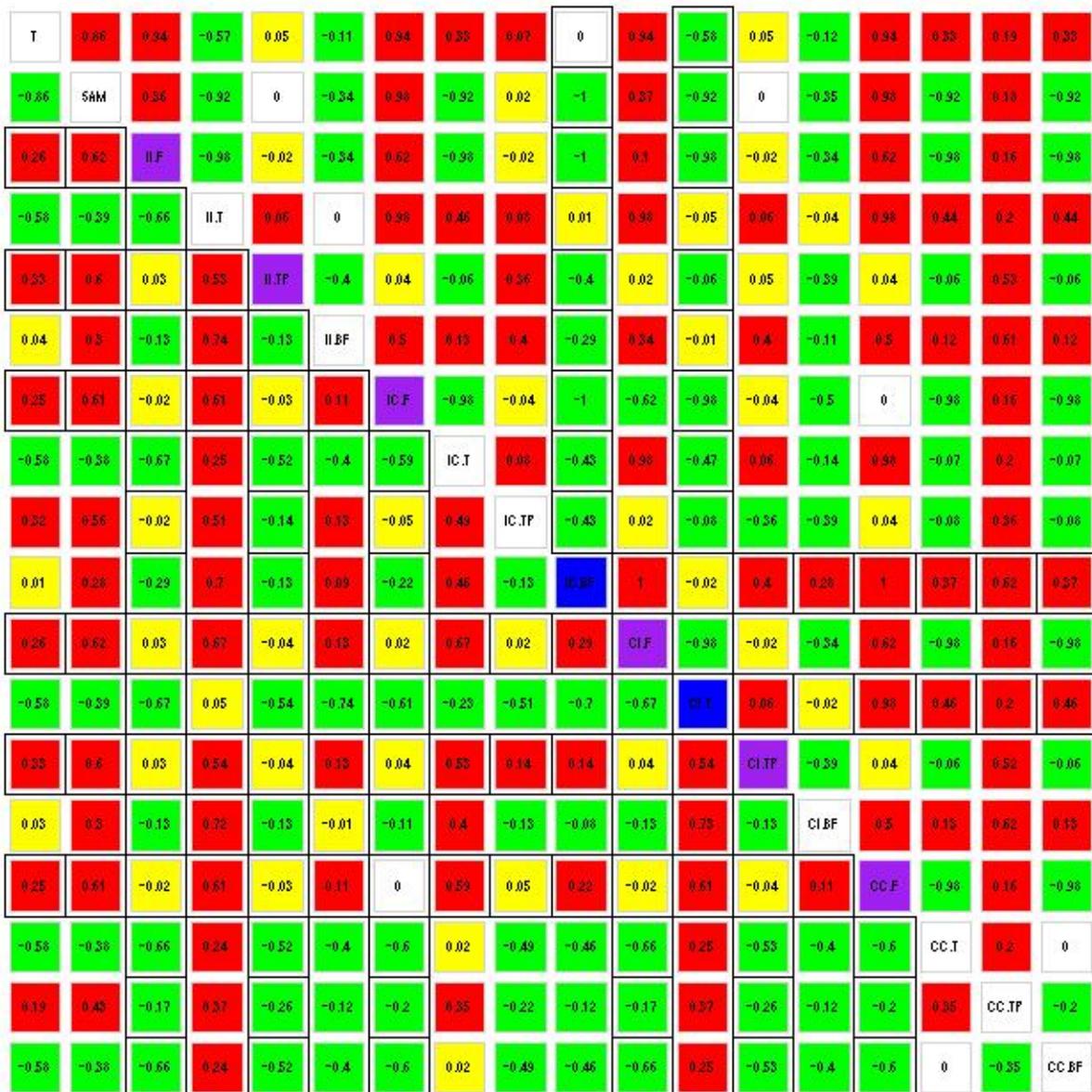


Figure 115: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 100 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

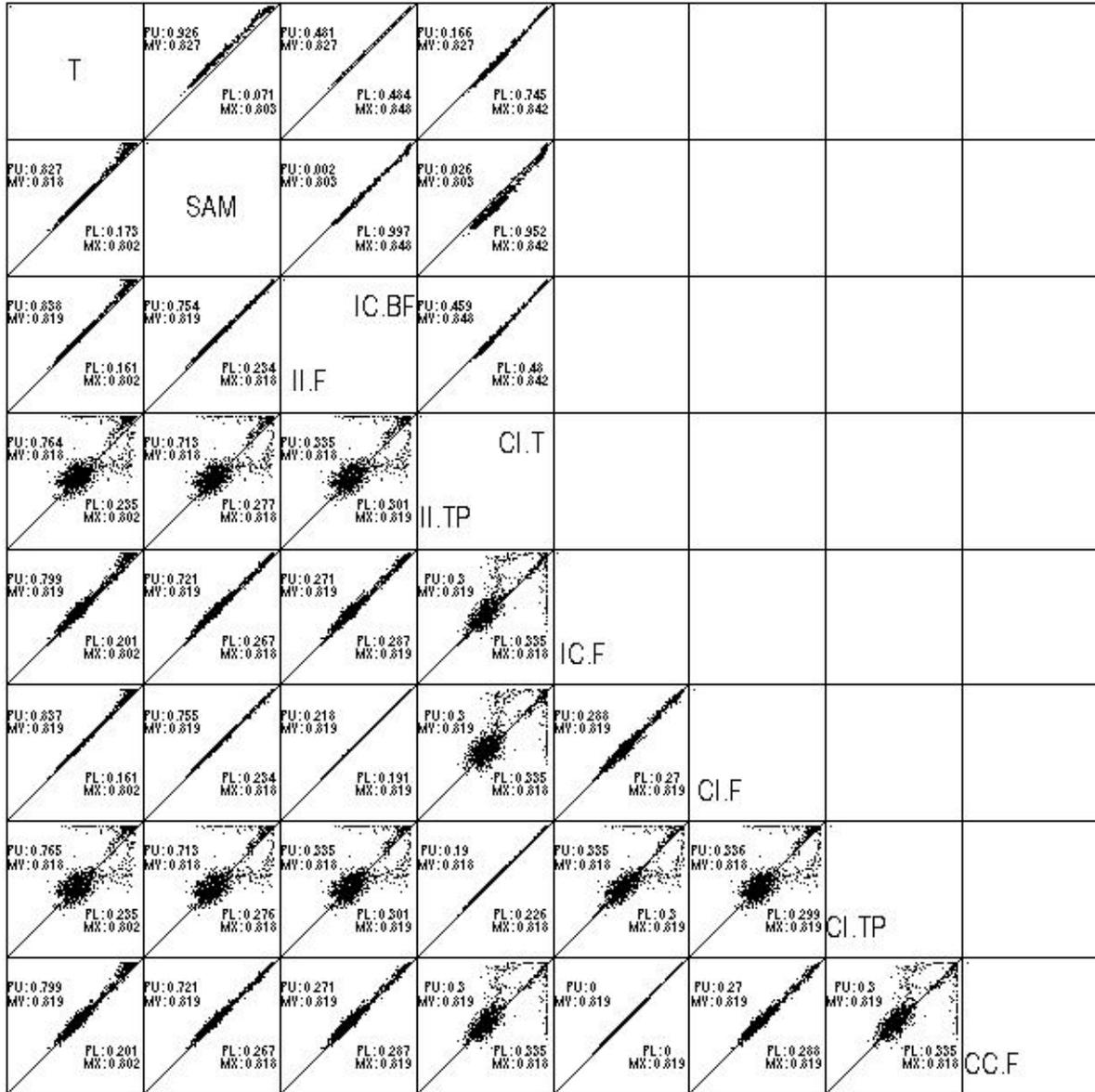


Figure 116: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

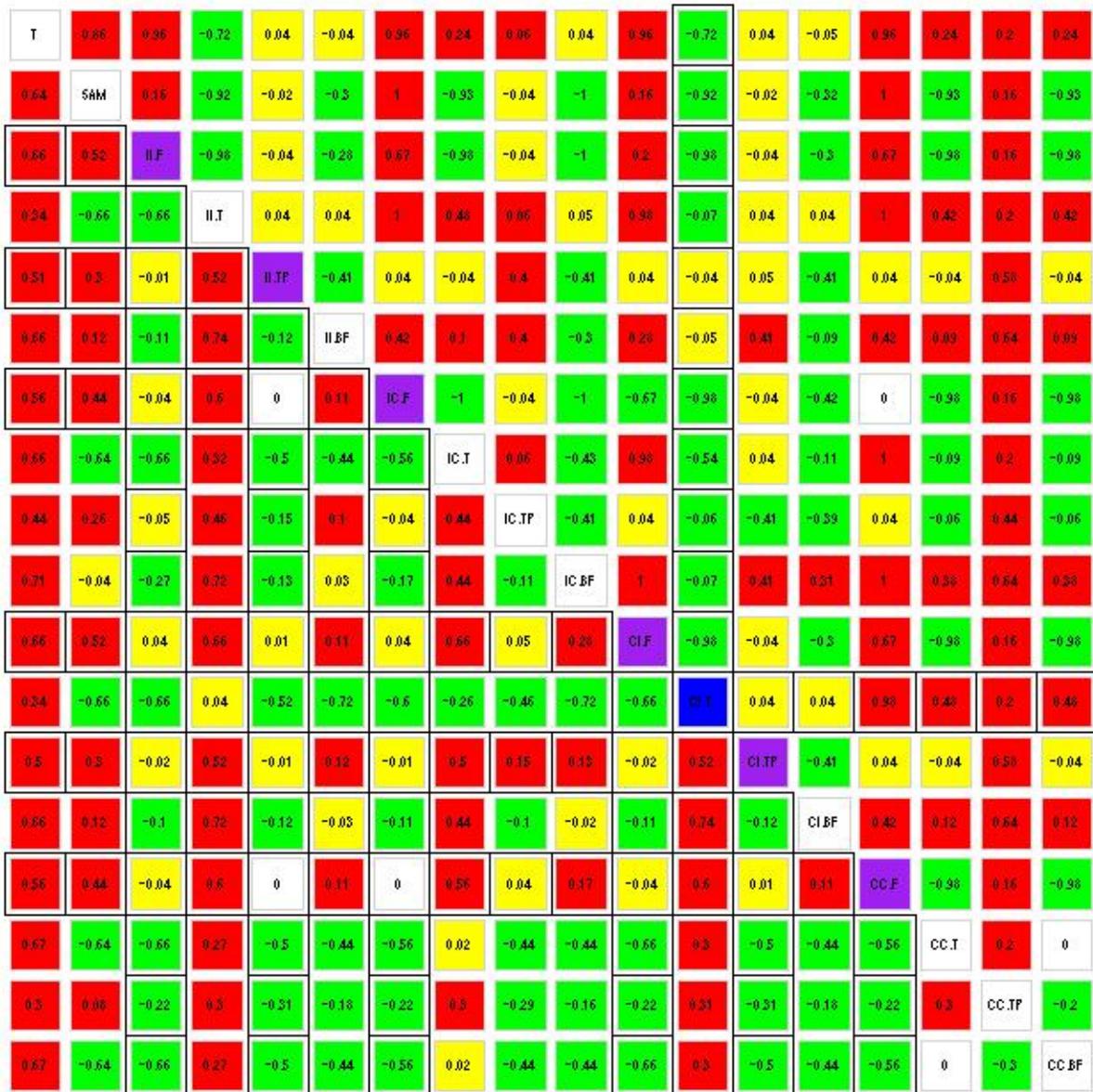


Figure 117: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 100 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

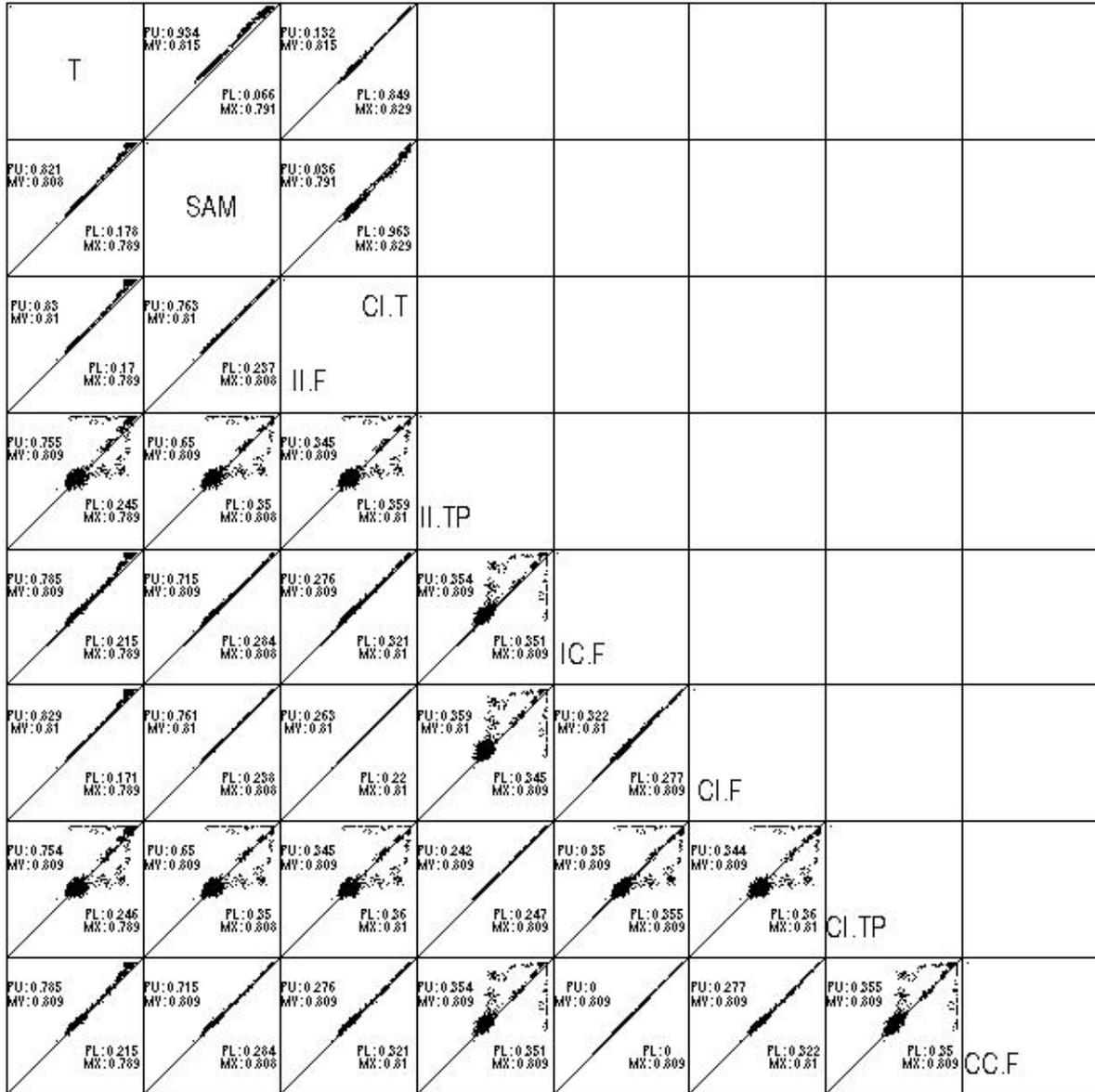


Figure 118: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

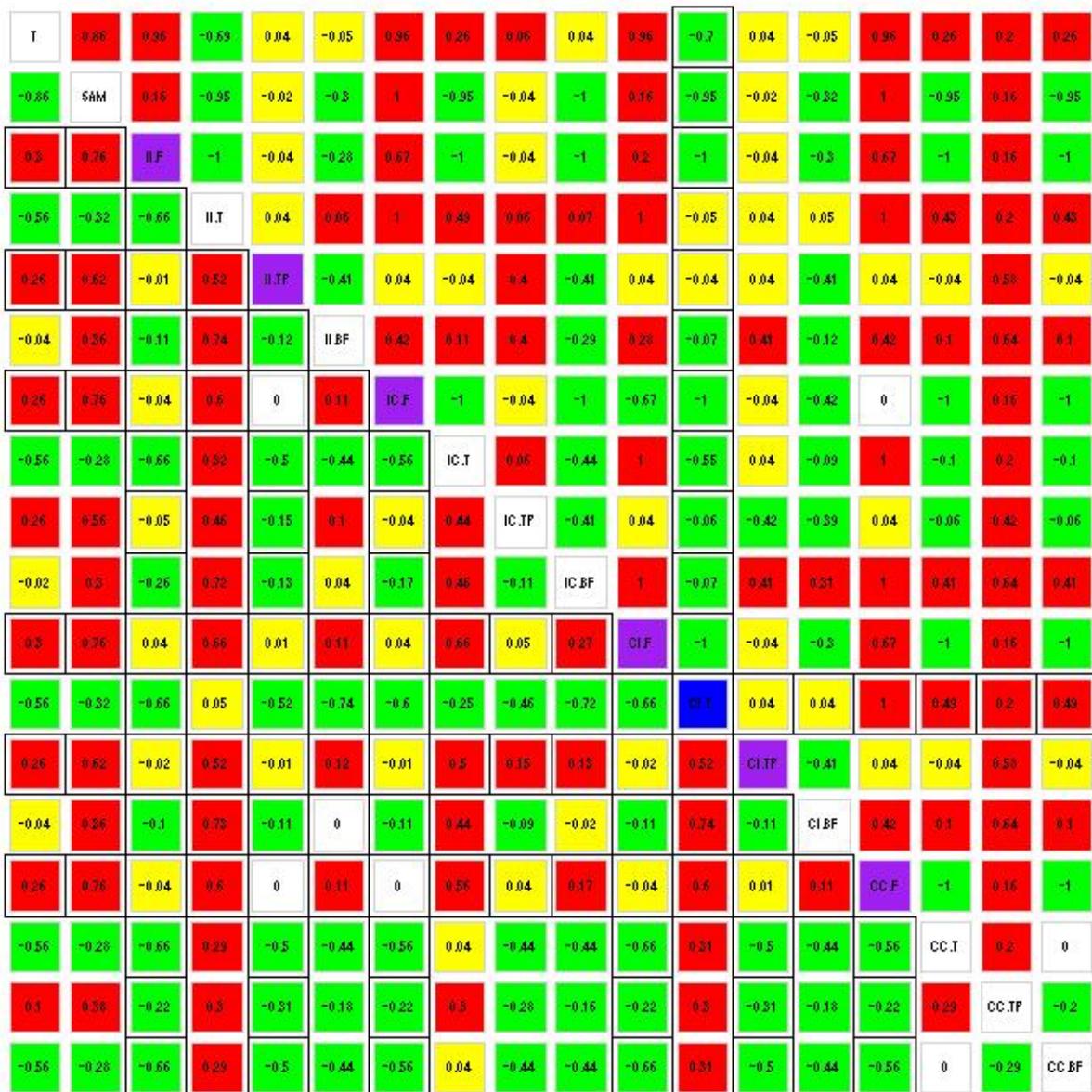


Figure 119: Heat map summary for simulations based on Independence of Signal and Noise model with number of replicates equal to 100 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

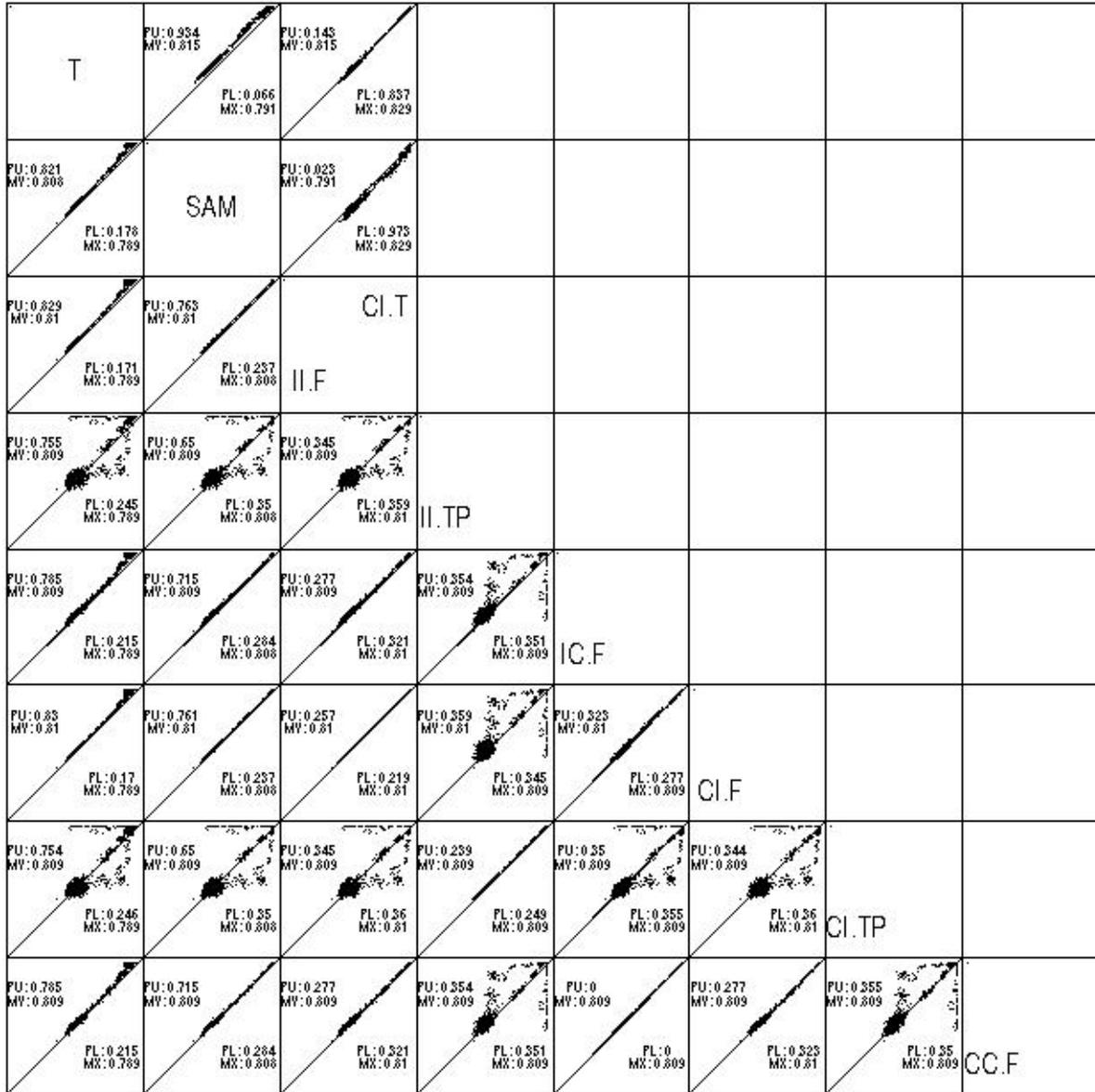


Figure 120: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Signal and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

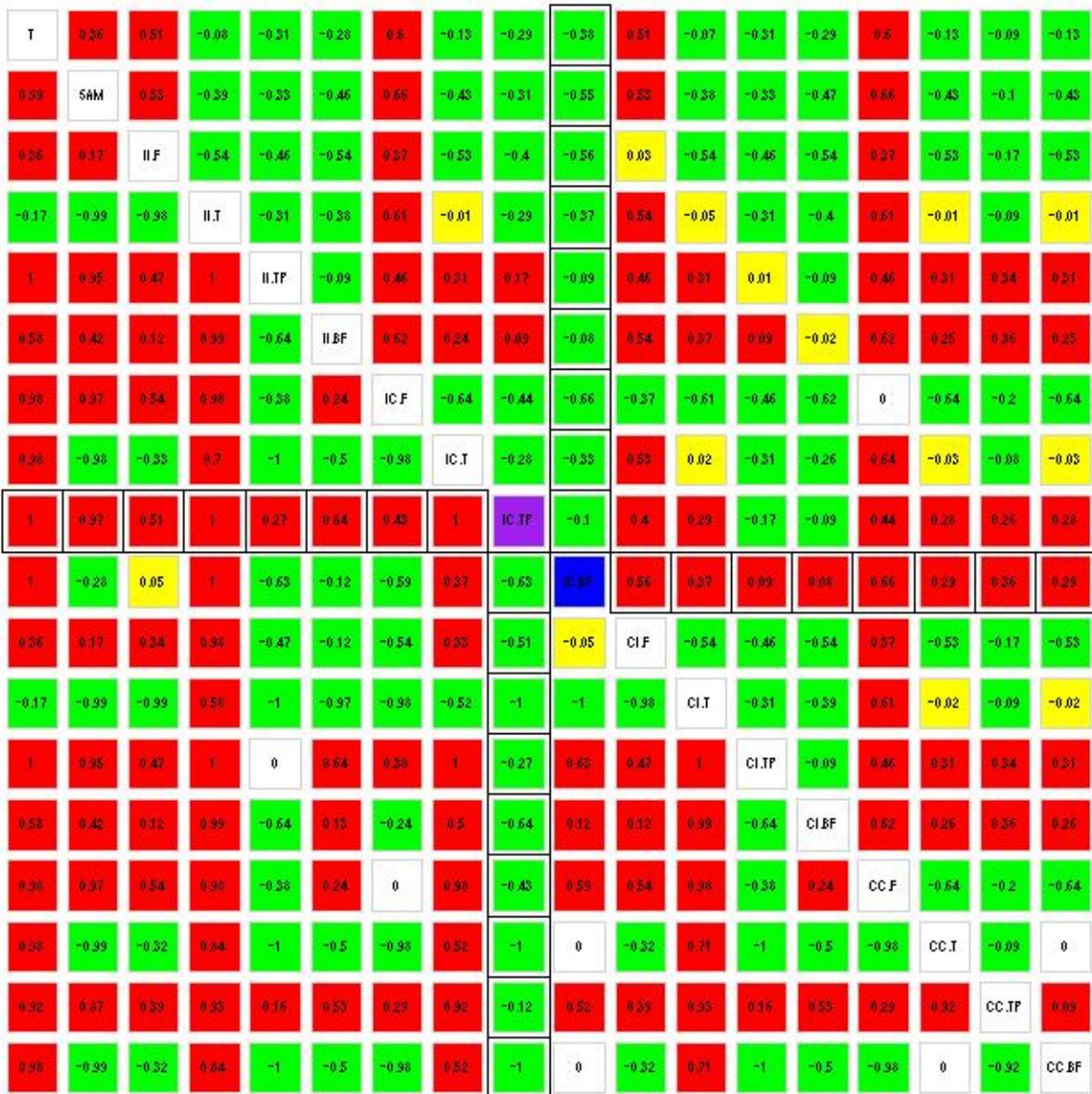


Figure 121: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 100 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

