

Research

Open Access

## Improving the power for detecting overlapping genes from multiple DNA microarray-derived gene lists

Xutao Deng<sup>1,2</sup>, Jun Xu<sup>1</sup> and Charles Wang\*<sup>1,2</sup>

Address: <sup>1</sup>Transcriptional Genomics Core, Burns Allen Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA and <sup>2</sup>Department of Medicine, UCLA David Geffen School of Medicine, Los Angeles, CA, USA

Email: Xutao Deng - dengx@ucla.edu; Jun Xu - xuj@cshs.org; Charles Wang\* - charles.wang@cshs.org

\* Corresponding author

from Symposium of Computations in Bioinformatics and Bioscience (SCBB07)  
Iowa City, Iowa, USA. 13–15 August 2007

Published: 28 May 2008

BMC Bioinformatics 2008, 9(Suppl 6):S14 doi:10.1186/1471-2105-9-S6-S14

This article is available from: <http://www.biomedcentral.com/1471-2105/9/S6/S14>

© 2008 Deng et al; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** In DNA microarray gene expression profiling studies, a fundamental task is to extract statistically significant genes that meet certain research hypothesis. Currently, Venn diagram is a frequently used method for identifying overlapping genes that meet the investigator's research hypotheses. However this simple operation of intersecting multiple gene lists, known as the Intersection-Union Tests (IUTs), is performed without knowing the incurred changes in Type I error rate and can lead to loss of discovery power.

**Results:** We developed an IUT adjustment procedure, called Relaxed IUT (RIUT), which is proved to be less conservative and more powerful for intersecting independent tests than the traditional Venn diagram approach. The advantage of the RIUT procedure over traditional IUT is demonstrated by empirical Monte-Carlo simulation and two real toxicogenomic gene expression case studies. Notably, the enhanced power of RIUT enables it to identify overlapping gene sets leading to identification of certain known related pathways which were not detected using the traditional IUT method.

**Conclusion:** We showed that traditional IUT via a Venn diagram is generally conservative, which may lead to loss discovery power in DNA microarray studies. RIUT is proved to be a more powerful alternative for performing IUTs in identifying overlapping genes from multiple gene lists derived from microarray gene expression profiling.

### Background

Nowadays many microarray-based studies adopt complex experimental design involving multiple treatments, cell lines/tissues, multiple dosages, time points, phenotypes and so on [1-4]. These studies are often involved with complex research hypotheses. For instance in one of our

previous studies [1], we were interested in identifying differentially expressed genes (DEGs) in responding to common bile duct ligation in bone marrow stem cells (BMSCs) compared against primary hepatocytes. Two DEG sets from BMSCs and hepatocytes were identified respectively, and the overlapping genes across the two cell

types were obtained. The overlapping genes produced across the two cell types allowed the identification of common biological pathways, ontological classes, and biological mechanisms across the two cell types in responding to the treatment.

The intersection operations on multiple gene lists are equivalent to performing multiple tests for the combined hypotheses on every single gene. Although there are many statistical tests proposed for gene expression studies [5-8], the problems of obtaining overlapping gene sets based on multiple tests were overlooked in microarray-based studies. To obtain genes that satisfy the specific hypotheses, researchers simply overlap the gene sets from multiple gene sets and visualized them in Venn diagrams. However, because of lacking multiplicity adjustment, this procedure overlooks the changes of statistical properties, i.e., power, type 1 error rate, p-values, during the intersection operations. This type of multiple testing for finding overlapping genes is known as the Intersection-Union Test (IUT). Despite some early efforts [9-11], the statistical properties and adjustment algorithms of IUT are not well established. Berger has proved that IUT without multiplicity adjustment is a level- $\alpha$  test [10], when the individual tests were controlled at type 1 error rate  $\alpha$ . However, the family wise error rate (FWER) for IUT  $\alpha'$  is generally much smaller than  $\alpha$ . Therefore, performing IUT without multiplicity adjustment would be very conservative and result in too many false negatives.

In this paper, we show that current overlapping operation, applying no p-value adjustment for IUT, is overly conservative in general. As a result, current microarray studies suffer from low power in detecting overlapping genes and therefore limit its use in biological data mining. We developed an analytical solution, named as Relaxed IUT (RIUT) for the multiplicity adjustment of IUTs under certain conditions. We theoretically proved that our proposed method is a less conservative and more powerful than current approaches. We demonstrated the superiority of RIUT for detecting overlapping genes in simulated data sets and complex microarray-based toxicogenomic studies.

## Results

### Monte-Carlo simulation results of RIUT

As an example to showcase the power of RIUT, the mRNA expression of a given gene is tested whether it is significantly altered by a drug treatment in multiple tissues. Suppose gene expressions were measured in  $m$  different tissues and one is interested in the overlapping DEGs. For each tissue, a two-sample t-test is performed between a treatment group and a control group, each containing  $n$  replicates to obtain a list of significant genes for that tis-

sue. Then we have an IUT that is constructed by  $m$  individual tests, each for a different tissue:

$H_{0i}$ : the drug has no effect in the  $i$ th tissue, i.e.,  $\mu_{ti} = \mu_{ci}$

$H_{Ai}$ : the drug has effect in the  $i$ th tissue, i.e.,  $\mu_{ti} \neq \mu_{ci}$ ,  $1 \leq i \leq m$ ,

where  $\mu_{ti}$  and  $\mu_{ci}$  denote the expression mean of the treatment and the control groups respectively of the  $i$ th tissue. The hypotheses for IUT are  $H_0$ : the drug shows no effect on at least one tissue vs.  $H_A$ : the drug shows effect on all tissues.

For the Monte-Carlo simulation, the expression data for the treatment and control groups were modeled as normal distributions  $N(\mu_{ti}, 1)$  and  $N(\mu_{ci}, 1)$  respectively, where  $\mu_{ci} = 0$ ,  $1 \leq i \leq m$ . We then drew  $n = 5$  samples from each of these distributions and apply RIUT and BIUT to these simulated data. Table 1 shows the estimated type 1 error rate by 10000 simulated instances of IUT formed by 2 individual tests. Expression mean  $\mu_{t1}$  is fixed at 0 for the first tissue and different values of  $\mu_{t2}$  were used to represent the drug having diverse effects ( $\mu_{t2} = 0, 0.5, \dots, 4.5, 5.0$ ) on the second tissue. Overall, the drug has no effect on the first tissue and the null hypothesis  $H_0$  of IUT is true. Results show that both RIUT and traditional Berger's IUT (BIUT) are bounded by nominal  $\alpha$  at which the individual hypotheses were tested. The actual type 1 error for IUT is generally smaller than  $\alpha$ . RIUT is less conservative than BIUT as it achieves a type 1 error rate that is closer to  $\alpha$ . To prove concept, we also tested two meta-analysis methods for combining independent tests, the Fisher's method [15] and Stouffer's method [16]. The results show that their actual type 1 rate can be so much higher than the nominal one that these methods are not suitable for the IUTs.

At  $\mu_{t1} = \mu_{t2} = 0$ , the resultant p-value distributions generated from 10000 instances were illustrated in Figure 1. Theoretically, the p-values originated from null hypothesis should appear approximately uniformly distributed (the dashed line). The adjusted  $p'$  using RIUT achieved the desired distribution as shown in Figure 1a. However, Figure 1b shows that the unadjusted  $p$  is seriously skewed to the right, indicating that the test is overly conservative.

Table 2 shows the power estimate of IUT consisting of