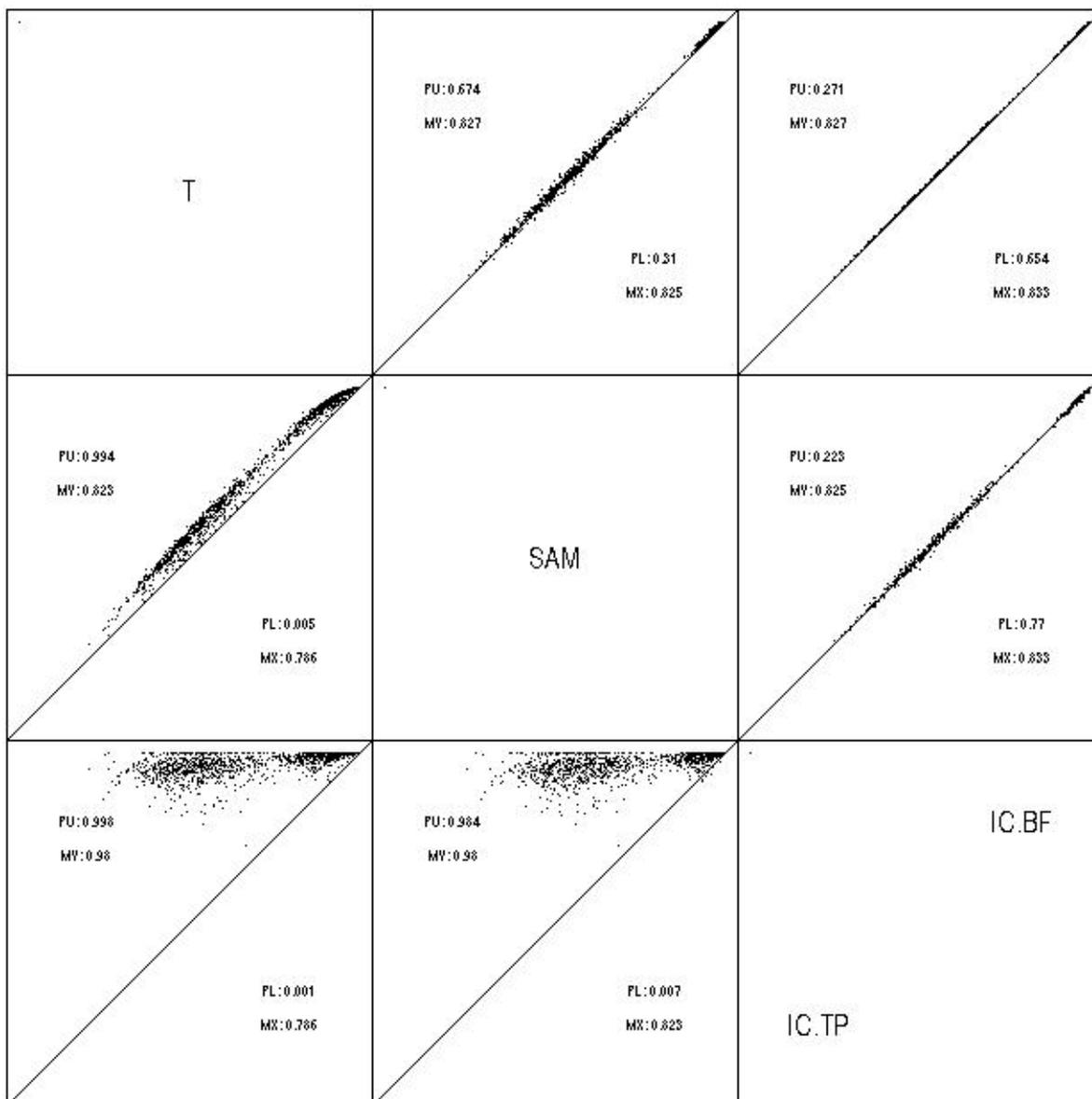


Figure 122: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

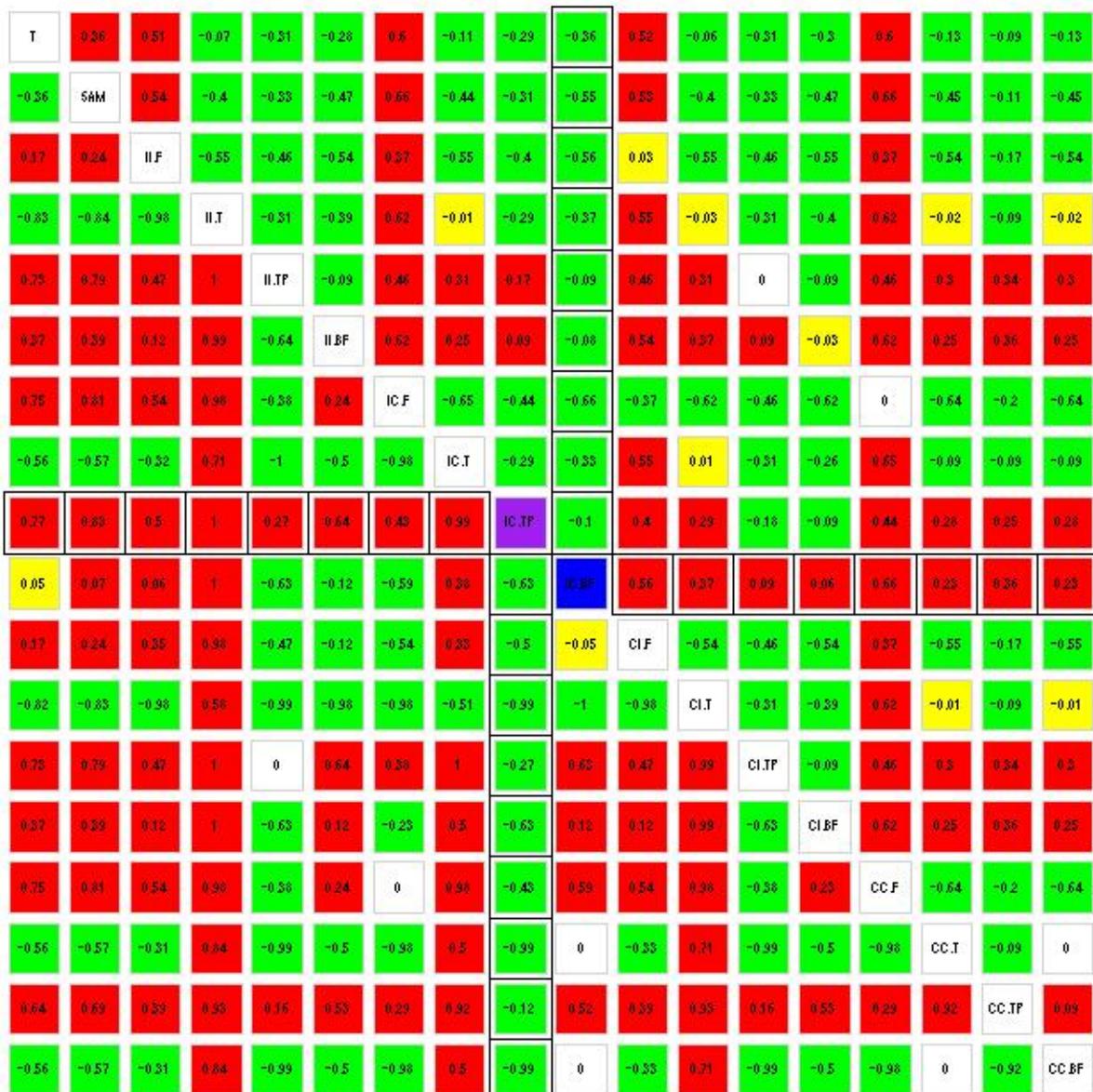


Figure 123: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 100 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

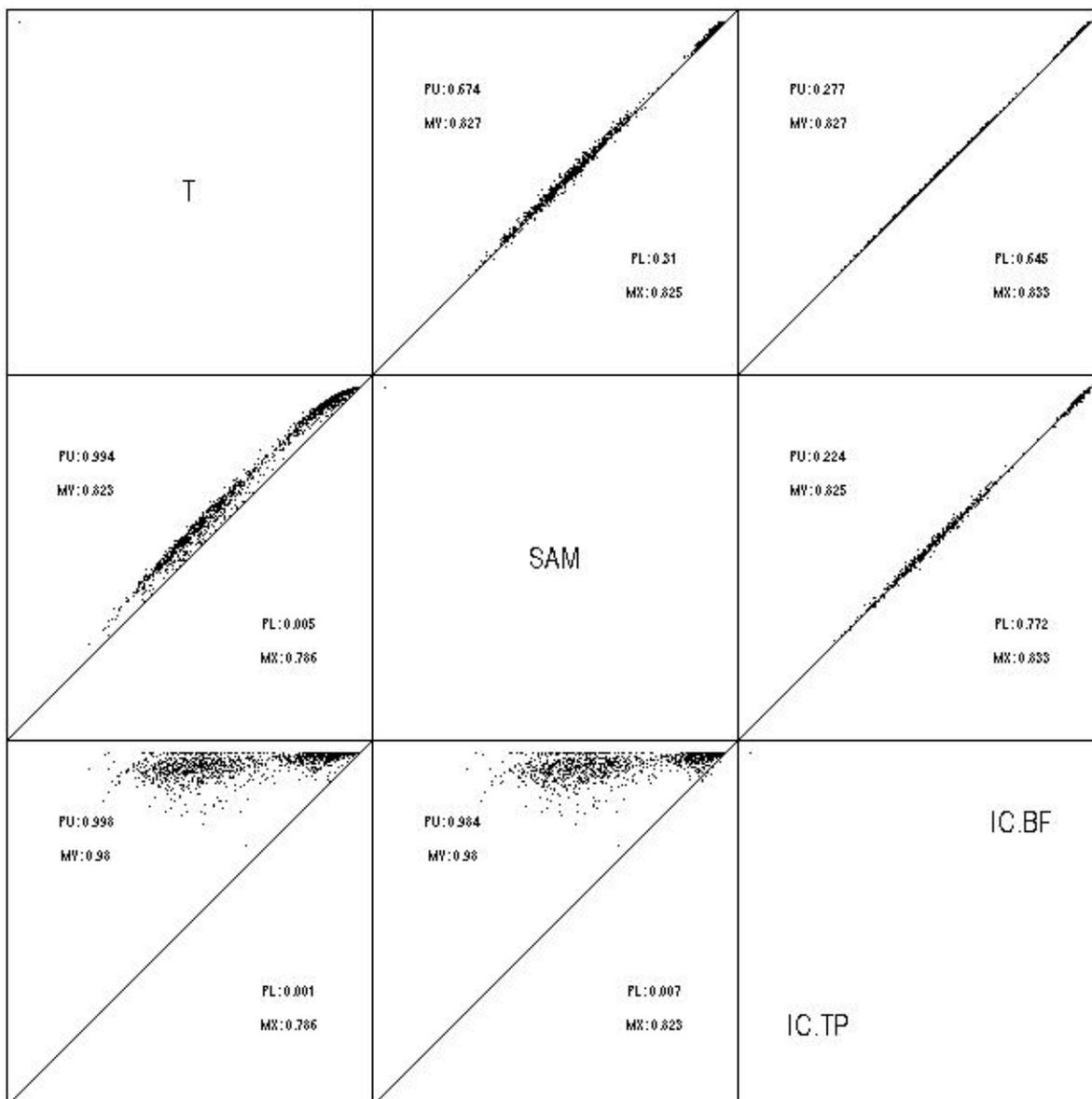


Figure 124: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

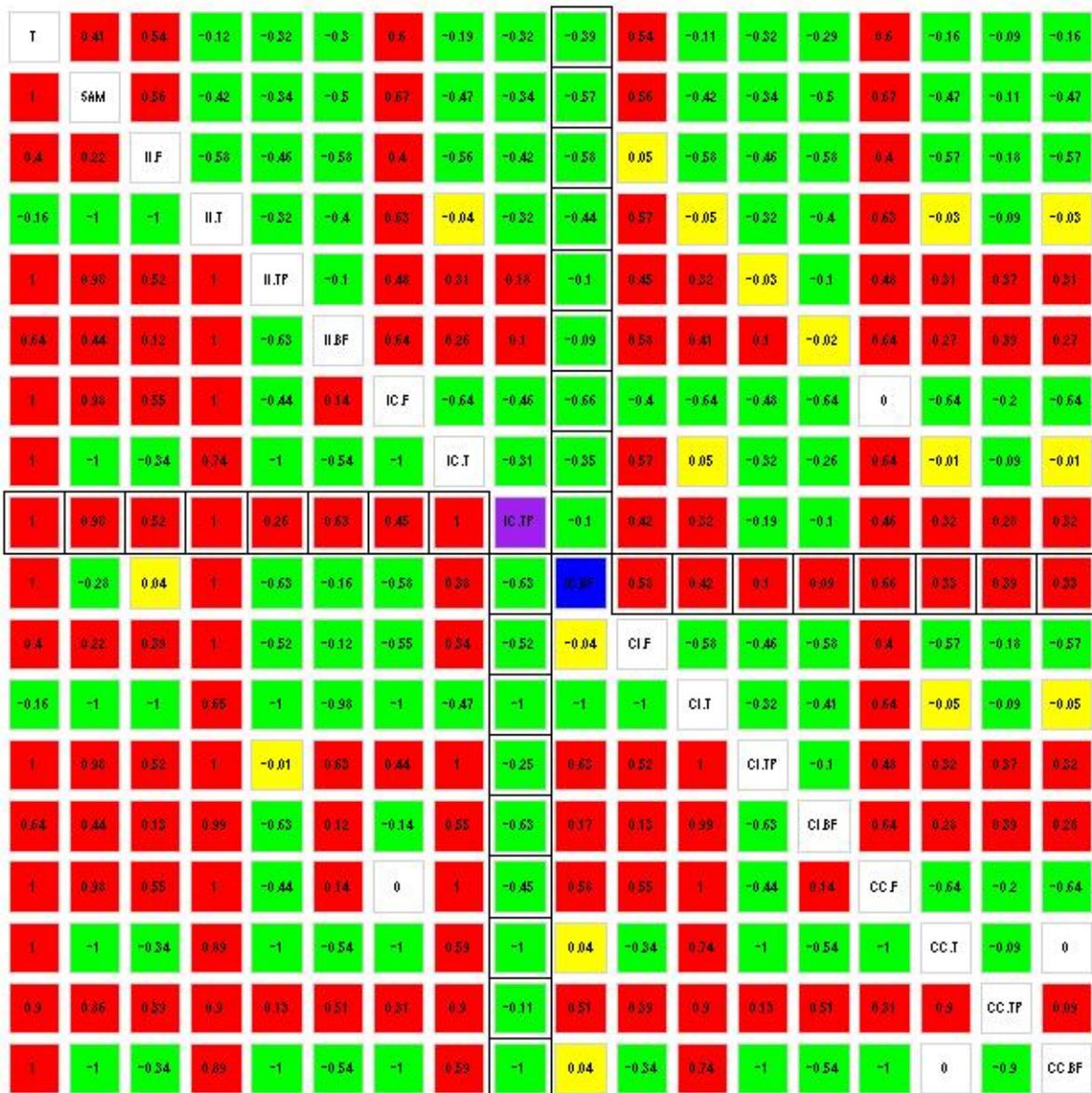


Figure 125: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 100 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

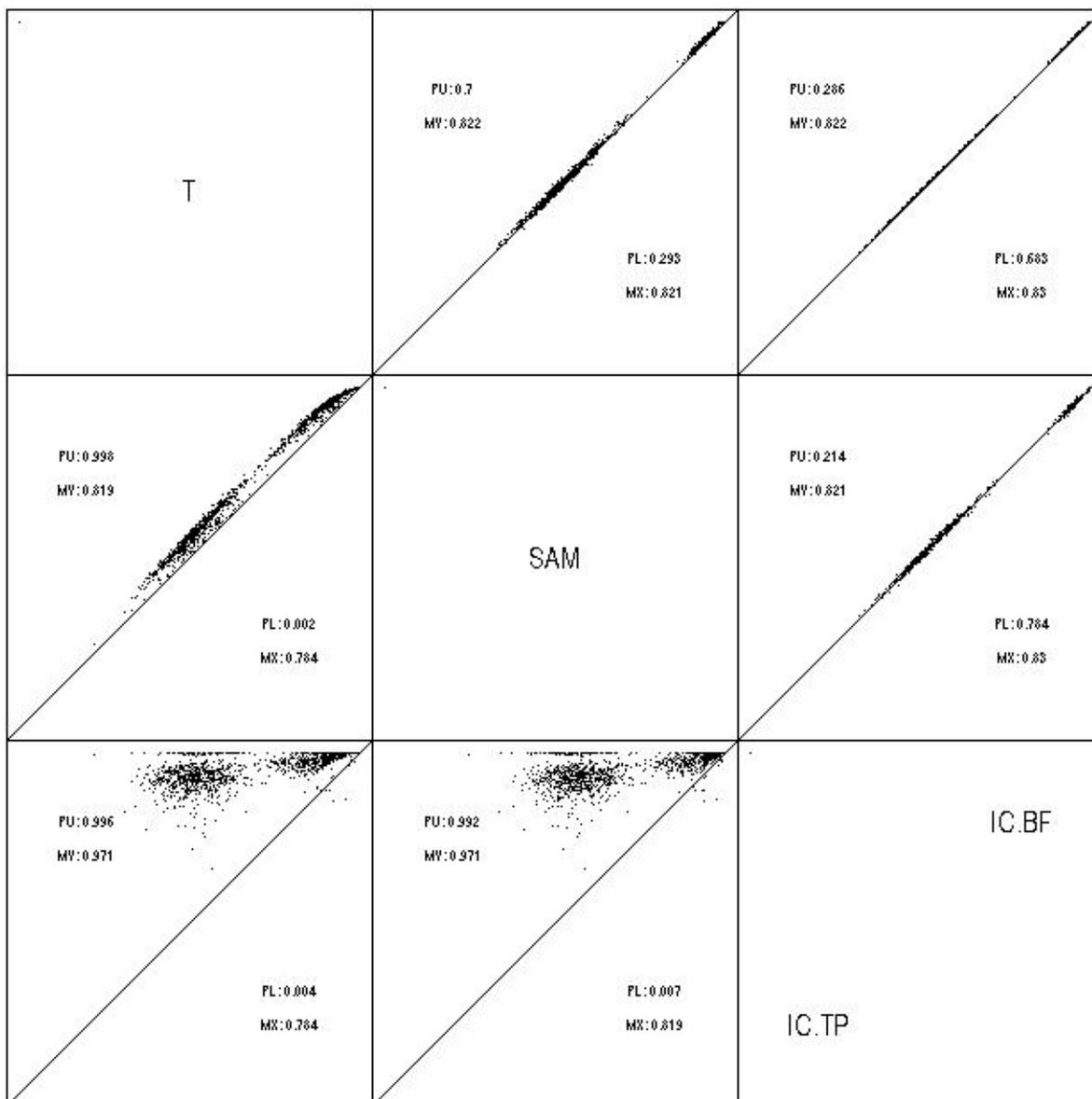


Figure 126: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

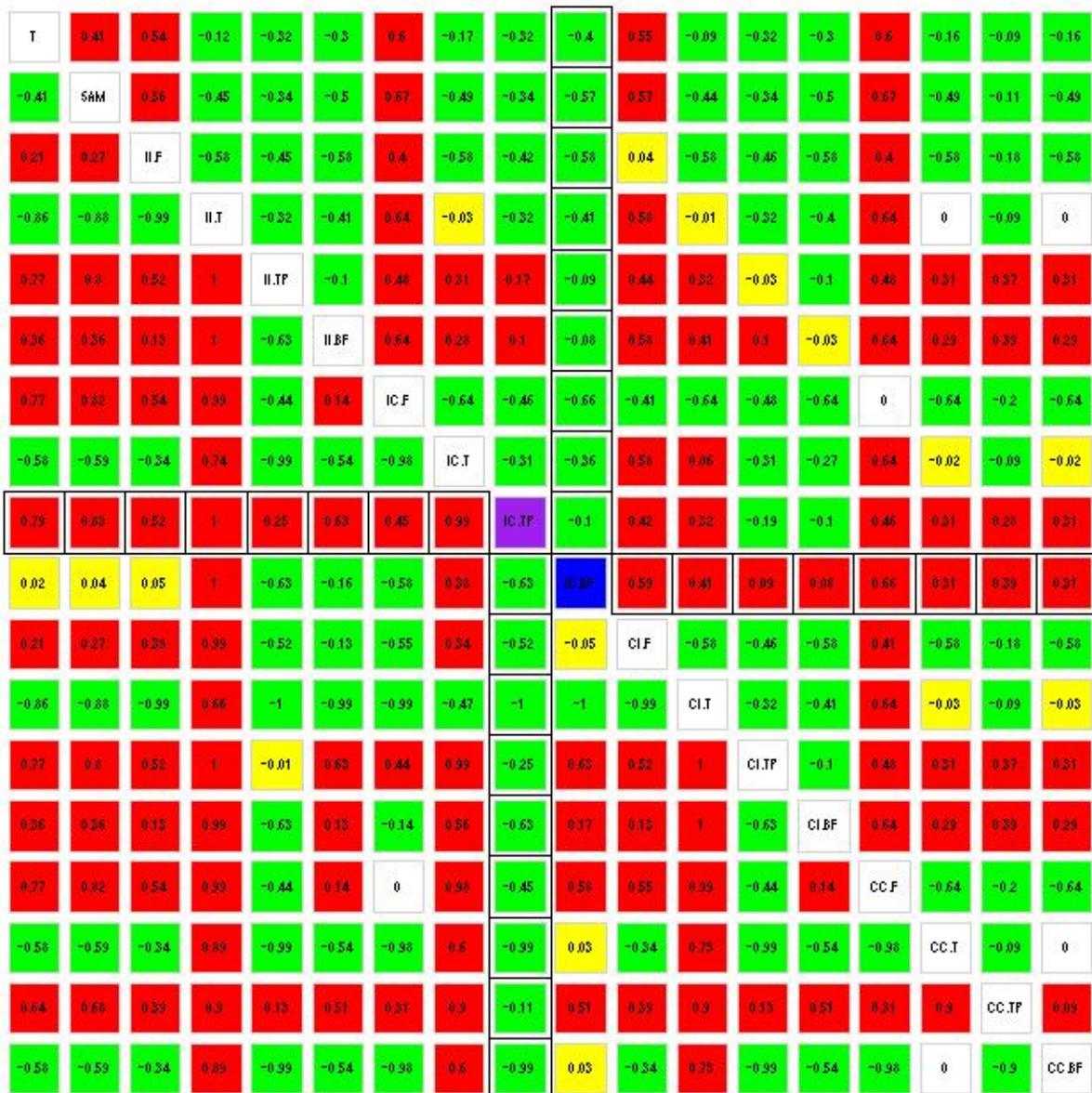


Figure 127: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 100 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

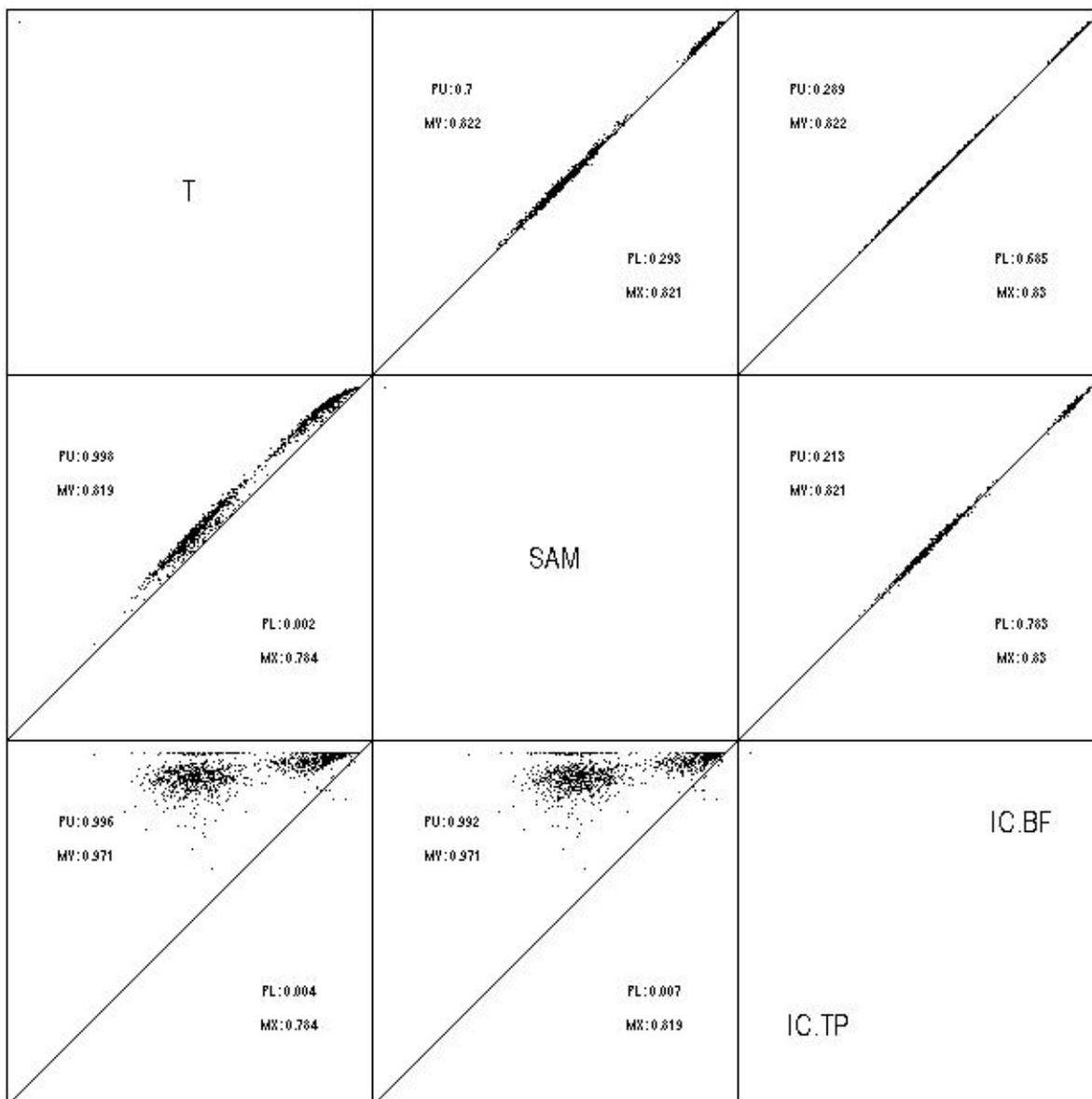


Figure 128: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

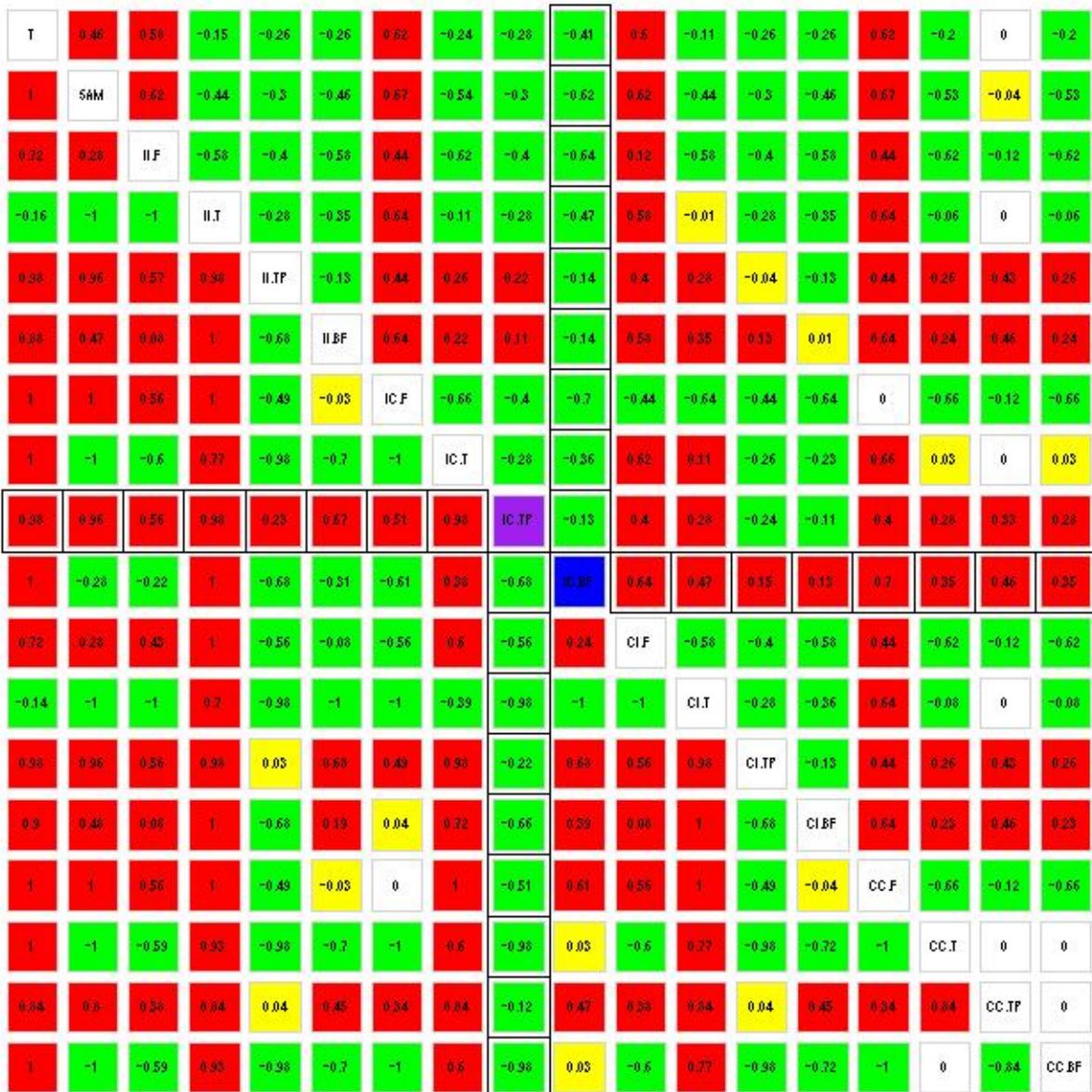


Figure 129: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 100 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

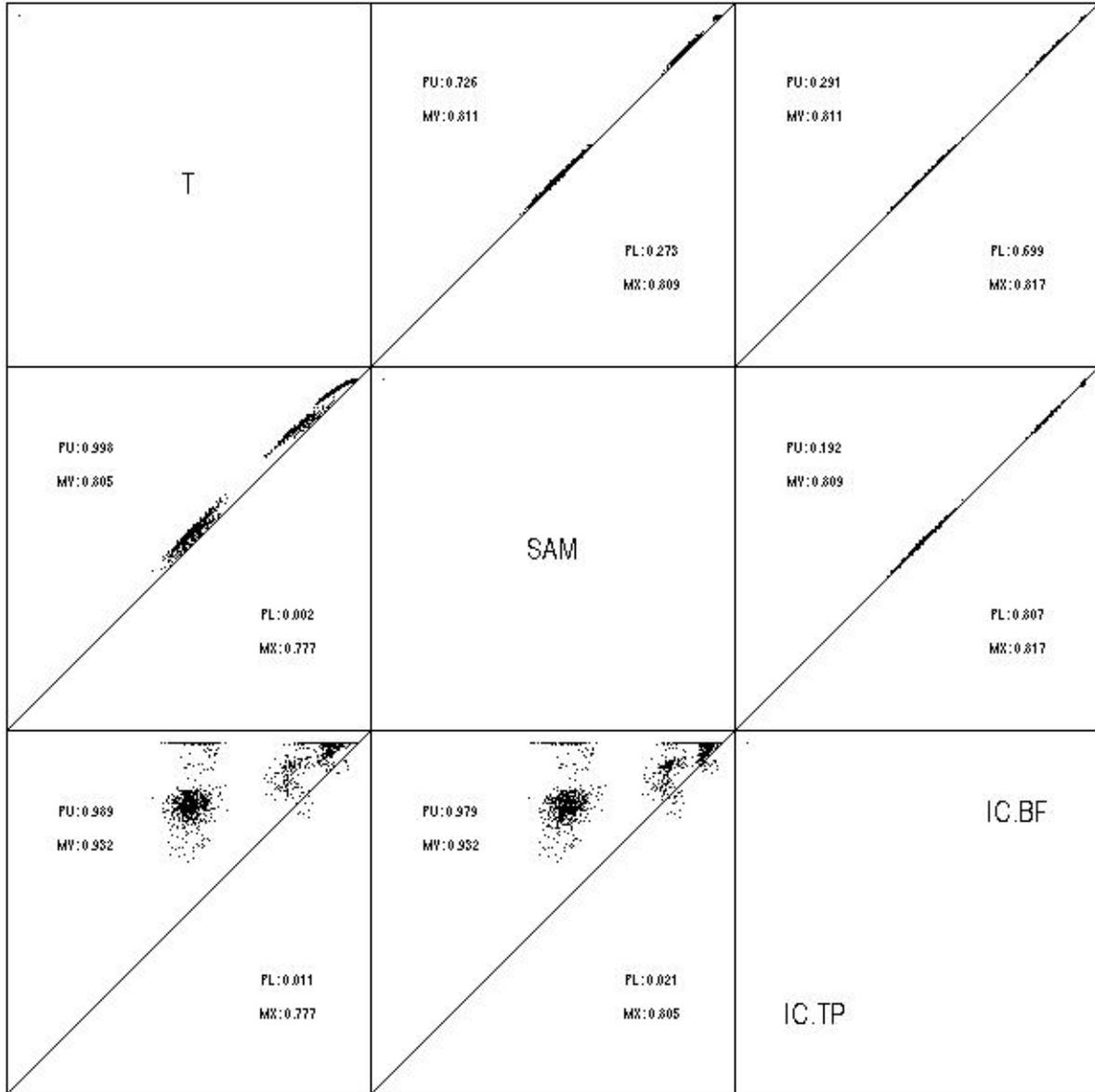


Figure 130: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

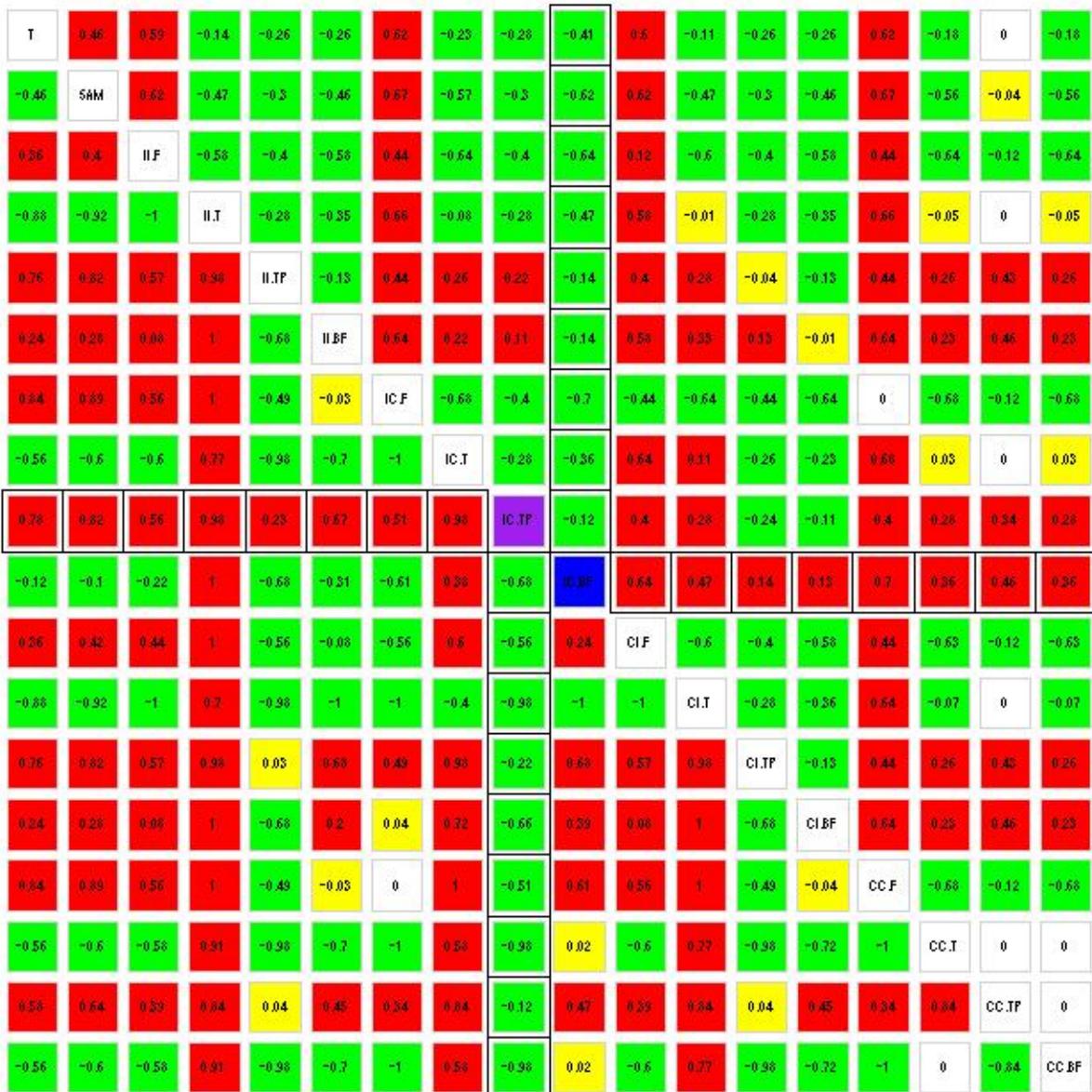


Figure 131: Heat map summary for simulations based on Independence of Abundance and Noise model with number of replicates equal to 100 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

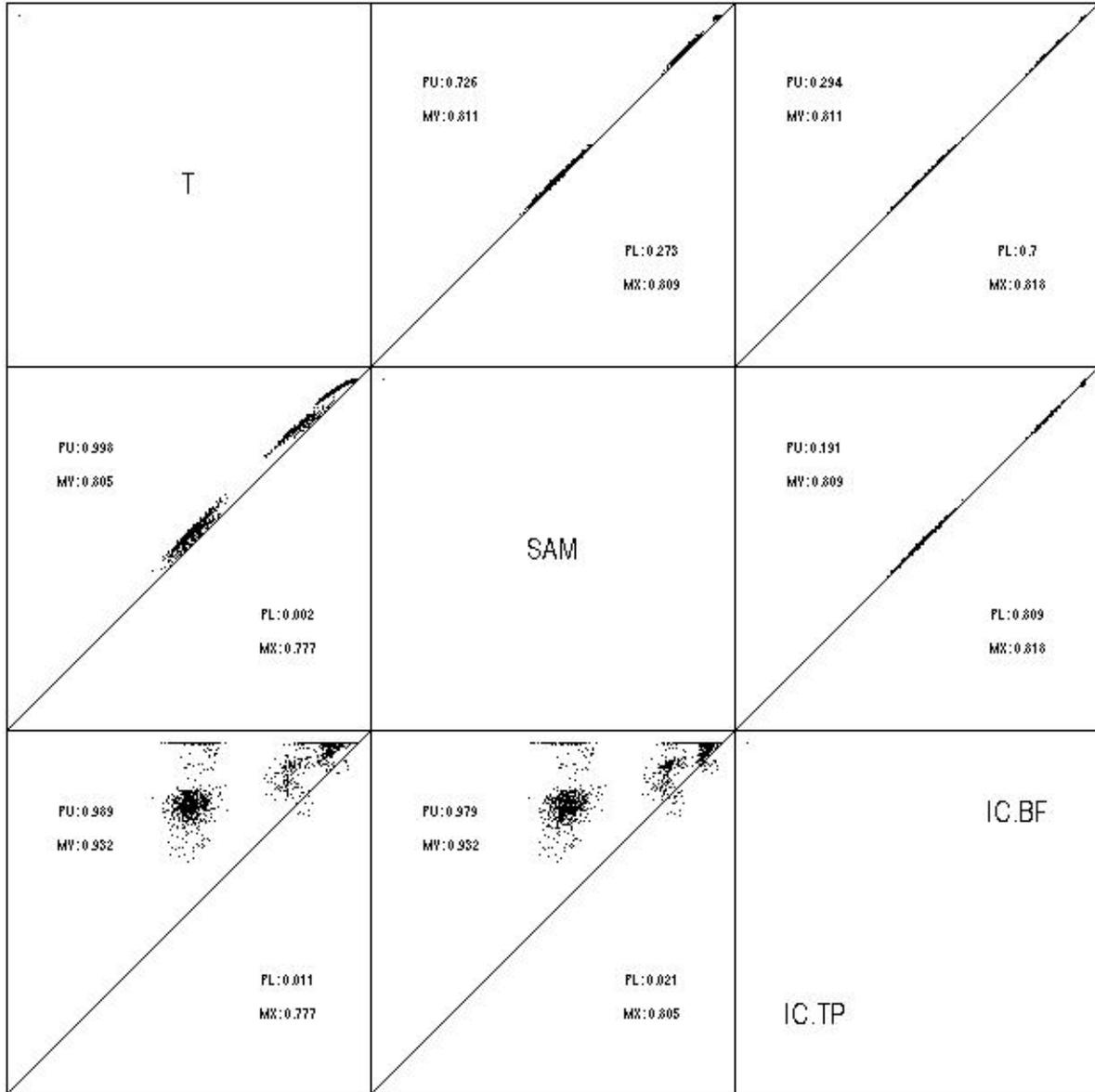


Figure 132: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Independence of Abundance and Noise model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

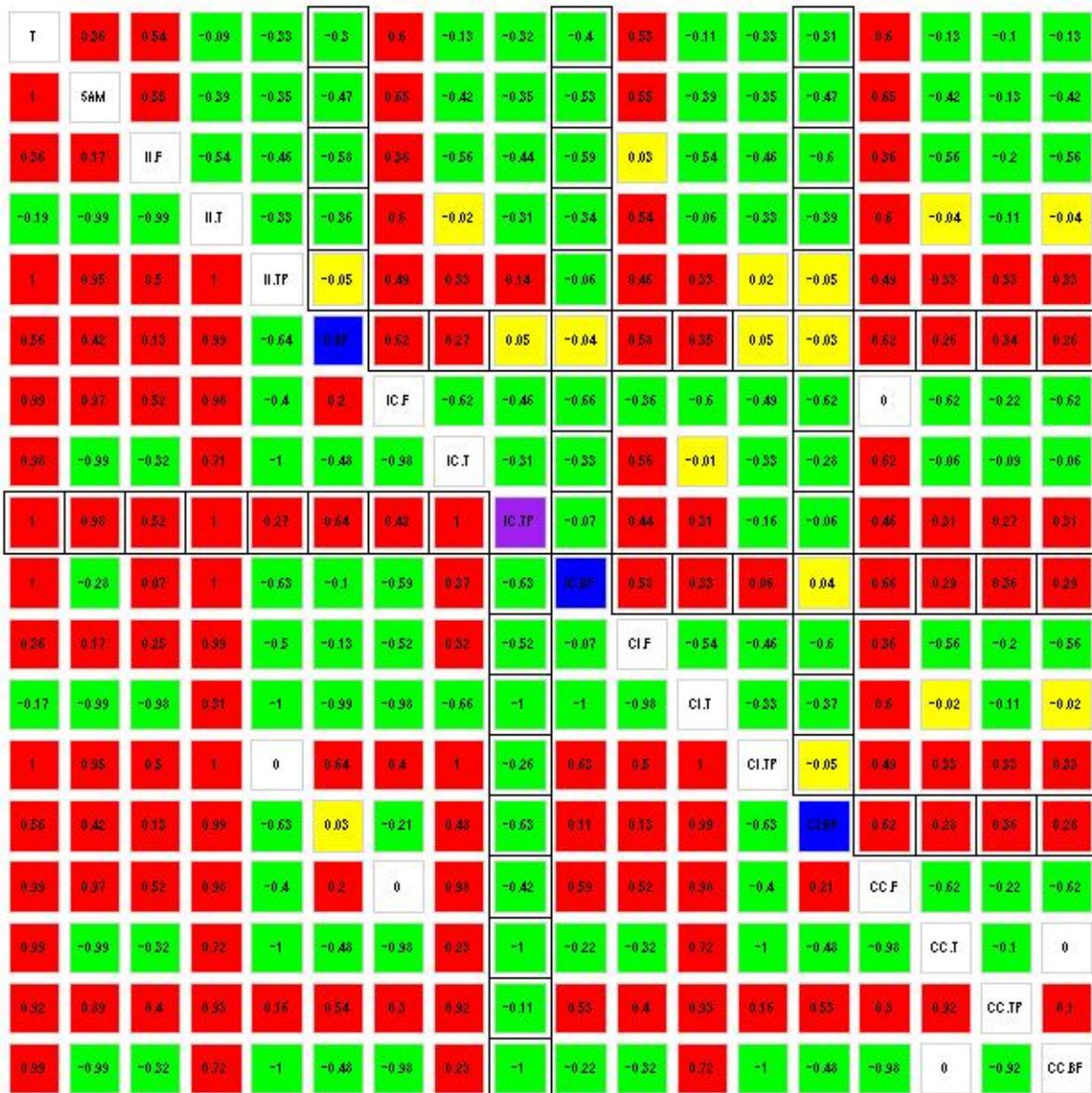


Figure 133: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 100 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

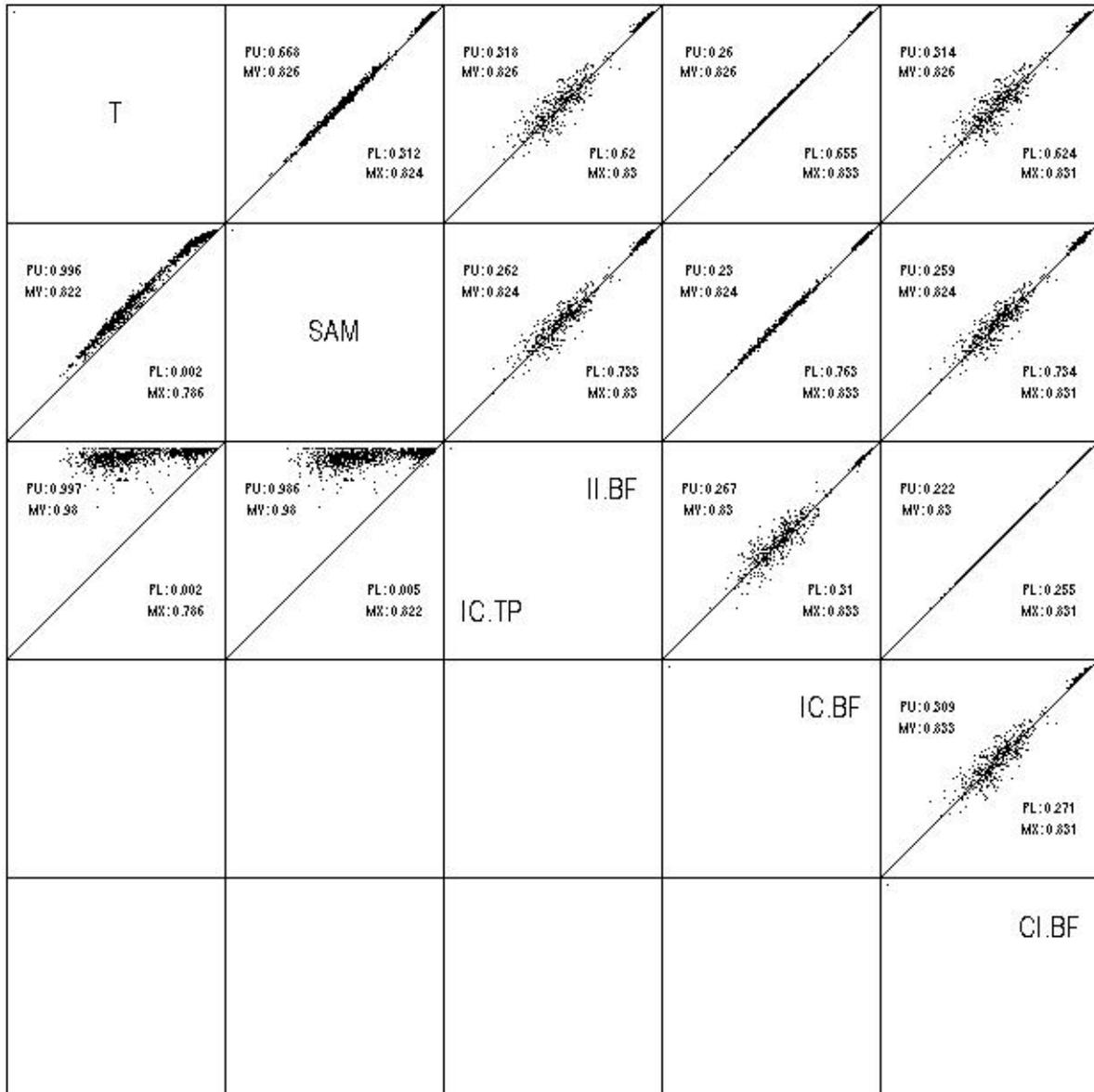


Figure 134: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 1% genes to be differential genes. Hyperparameters are estimated by method of moments.

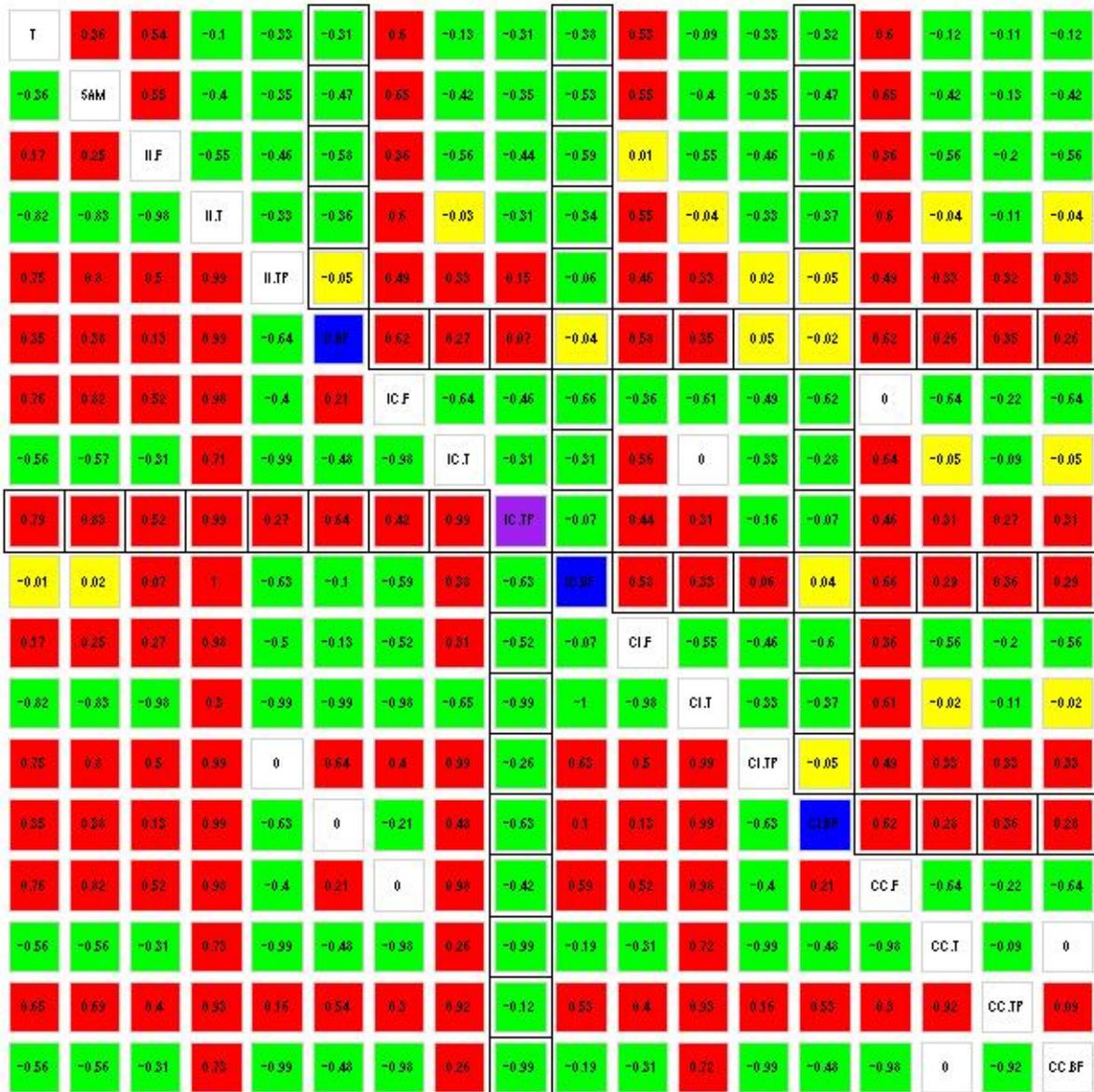


Figure 135: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 100 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

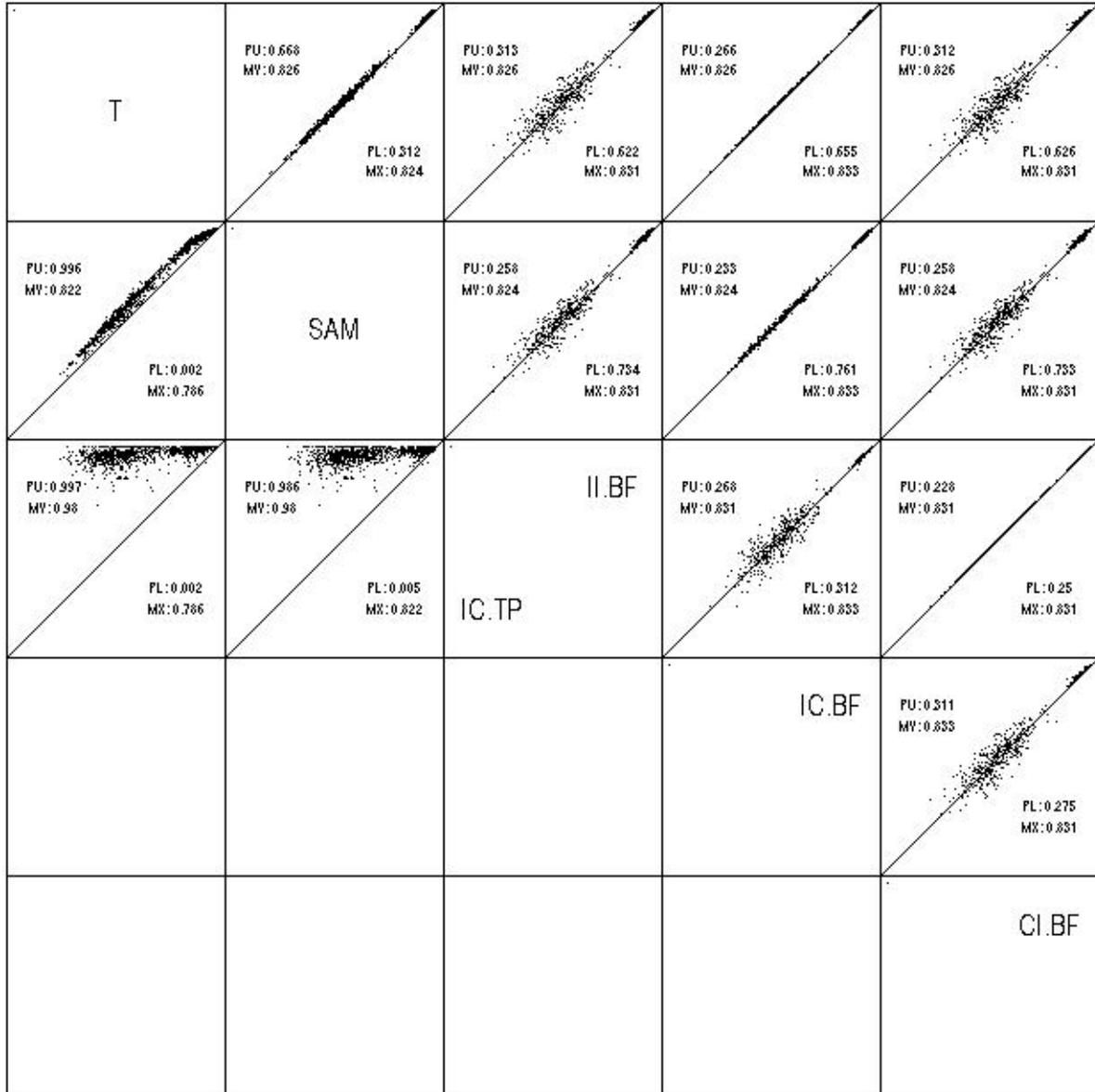


Figure 136: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 1% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

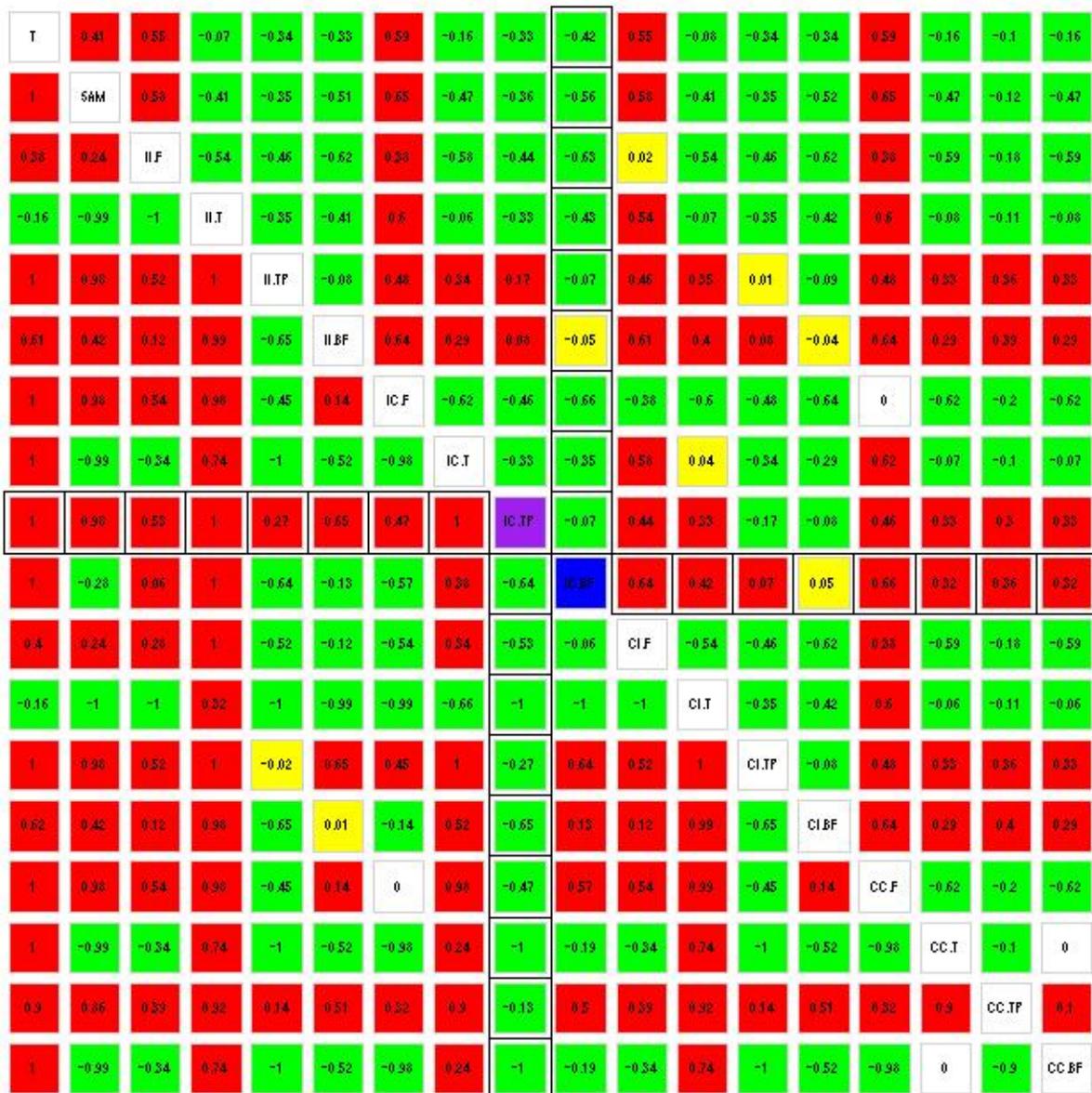


Figure 137: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 100 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

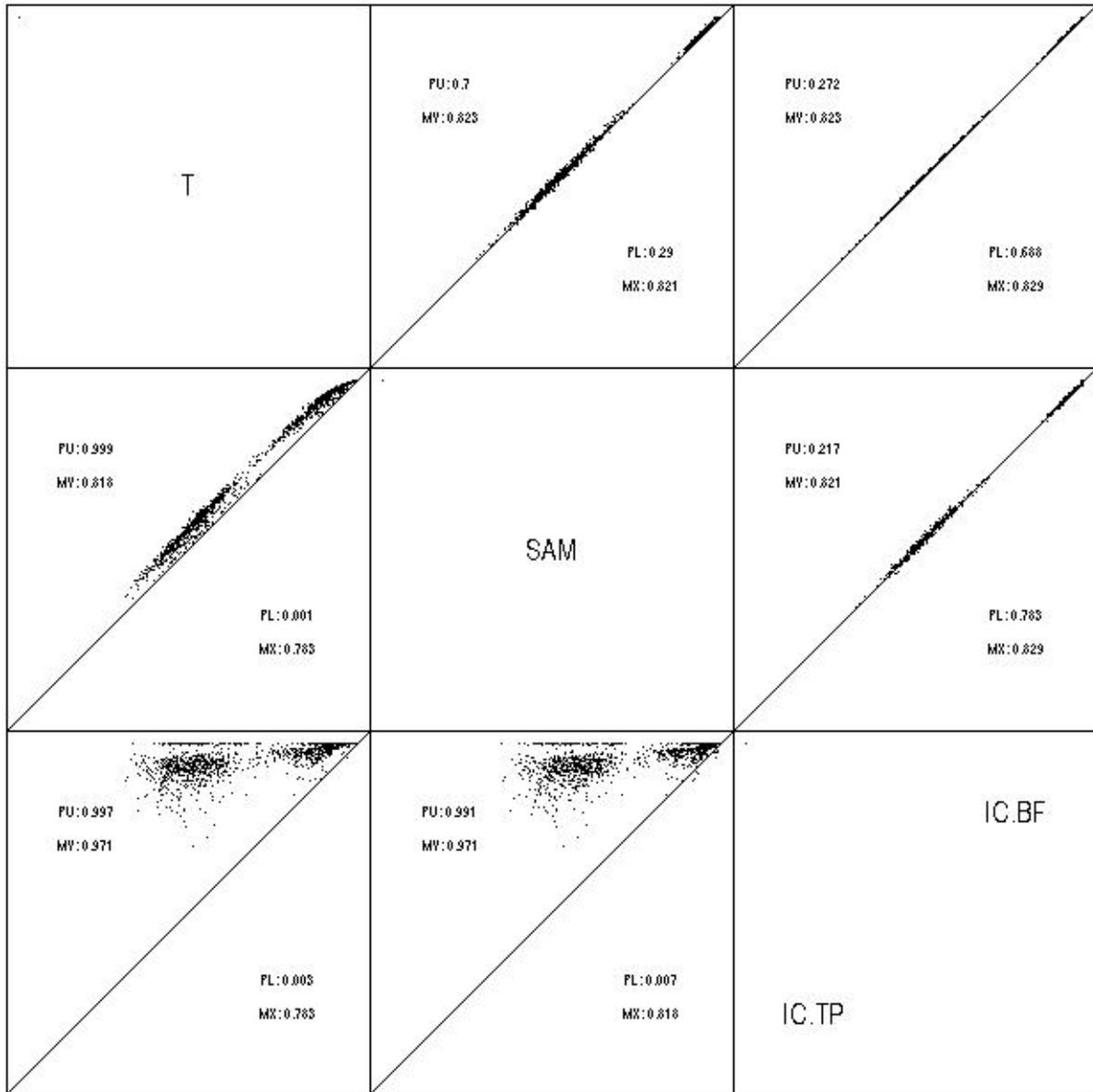


Figure 138: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 2% genes to be differential genes. Hyperparameters are estimated by method of moments.

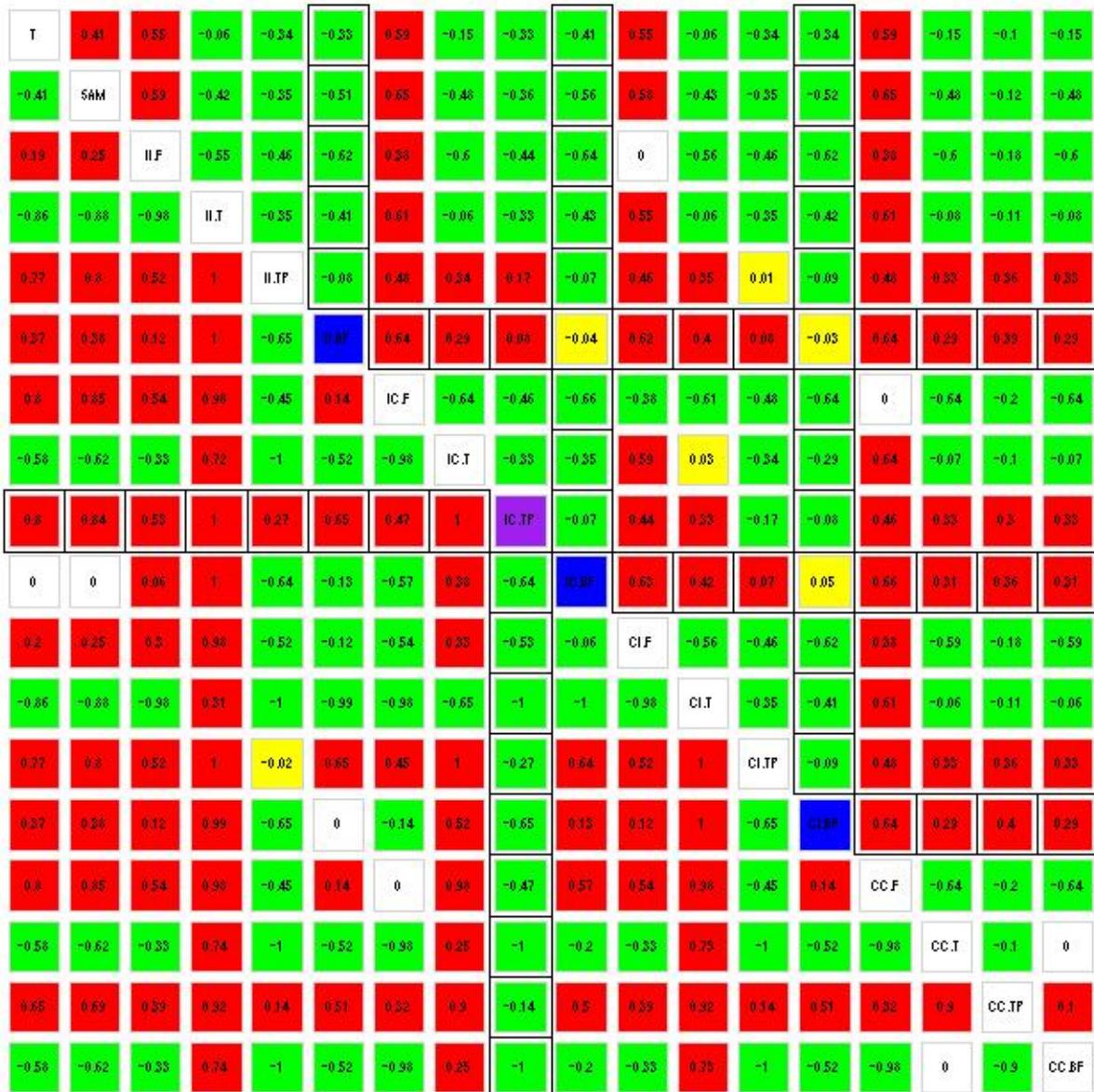


Figure 139: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 100 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

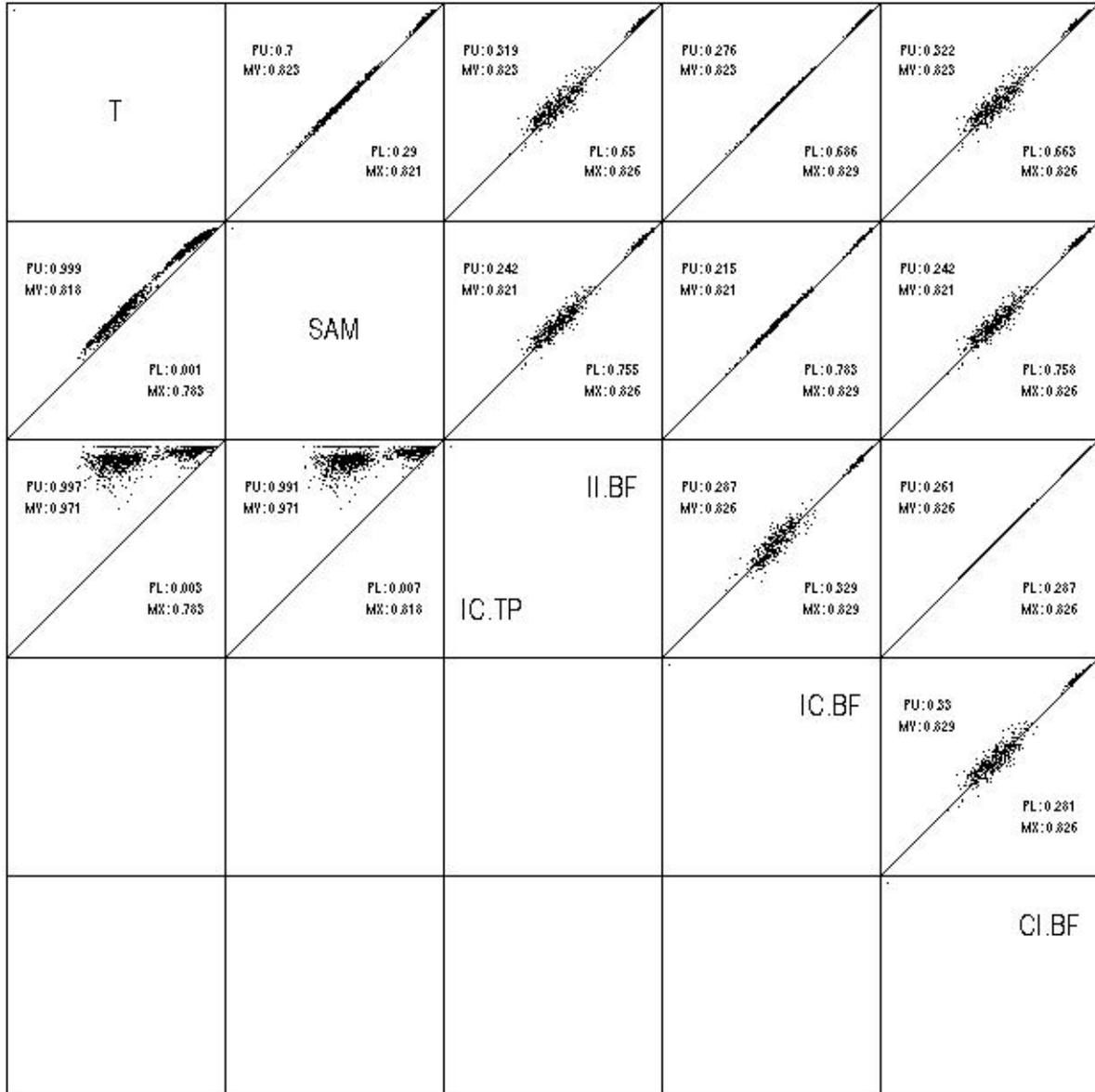


Figure 140: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 2% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

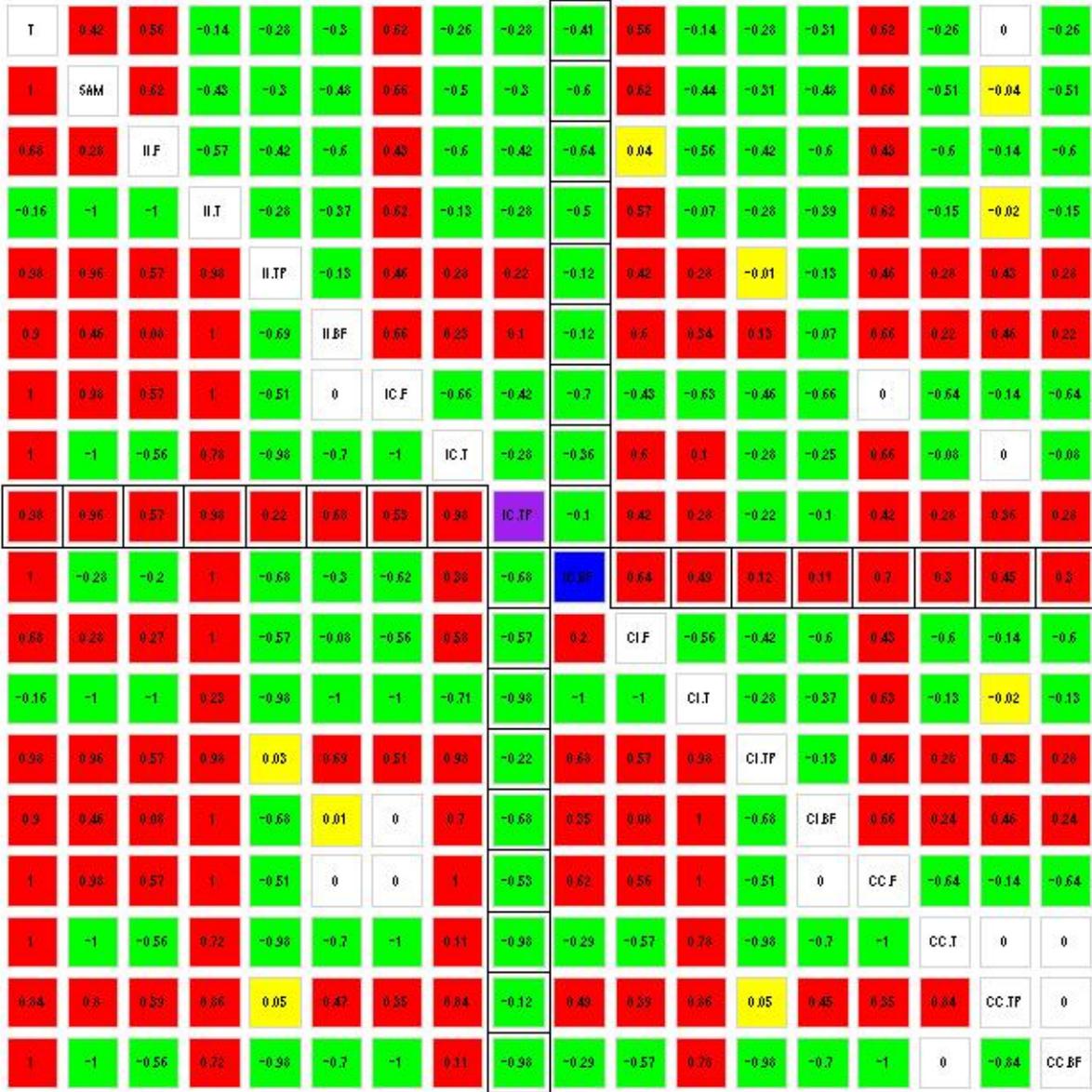


Figure 141: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 100 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

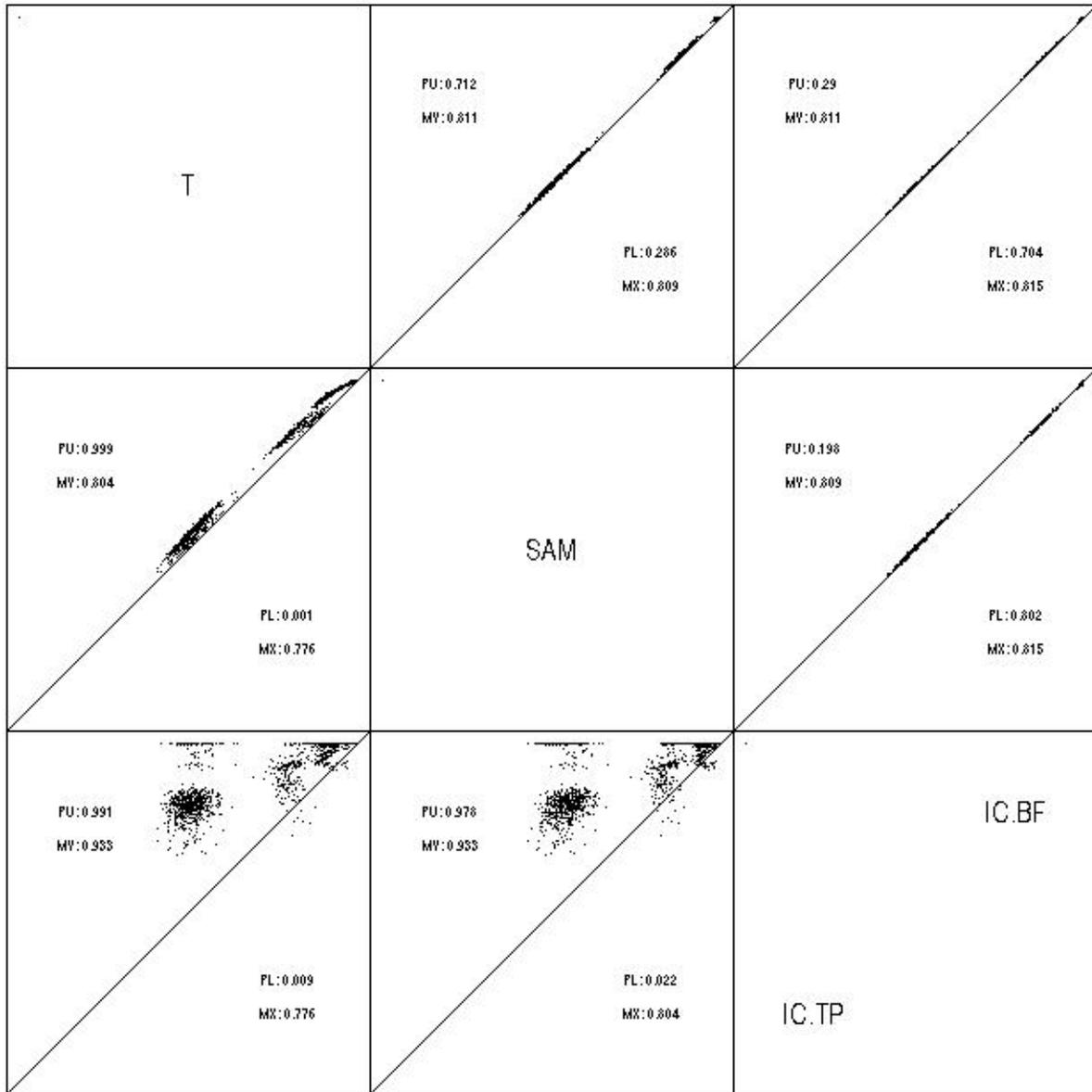


Figure 142: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 10% genes to be differential genes. Hyperparameters are estimated by method of moments.

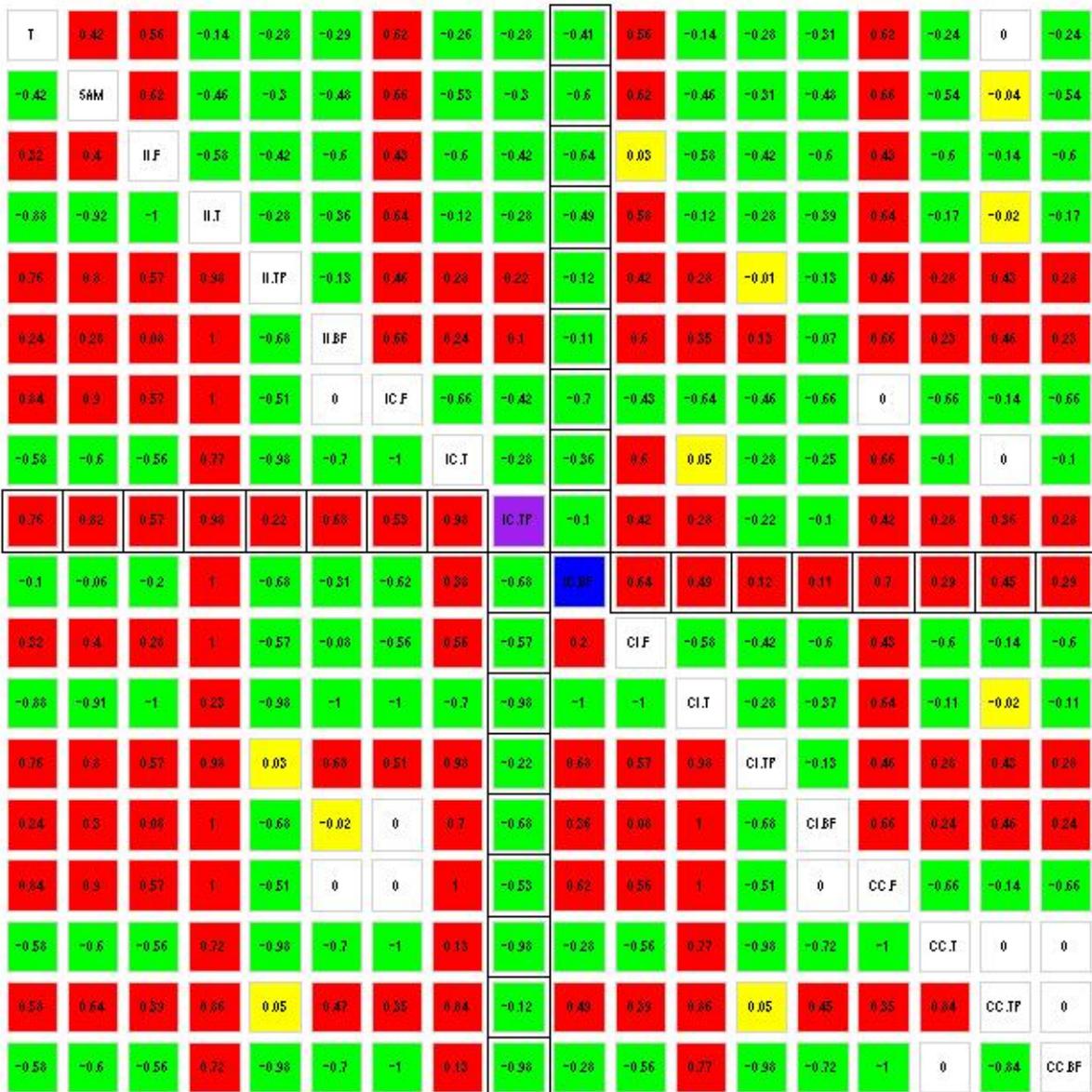


Figure 143: Heat map summary for simulations based on Complete Conjugacy model with number of replicates equal to 100 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

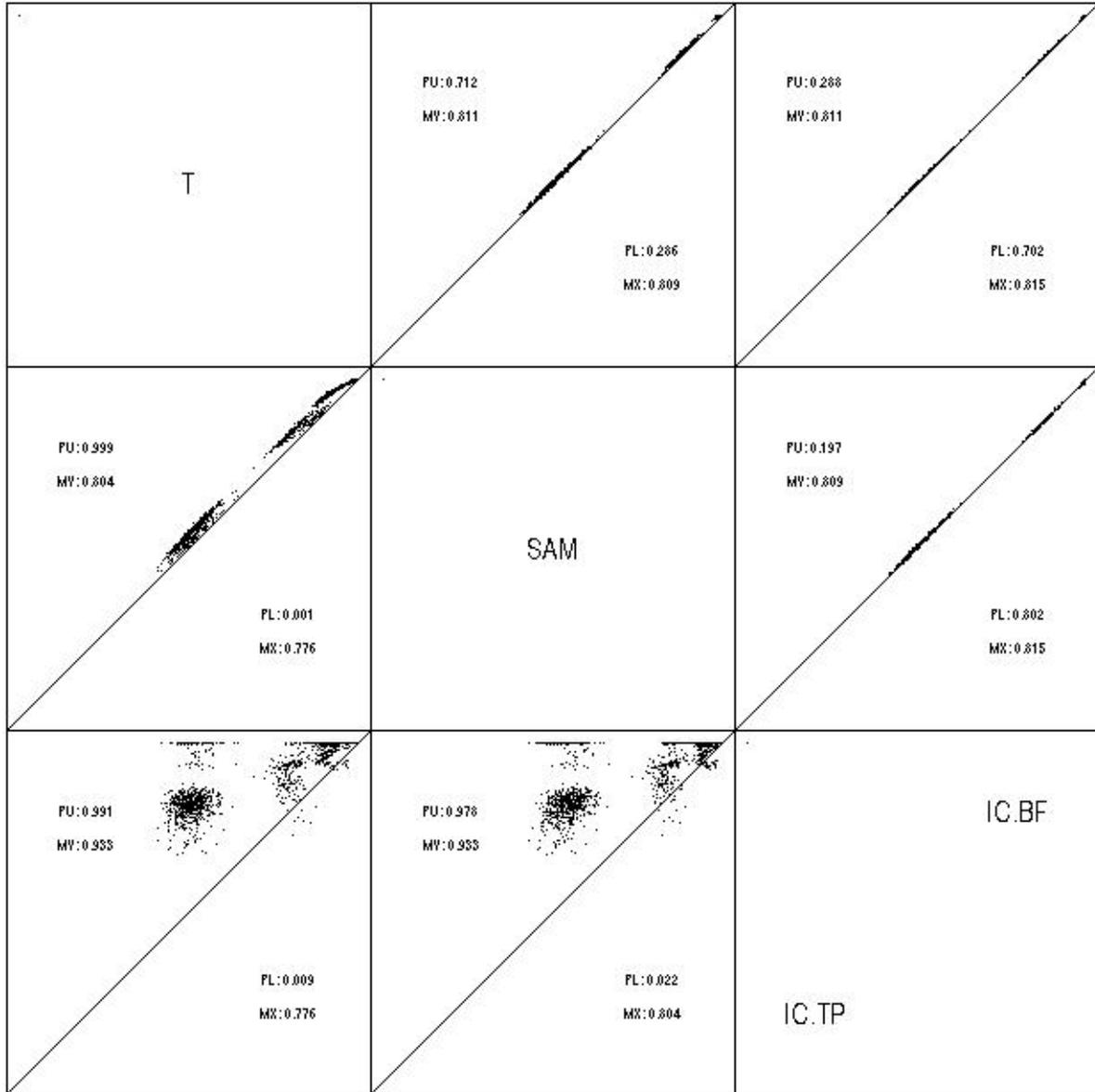


Figure 144: The best statistics among all 18 statistics are selected and compared to the simple T statistics and SAM score. This is the result of simulation based on Complete Conjugacy model with number of replicates equal to 10 and top 10% genes to be differential genes. Statistics based on multilevel models are calculated with true hyperparameters plugged in.

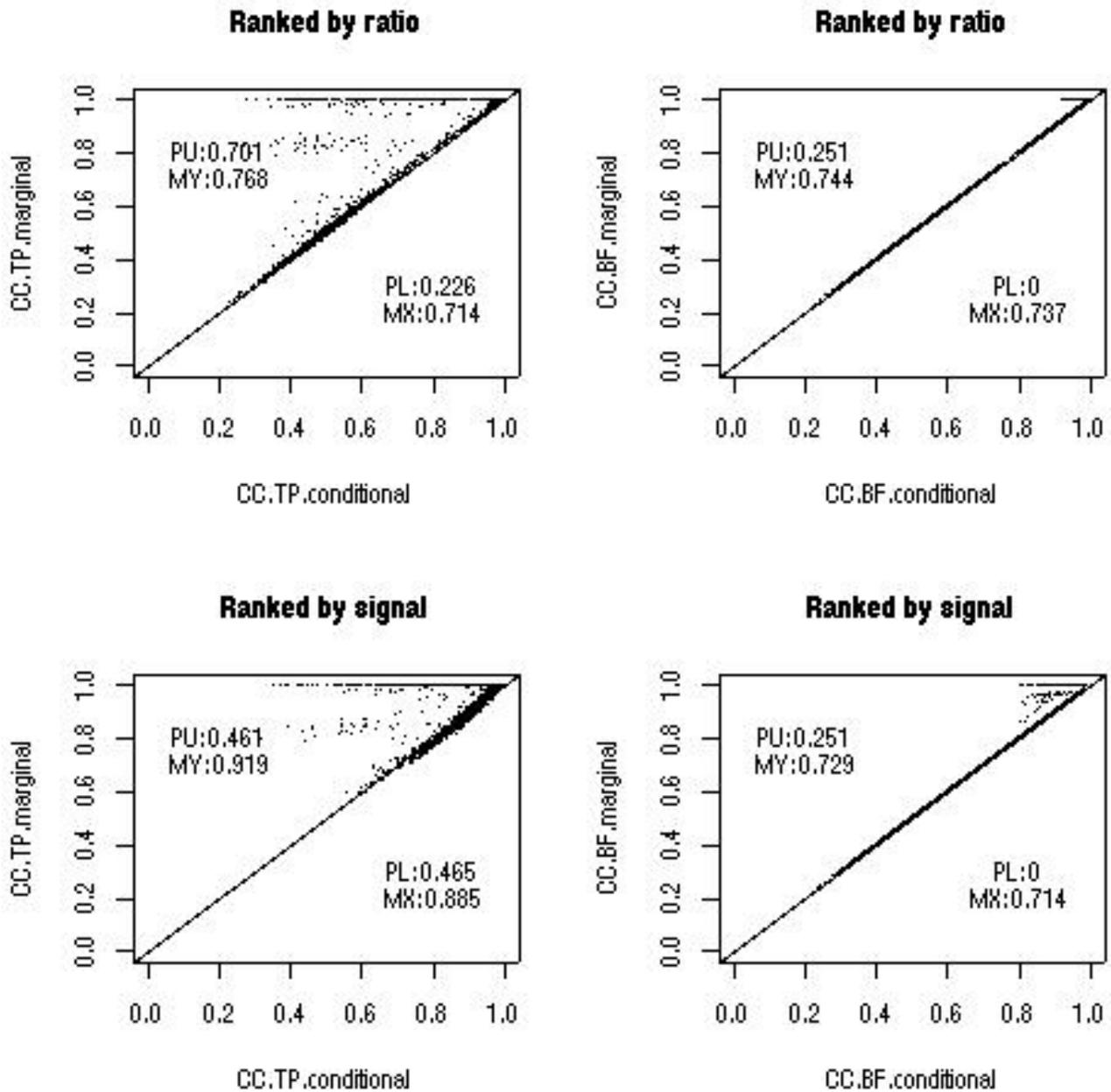


Figure 145: For Complete Conjugacy model, CC.TP and CC.BF are calculated in two ways, one based on conditional posterior of  $\delta_g$  with posterior mode of  $\sigma_g^2$  plugged in (X axis), the other based on marginal posterior of  $\delta_g$  integrating out  $\sigma_g^2$  (Y axis). The bottom row is the case to identify genes with large signal, the top row is the case to identify genes with reliable measure. Simulation are based on number of replicates equal to 3 and top 1% genes are considered to be differential genes. Hyperparameters are estimated by method of moments.

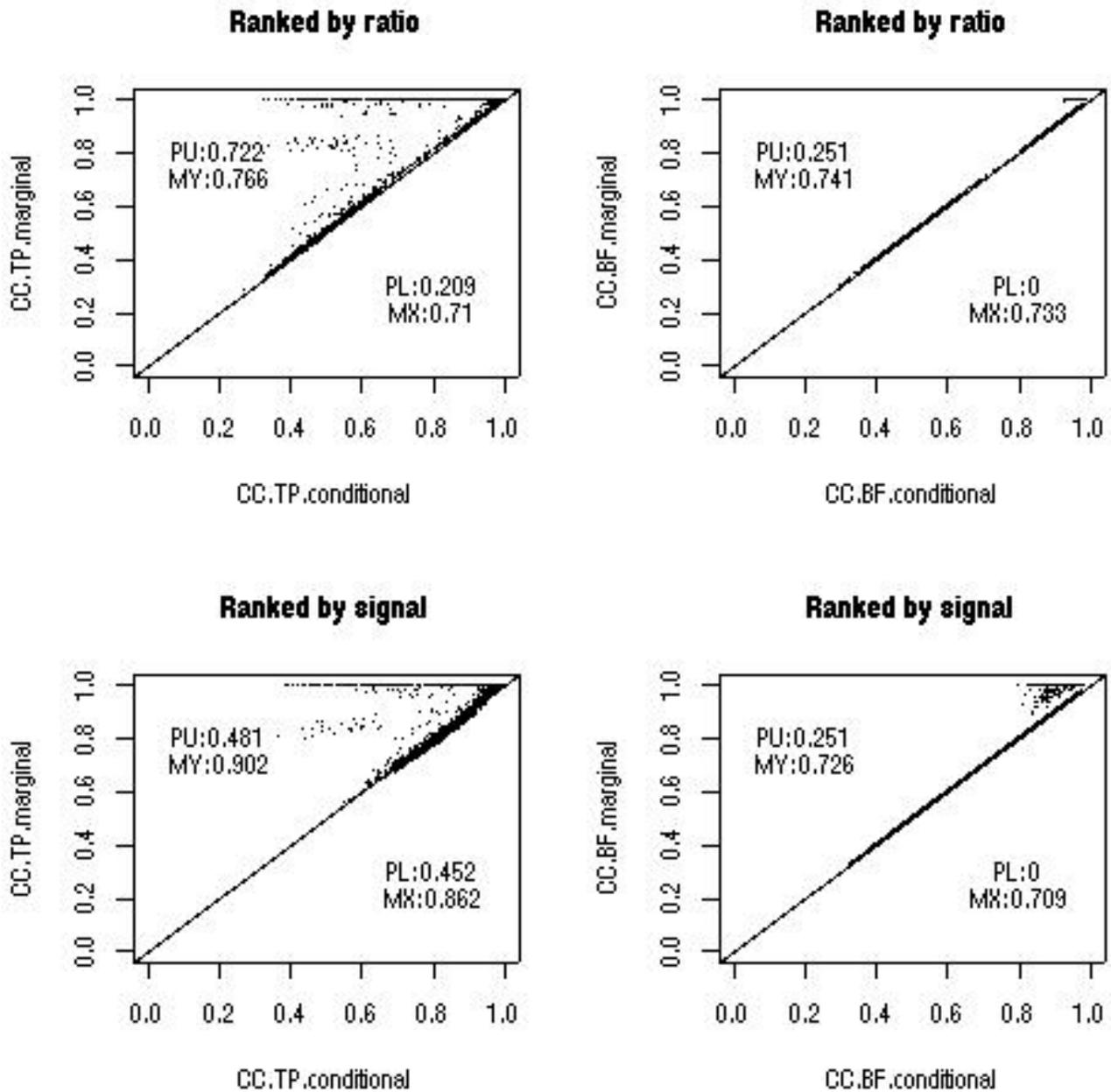


Figure 146: For Complete Conjugacy model, CC.TP and CC.BF are calculated in two ways, one based on conditional posterior of  $\delta_g$  with posterior mode of  $\sigma_g^2$  plugged in (X axis), the other based on marginal posterior of  $\delta_g$  integrating out  $\sigma_g^2$  (Y axis). The bottom row is the case to identify genes with large signal, the top row is the case to identify genes with reliable measure. Simulation are based on number of replicates equal to 3 and top 2% genes are considered to be differential genes. Hyperparameters are estimated by method of moments.

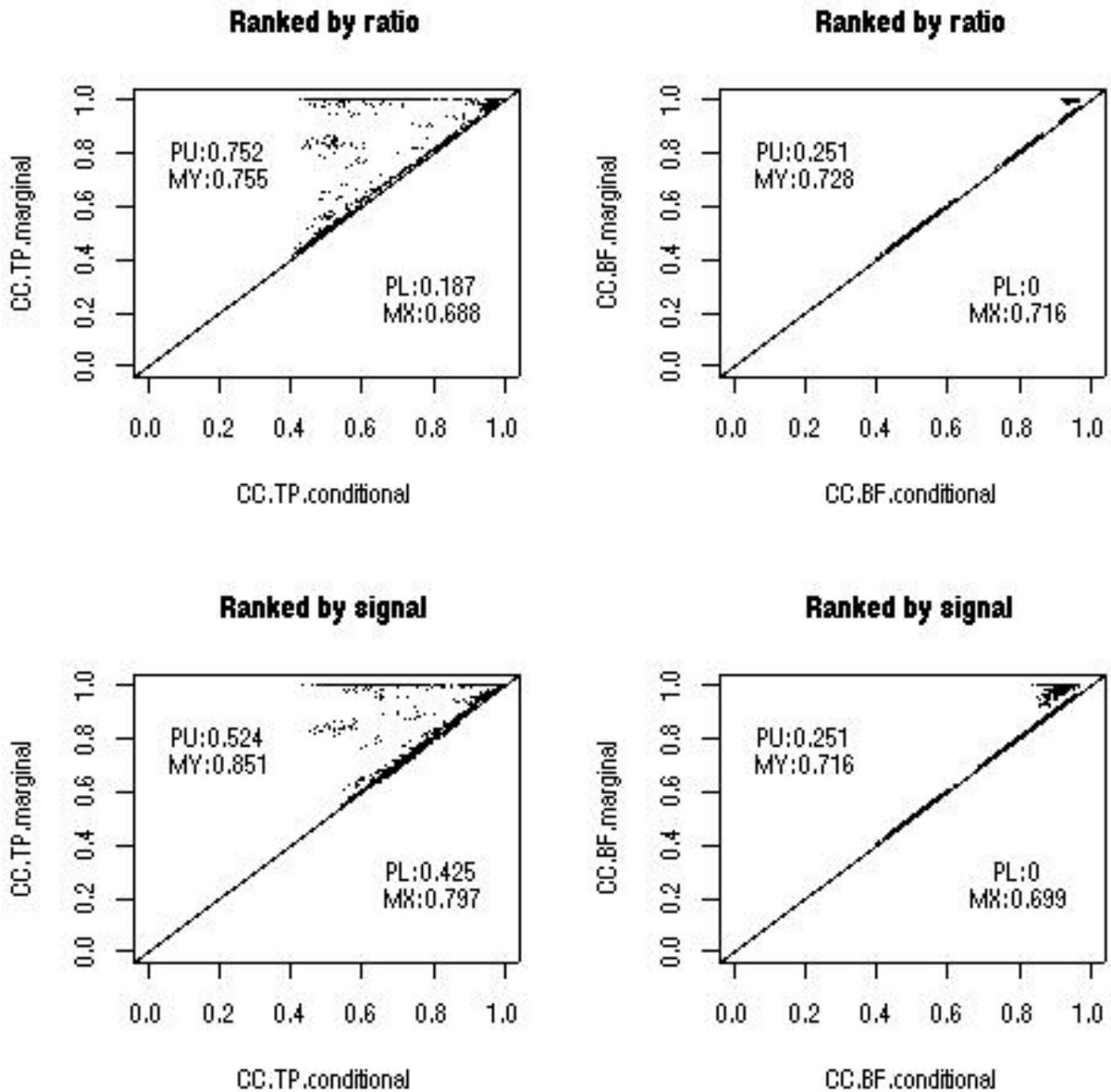


Figure 147: For Complete Conjugacy model, CC.TP and CC.BF are calculated in two ways, one based on conditional posterior of  $\delta_g$  with posterior mode of  $\sigma_g^2$  plugged in (X axis), the other based on marginal posterior of  $\delta_g$  integrating out  $\sigma_g^2$  (Y axis). The bottom row is the case to identify genes with large signal, the top row is the case to identify genes with reliable measure. Simulation are based on number of replicates equal to 3 and top 10% genes are considered to be differential genes. Hyperparameters are estimated by method of moments.

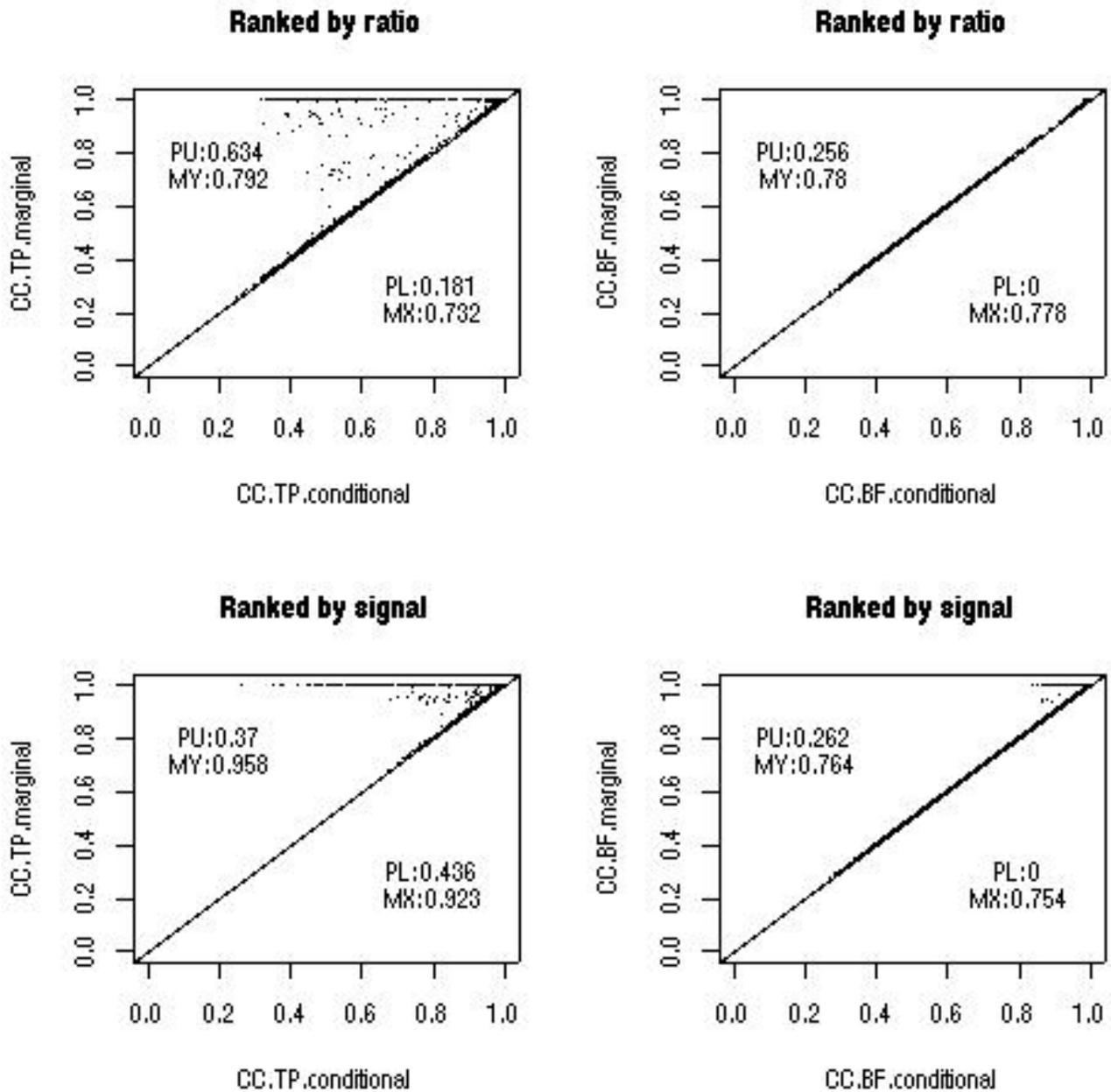


Figure 148: For Complete Conjugacy model, CC.TP and CC.BF are calculated in two ways, one based on conditional posterior of  $\delta_g$  with posterior mode of  $\sigma_g^2$  plugged in (X axis), the other based on marginal posterior of  $\delta_g$  integrating out  $\sigma_g^2$  (Y axis). The bottom row is the case to identify genes with large signal, the top row is the case to identify genes with reliable measure. Simulation are based on number of replicates equal to 10 and top 1% genes are considered to be differential genes. Hyperparameters are estimated by method of moments.

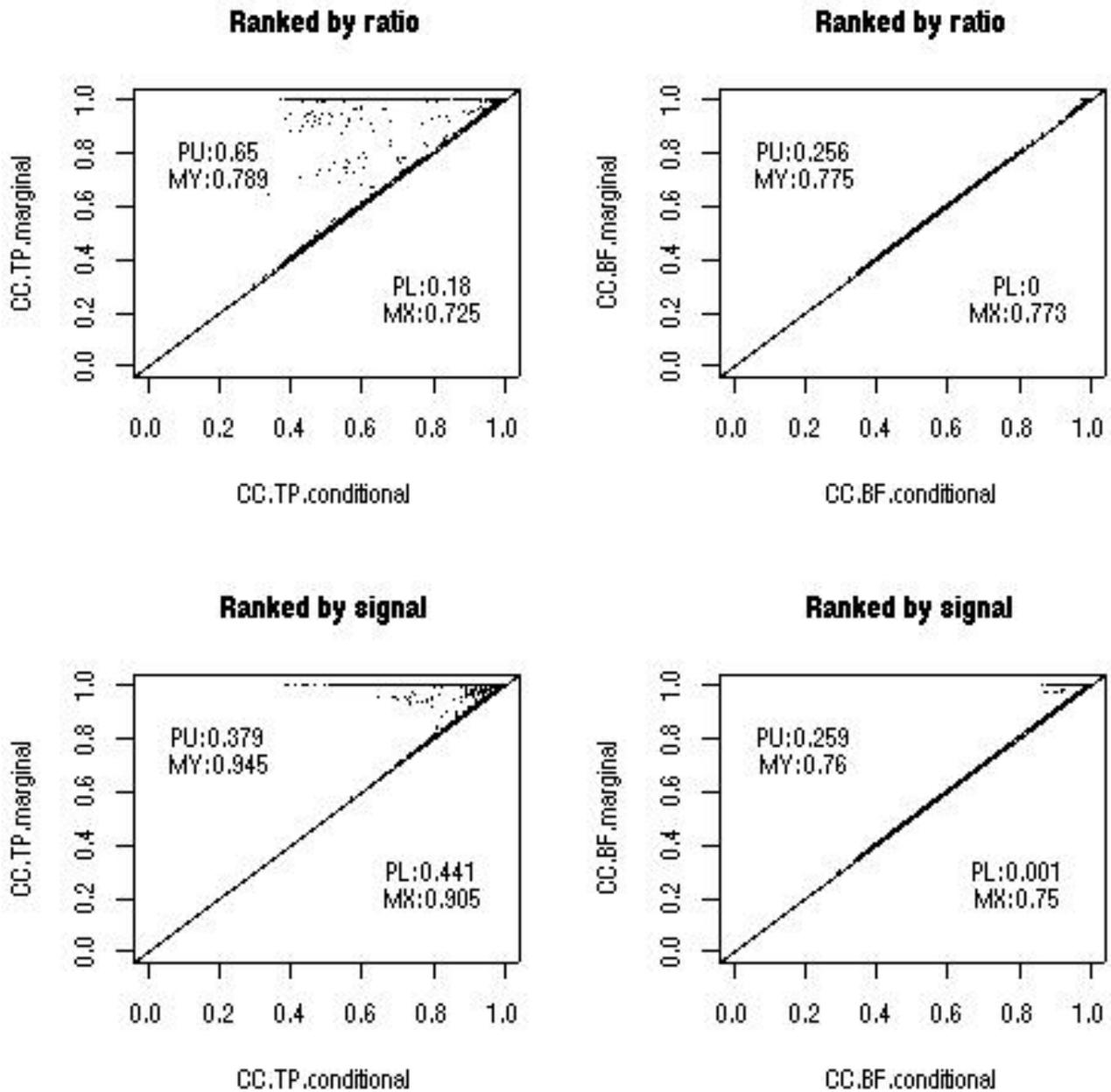


Figure 149: For Complete Conjugacy model, CC.TP and CC.BF are calculated in two ways, one based on conditional posterior of  $\delta_g$  with posterior mode of  $\sigma_g^2$  plugged in (X axis), the other based on marginal posterior of  $\delta_g$  integrating out  $\sigma_g^2$  (Y axis). The bottom row is the case to identify genes with large signal, the top row is the case to identify genes with reliable measure. Simulation are based on number of replicates equal to 10 and top 2% genes are considered to be differential genes. Hyperparameters are estimated by method of moments.

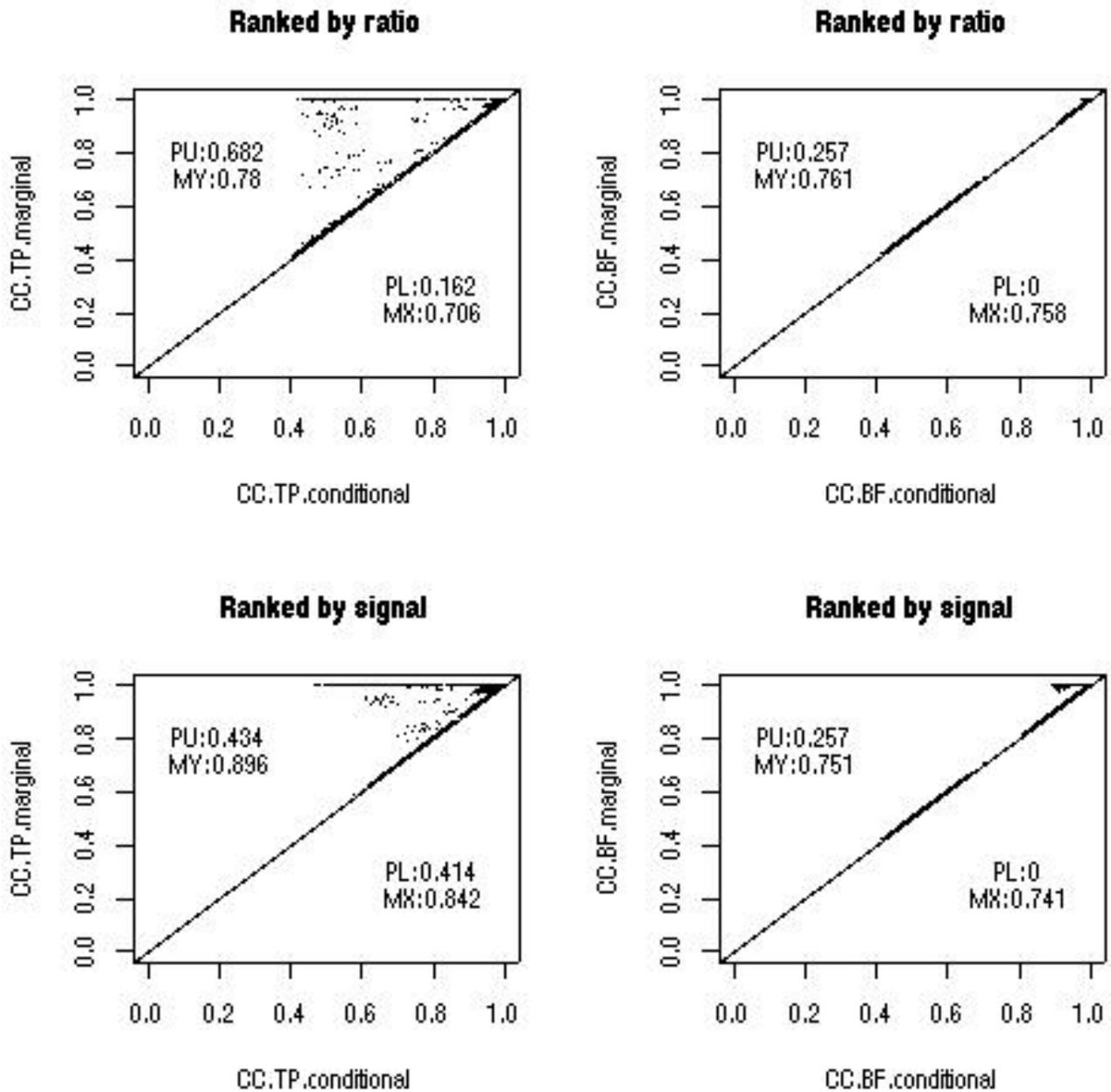


Figure 150: For Complete Conjugacy model, CC.TP and CC.BF are calculated in two ways, one based on conditional posterior of  $\delta_g$  with posterior mode of  $\sigma_g^2$  plugged in (X axis), the other based on marginal posterior of  $\delta_g$  integrating out  $\sigma_g^2$  (Y axis). The bottom row is the case to identify genes with large signal, the top row is the case to identify genes with reliable measure. Simulation are based on number of replicates equal to 10 and top 10% genes are considered to be differential genes. Hyperparameters are estimated by method of moments.

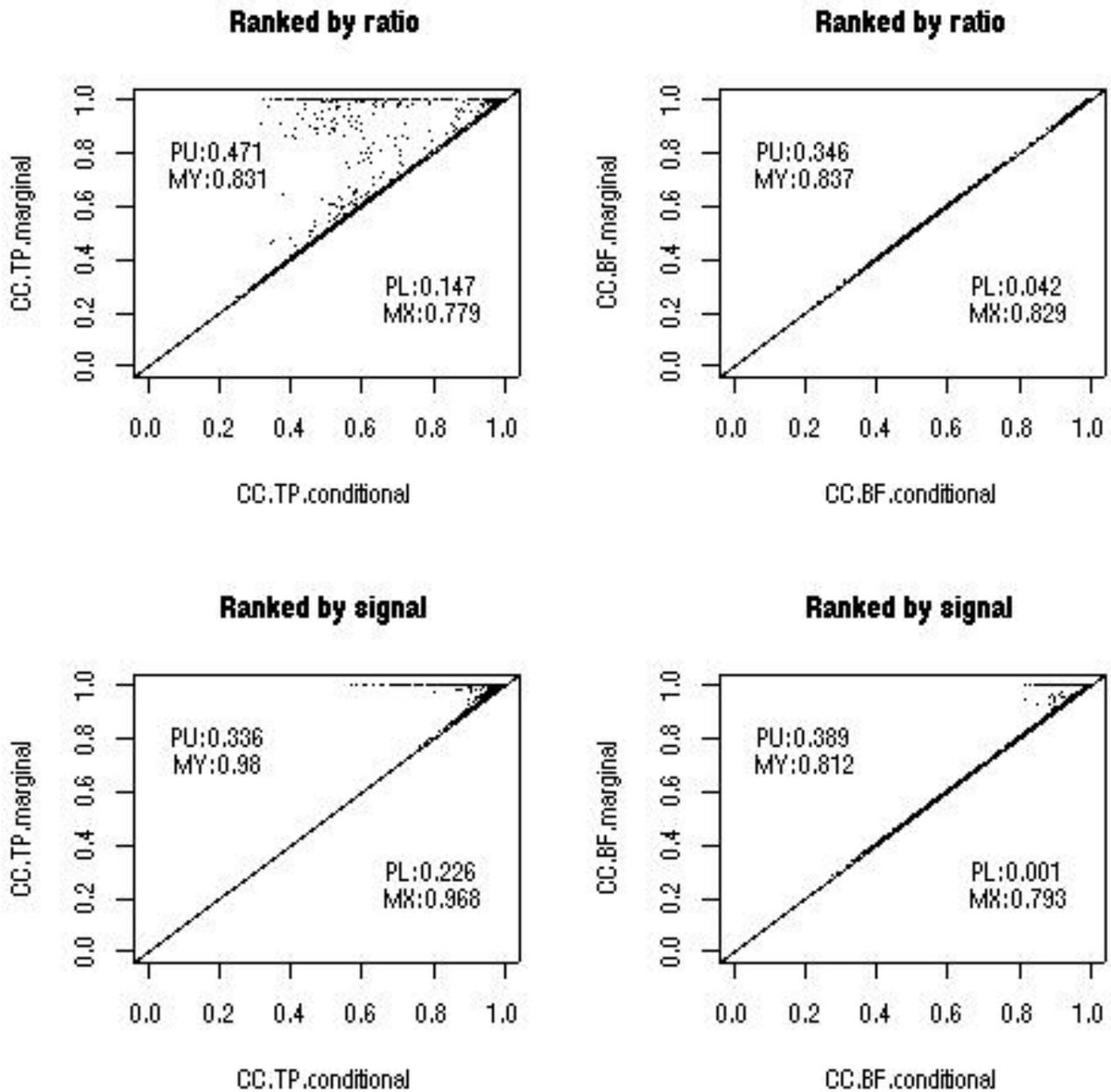


Figure 151: For Complete Conjugacy model, CC.TP and CC.BF are calculated in two ways, one based on conditional posterior of  $\delta_g$  with posterior mode of  $\sigma_g^2$  plugged in (X axis), the other based on marginal posterior of  $\delta_g$  integrating out  $\sigma_g^2$  (Y axis). The bottom row is the case to identify genes with large signal, the top row is the case to identify genes with reliable measure. Simulation are based on number of replicates equal to 100 and top 1% genes are considered to be differential genes. Hyperparameters are estimated by method of moments.

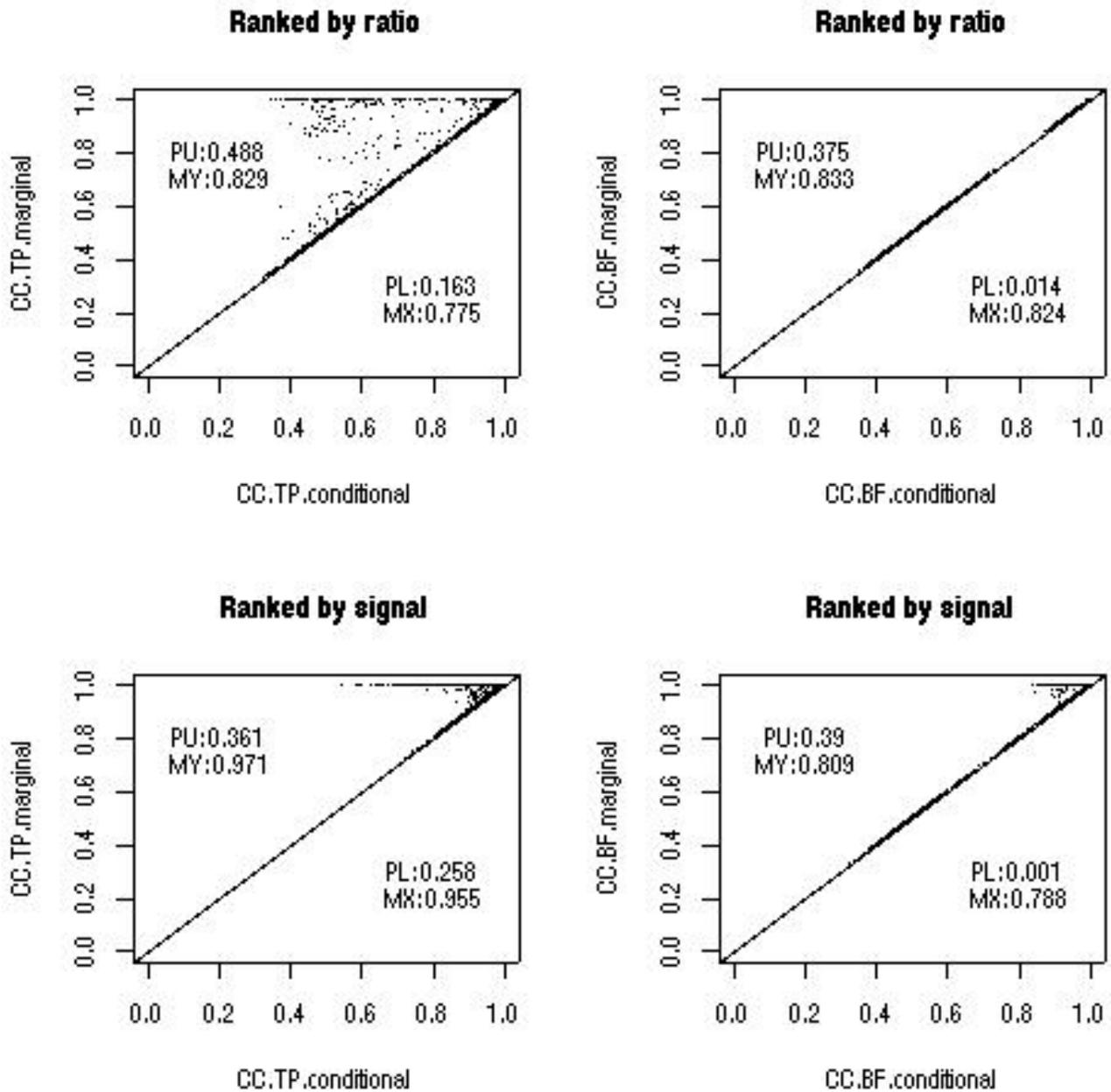


Figure 152: For Complete Conjugacy model, CC.TP and CC.BF are calculated in two ways, one based on conditional posterior of  $\delta_g$  with posterior mode of  $\sigma_g^2$  plugged in (X axis), the other based on marginal posterior of  $\delta_g$  integrating out  $\sigma_g^2$  (Y axis). The bottom row is the case to identify genes with large signal, the top row is the case to identify genes with reliable measure. Simulation are based on number of replicates equal to 100 and top 2% genes are considered to be differential genes. Hyperparameters are estimated by method of moments.

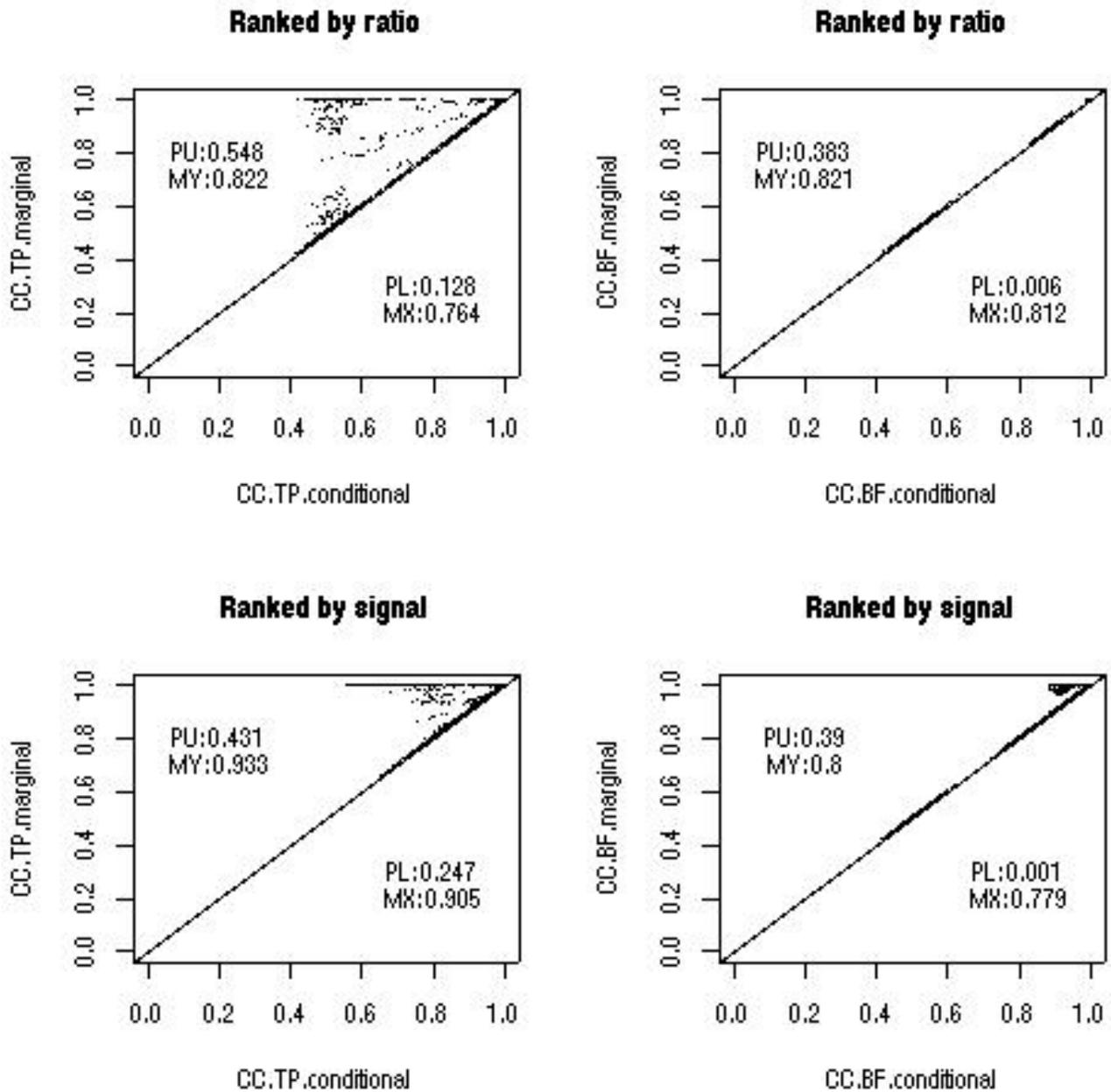


Figure 153: For Complete Conjugacy model, CC.TP and CC.BF are calculated in two ways, one based on conditional posterior of  $\delta_g$  with posterior mode of  $\sigma_g^2$  plugged in (X axis), the other based on marginal posterior of  $\delta_g$  integrating out  $\sigma_g^2$  (Y axis). The bottom row is the case to identify genes with large signal, the top row is the case to identify genes with reliable measure. Simulation are based on number of replicates equal to 100 and top 10% genes are considered to be differential genes. Hyperparameters are estimated by method of moments.

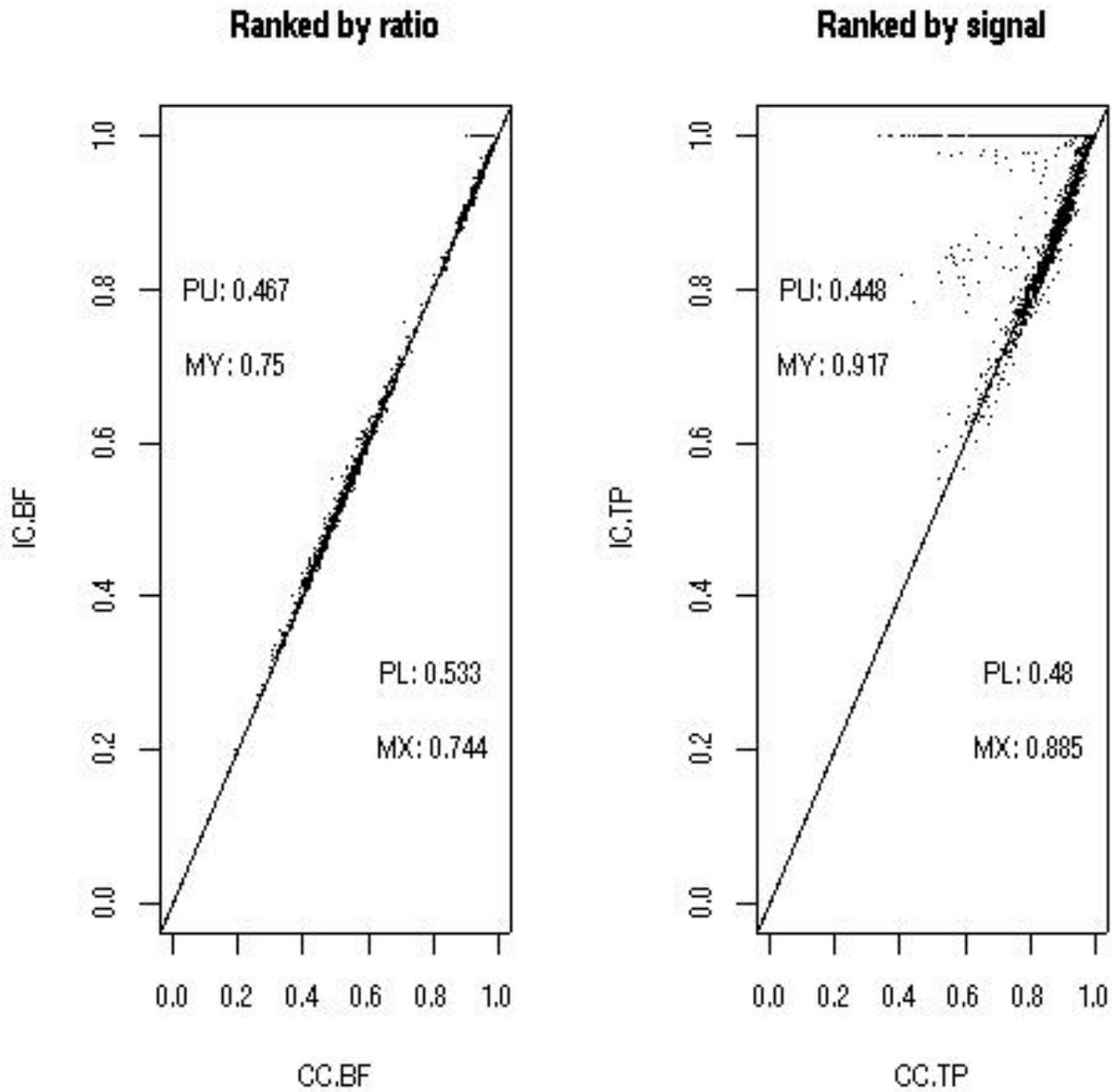


Figure 154: IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model, so does IC.BF to CC.BF. The simulation is based on Complete Conjugacy model with number of replicates equal to 3 and top 1% genes are differential genes.

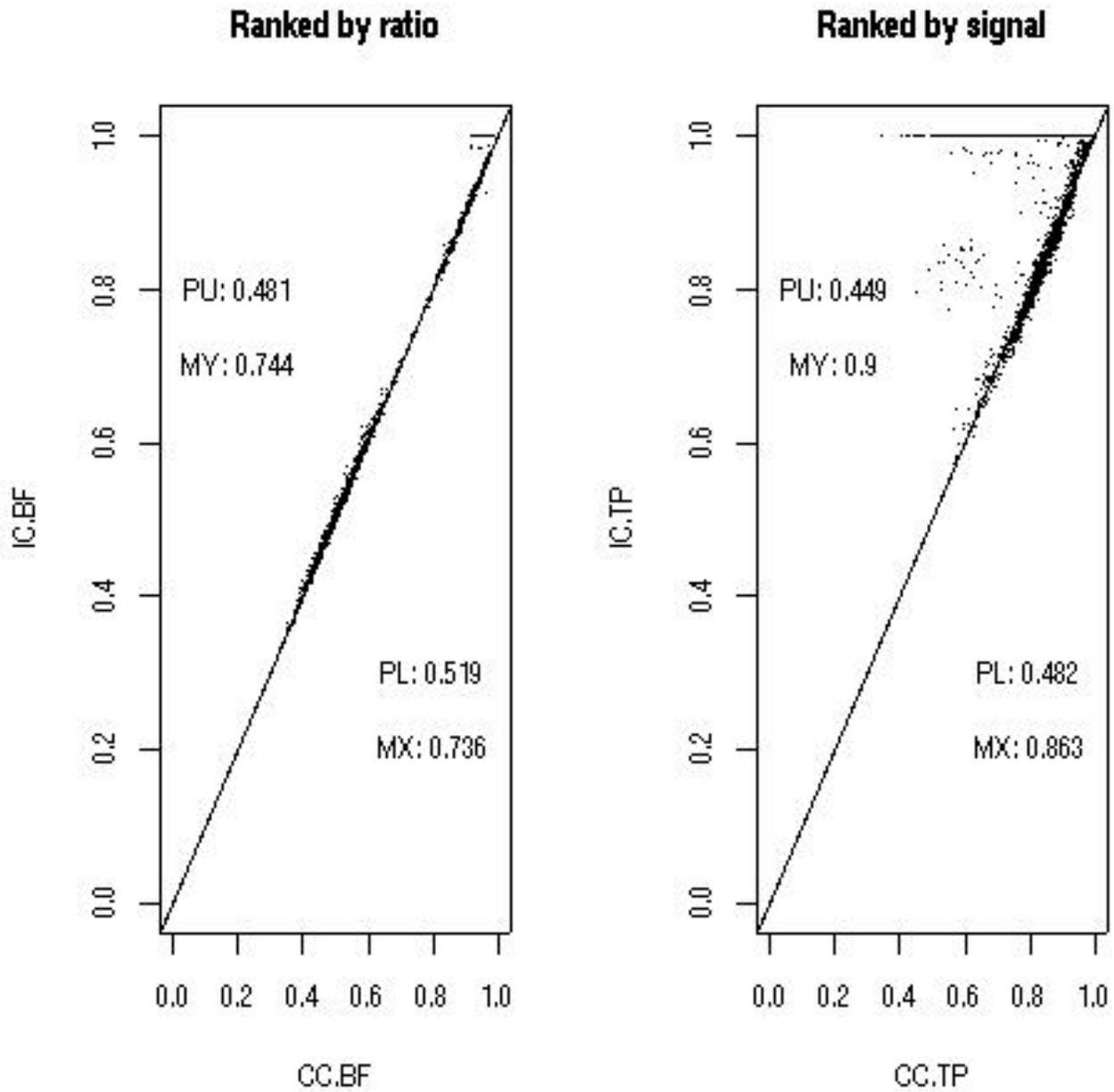


Figure 155: IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model, so does IC.BF to CC.BF. The simulation is based on Complete Conjugacy model with number of replicates equal to 3 and top 2% genes are differential genes.

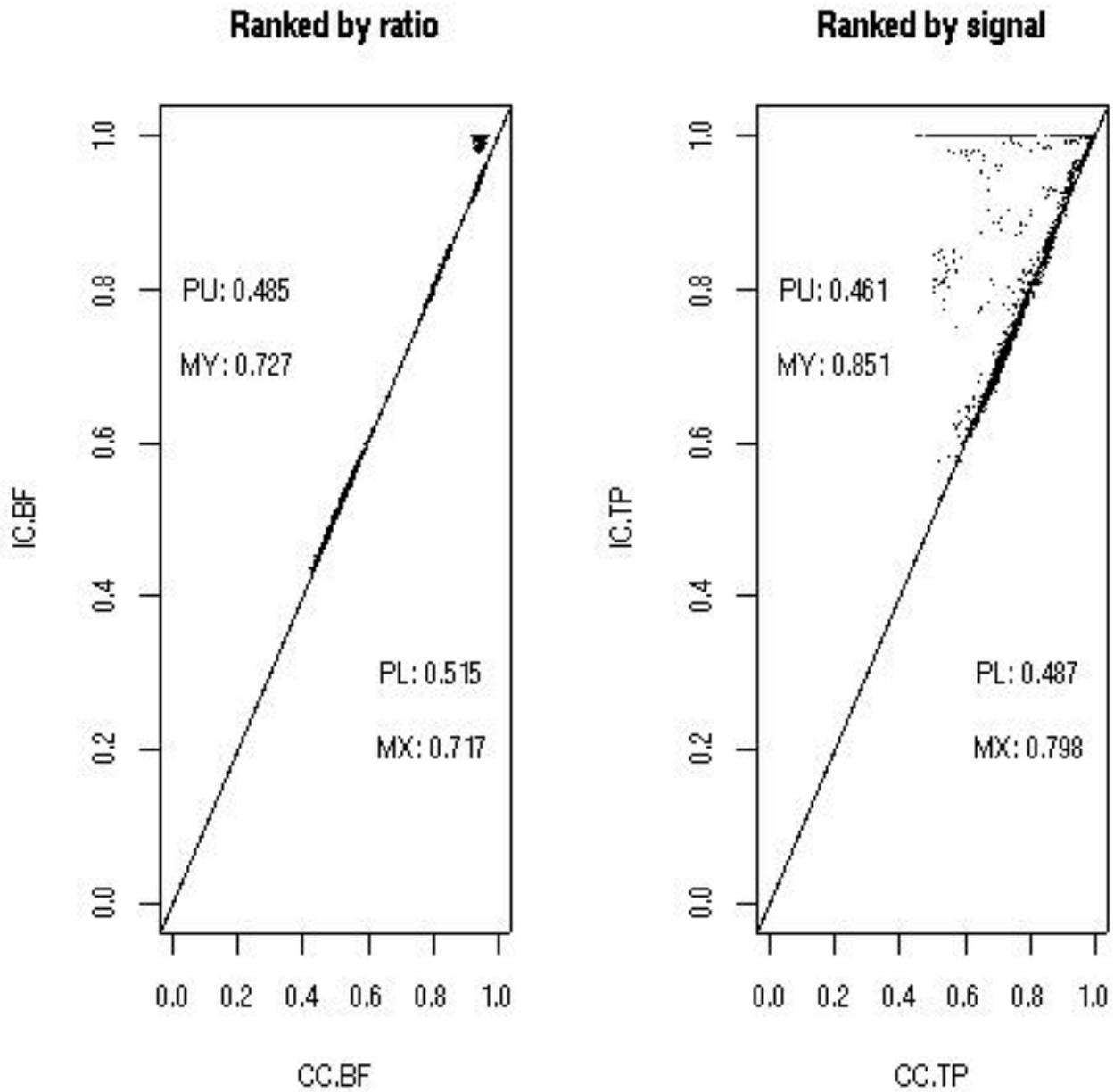


Figure 156: IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model, so does IC.BF to CC.BF. The simulation is based on Complete Conjugacy model with number of replicates equal to 3 and top 10% genes are differential genes.

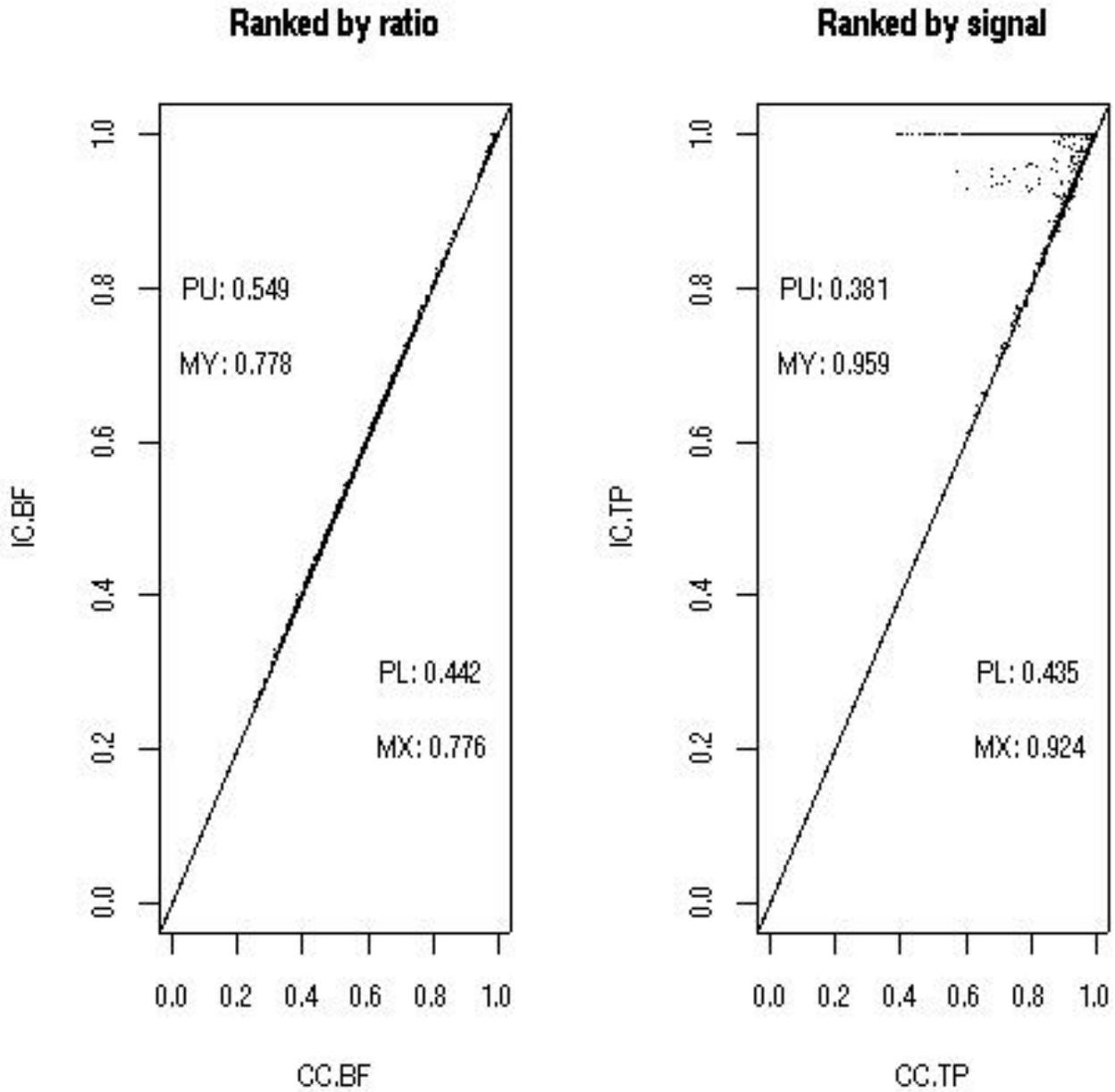


Figure 157: IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model, so does IC.BF to CC.BF. The simulation is based on Complete Conjugacy model with number of replicates equal to 10 and top 1% genes are differential genes.

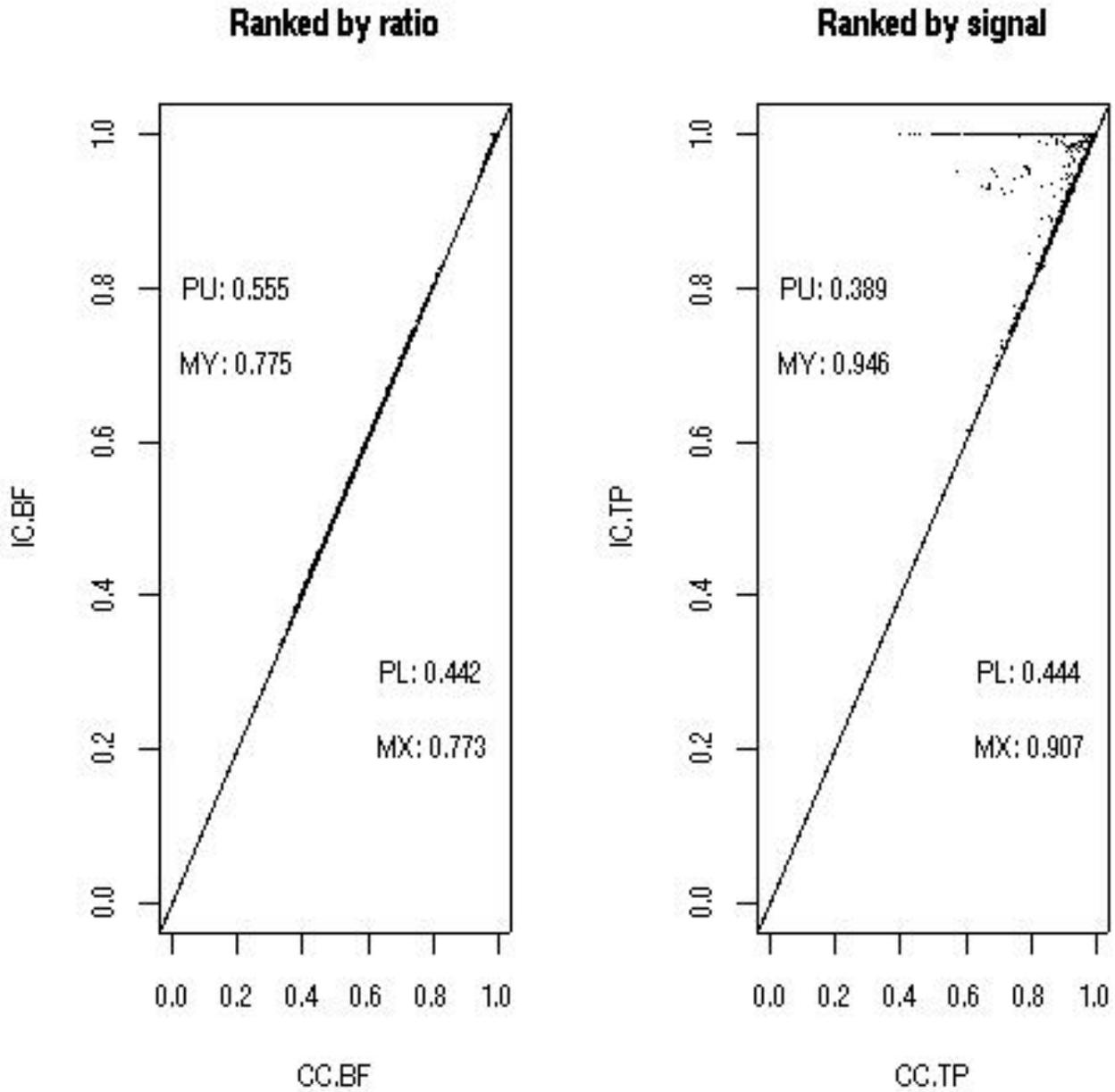


Figure 158: IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model, so does IC.BF to CC.BF. The simulation is based on Complete Conjugacy model with number of replicates equal to 10 and top 2% genes are differential genes.

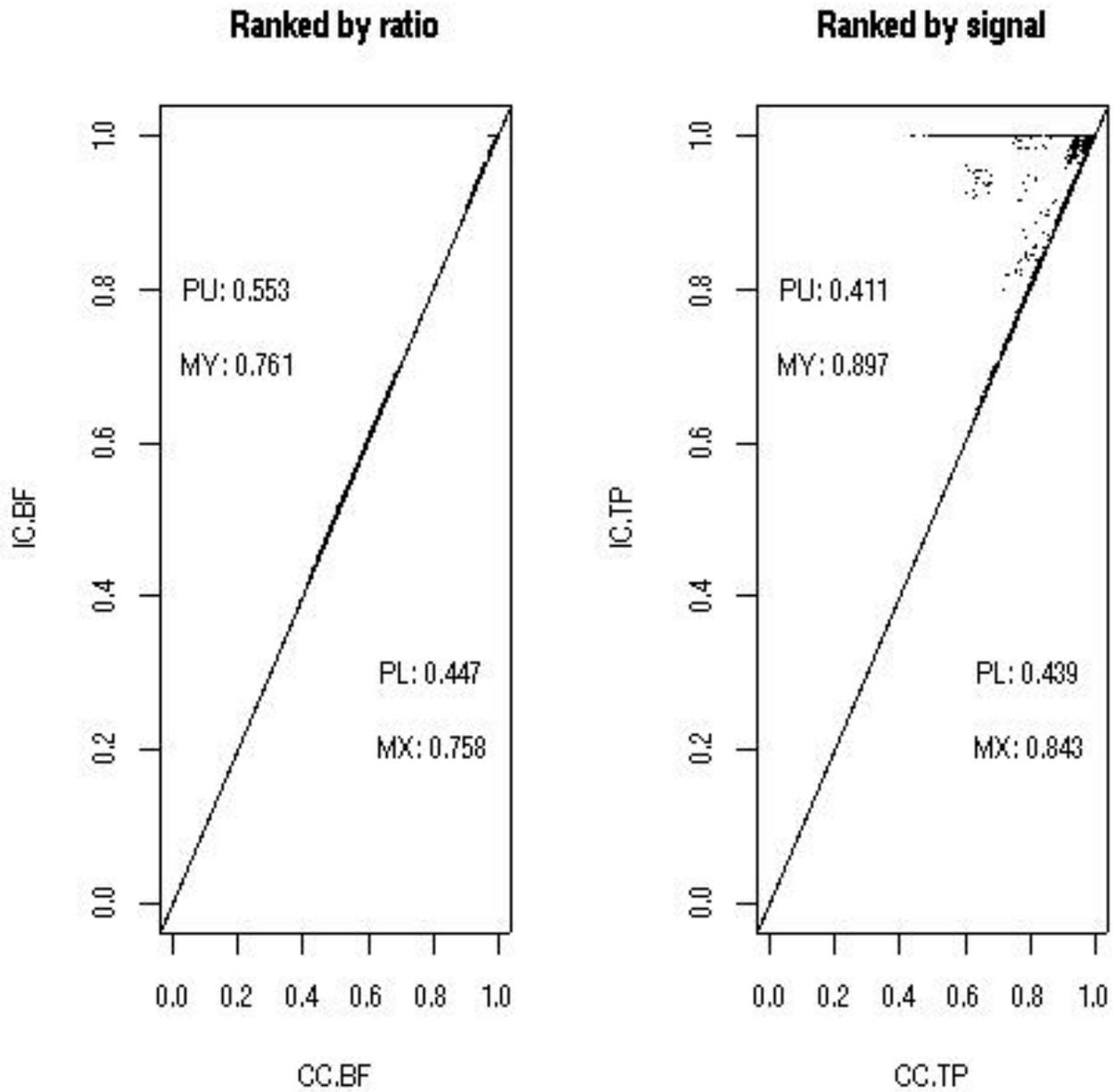


Figure 159: IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model, so does IC.BF to CC.BF. The simulation is based on Complete Conjugacy model with number of replicates equal to 10 and top 10% genes are differential genes.

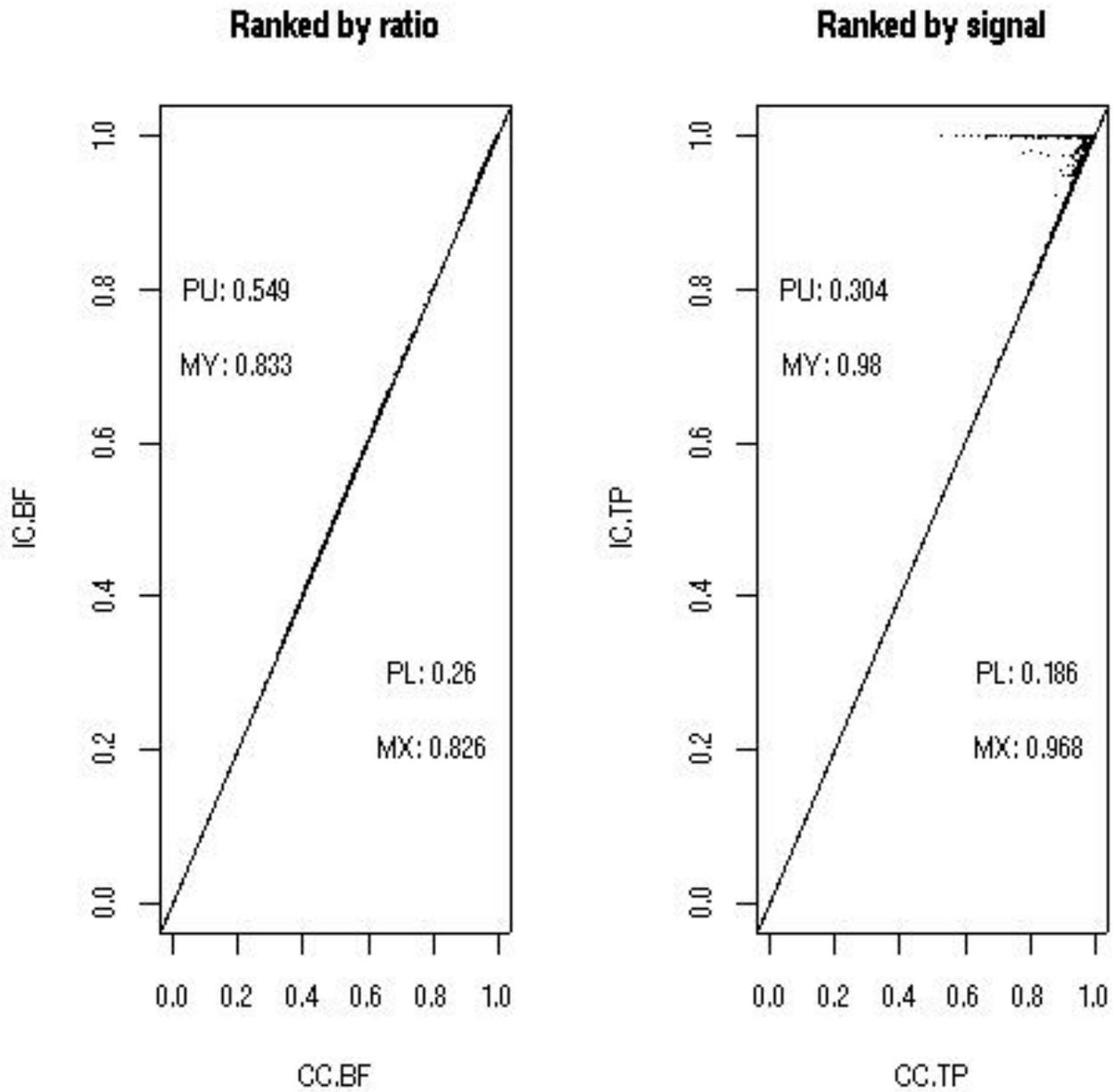


Figure 160: IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model, so does IC.BF to CC.BF. The simulation is based on Complete Conjugacy model with number of replicates equal to 100 and top 1% genes are differential genes.

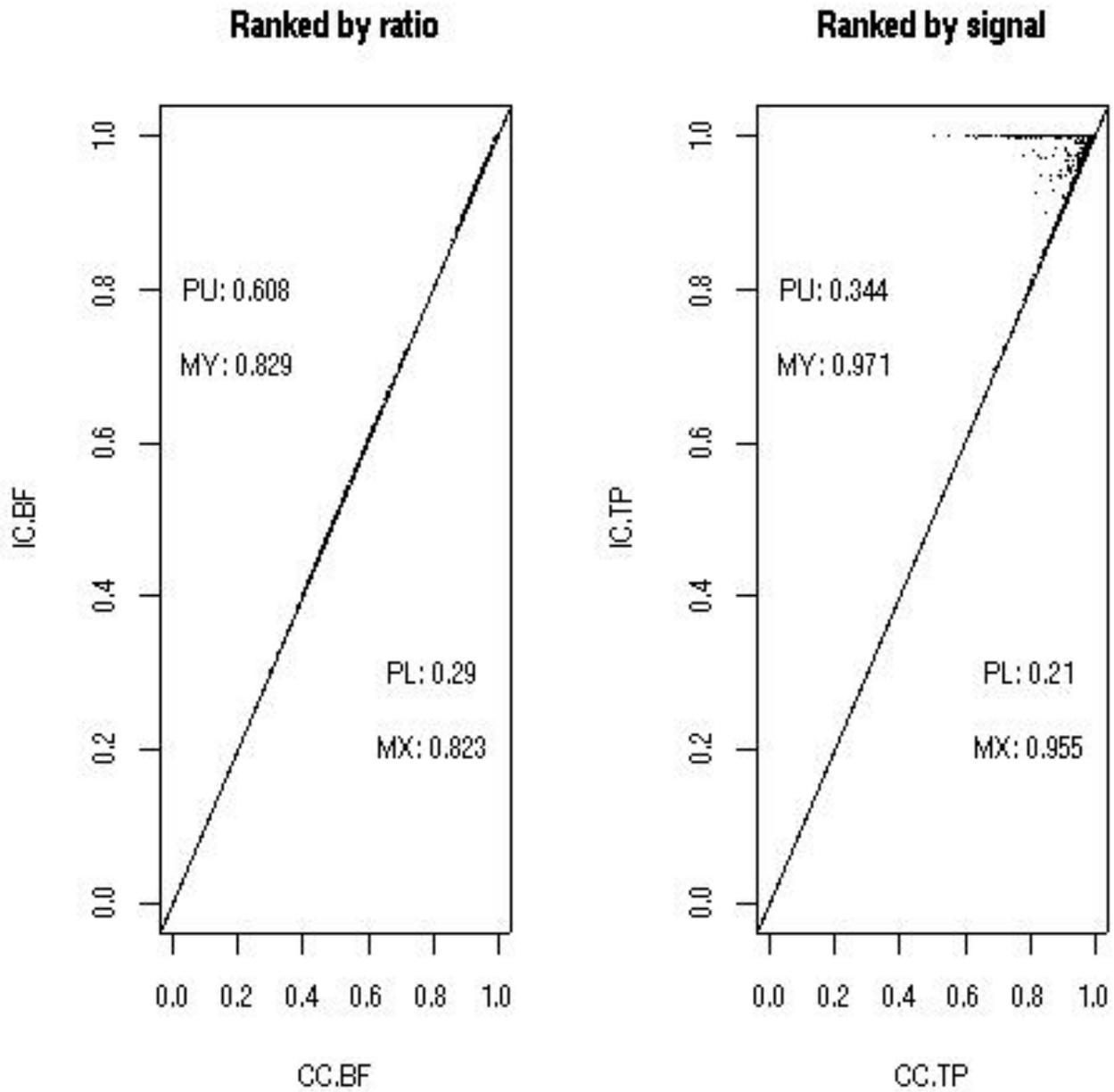


Figure 161: IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model, so does IC.BF to CC.BF. The simulation is based on Complete Conjugacy model with number of replicates equal to 100 and top 2% genes are differential genes.

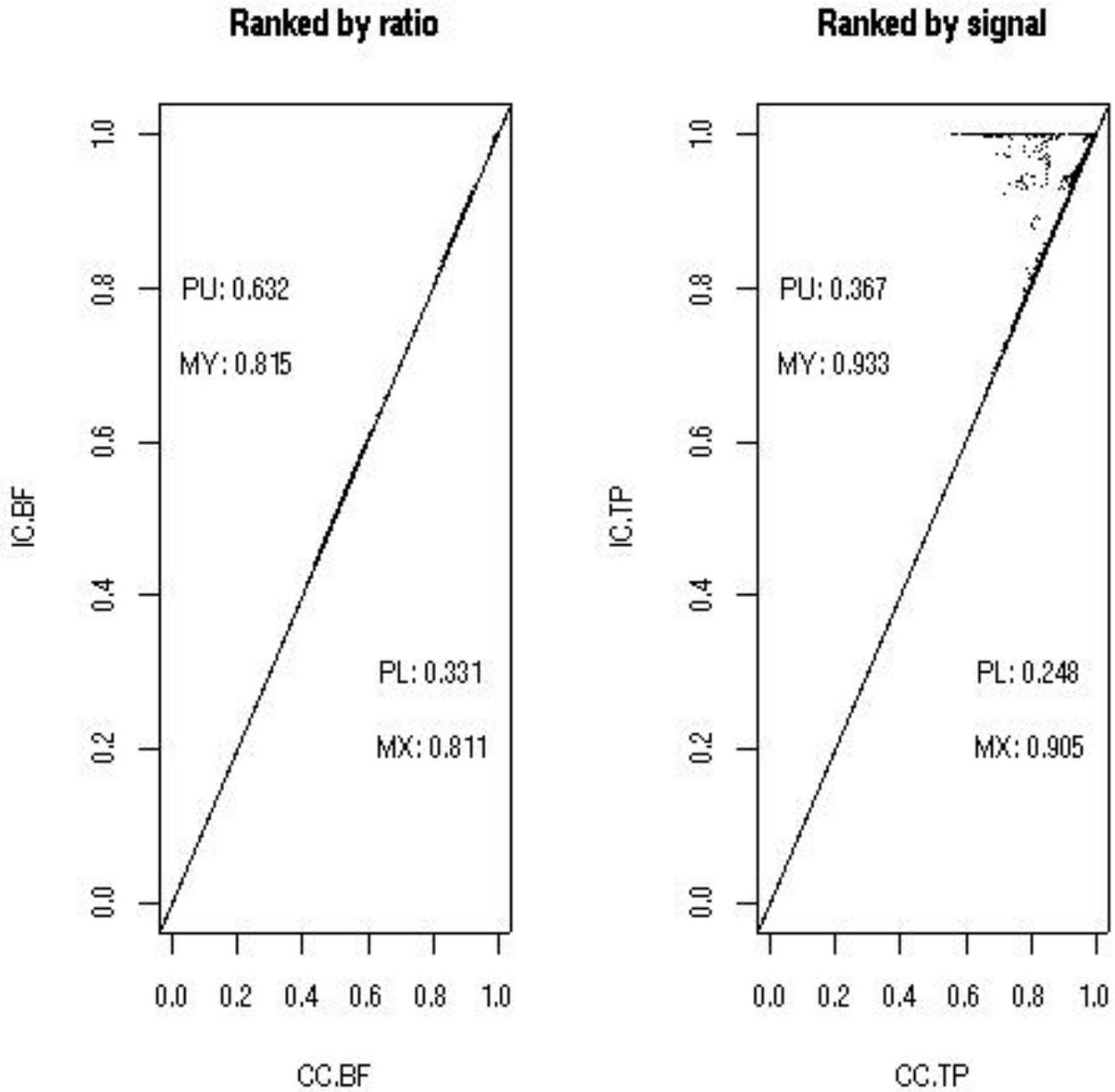


Figure 162: IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model, so does IC.BF to CC.BF. The simulation is based on Complete Conjugacy model with number of replicates equal to 100 and top 10% genes are differential genes.

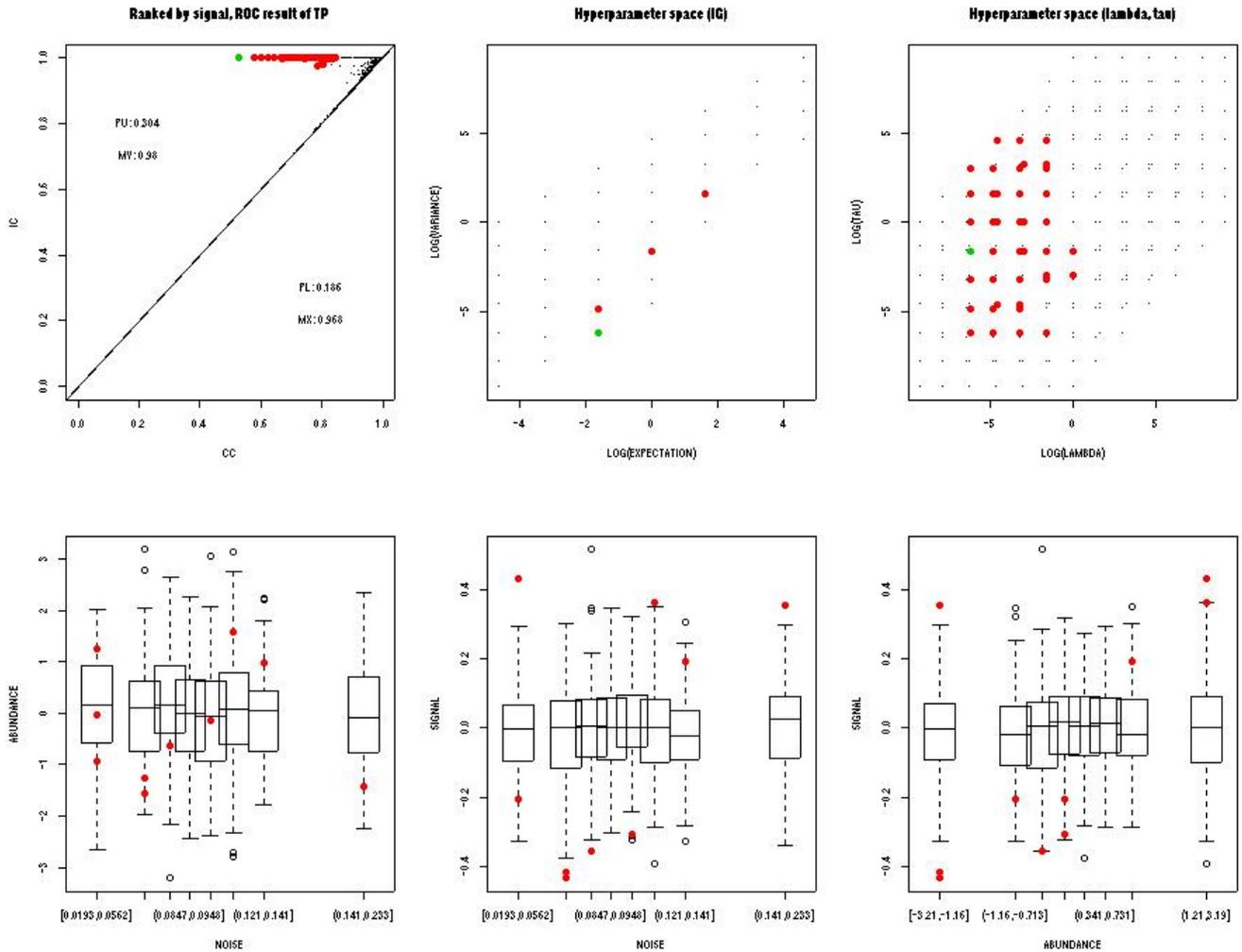


Figure 163: Diagnostic of the counter intuitive result that IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model. The first graph on the top row is the scatter plot of IC.TP vs. CC.TP for all data sets simulated from Complete Conjugacy model. Most of the time, IC.TP does equally well as CC.TP. There are cases (red and green points) that IC.TP does significantly better than CC.TP. The hyperparameter space in log scale are plotted in the second and third graph on the top row. Red and green points are mapped into these two graphs. The diagnostic plot of data simulated from Complete Conjugacy model with the worse hyperparameter scenario (green case) are plotted in the second row. The three boxplots show the relationship between abundance and noise, the relationship between signal and noise and the relationship between abundance and noise respectively. Red points in the second row are the genes with large signal.

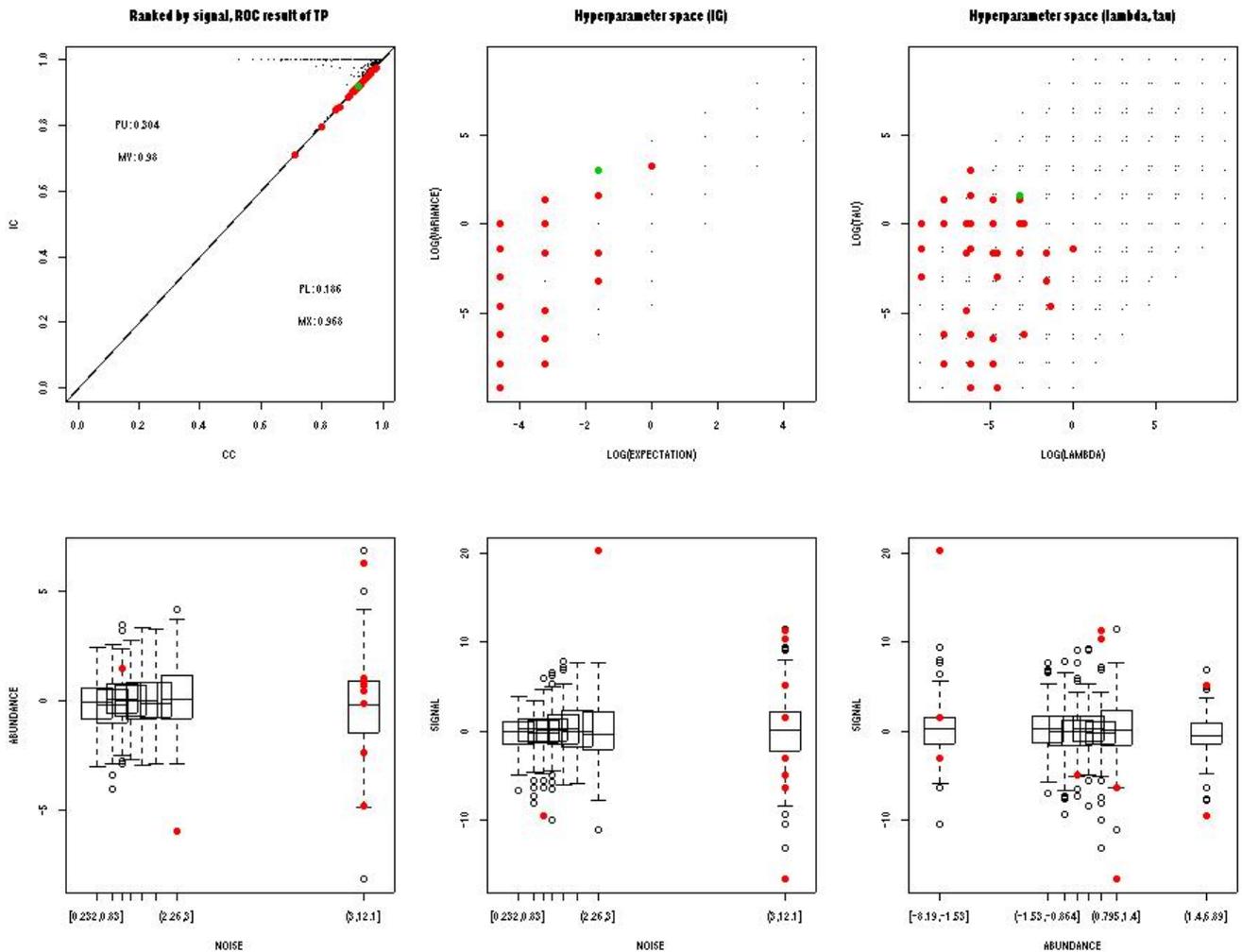


Figure 164: Diagnostic of the counter intuitive result that IC.TP outperforms CC.TP on data simulated from Complete Conjugacy model. The first graph on the top row is the scatter plot of IC.TP vs. CC.TP for all data sets simulated from Complete Conjugacy model. Most of the time, IC.TP does equally well as CC.TP. The red points indicate cases when CC.TP does better than IC.TP. The green points is the case where the difference between CC.TP and IC.TP is maximized. The hyperparameter space in log scale are plotted in the second and third graph on the top row. Red and green points are mapped into these two graphs. The diagnostic plot of data simulated from Complete Conjugacy model with the worse hyperparameter scenario (green case) are plotted in the second row. The three boxplots show the relationship between abundance and noise, the relationship between signal and noise and the relationship between abundance and noise respectively. Red points in the second row are the genes with large signal.

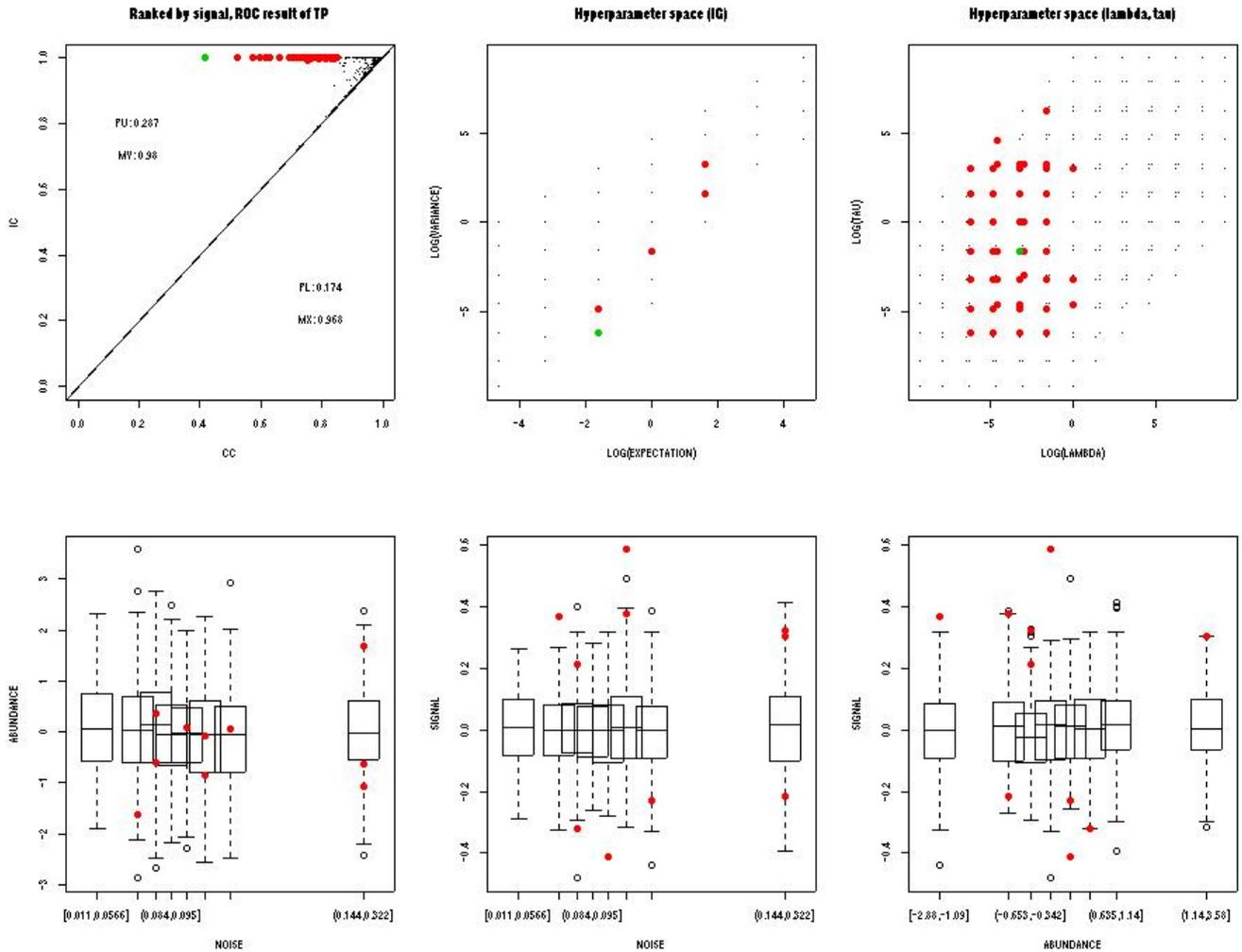


Figure 165: Same plot as Figure ?? except that simulation is done based on Independence of Abundance and Noise model. The first graph on the top row is the scatter plot of IC.TP vs. CC.TP. Most of the time, IC.TP does equally well as CC.TP. The red points indicate cases when IC.TP does better than CC.TP. The green points is the case where the difference between IC.TP and CC.TP is maximized. The hyperparameter space in log scale are plotted in the second and third graph on the top row. Red and green points are mapped into these two graphs. The diagnostic plot of data simulated from Complete Conjugacy model with the worse hyperparameter senario (green case) are plotted in the second row. The three boxplots show the relationship between abundance and noise, the relationship between signal and noise and the relationship between abundance and noise respectively. Red points in the second row are the genes with large signal.

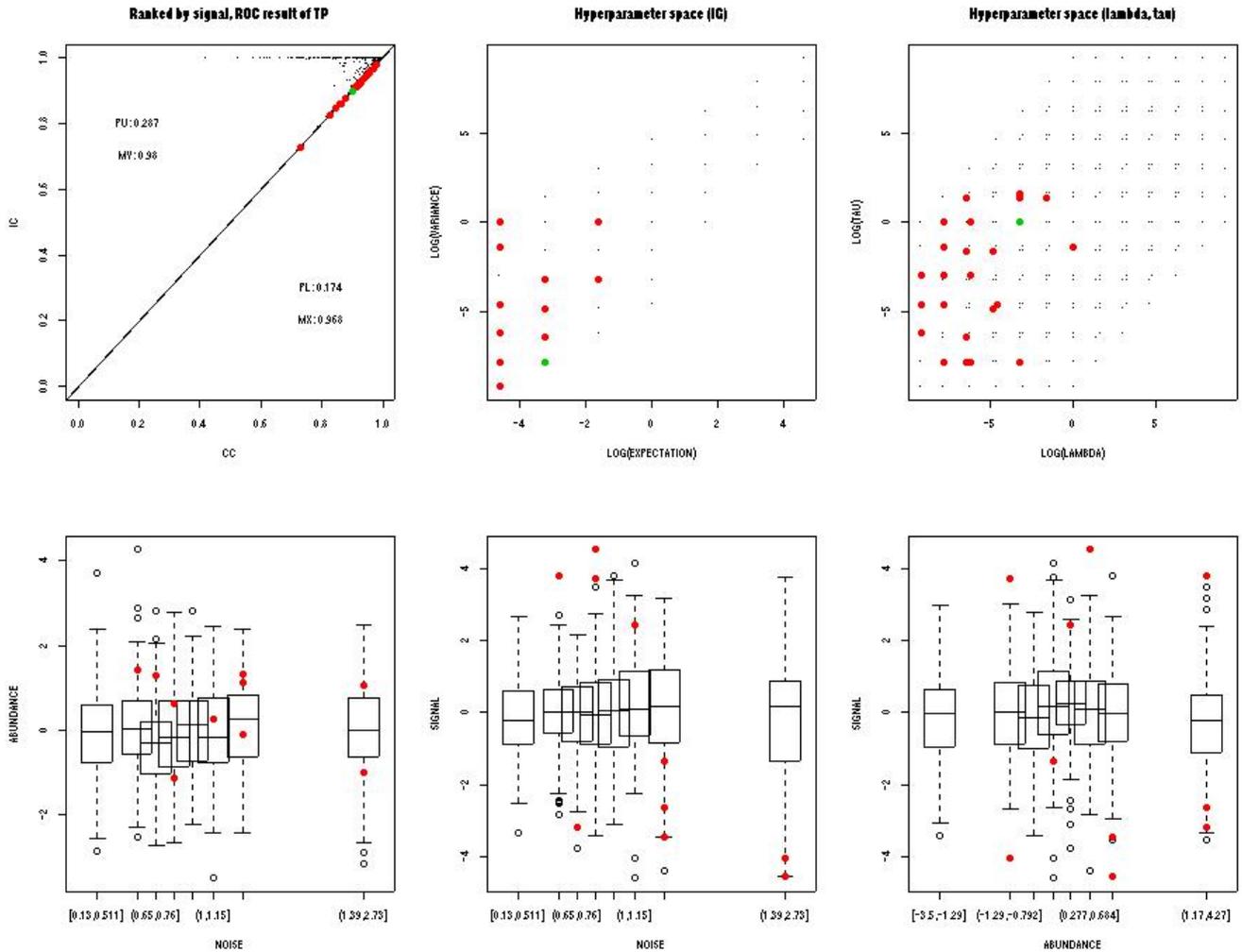


Figure 166: Same plot as Figure ?? except that simulation is done based on Independence of Abundance and Noise model. The first graph on the top row is the scatter plot of IC.TP vs. CC.TP. Most of the time, IC.TP does equally well as CC.TP. The red points indicate cases when CC.TP does better than IC.TP. The green points is the case where the difference between CC.TP and IC.TP is maximized. The hyperparameter space in log scale are plotted in the second and third graph on the top row. Red and green points are mapped into these two graphs. The diagnostic plot of data simulated from Complete Conjugacy model with the worse hyperparameter senario (green case) are plotted in the second row. The three boxplots show the relationship between abundance and noise, the relationship between signal and noise and the relationship between abundance and noise respectively. Red points in the second row are the genes with large signal.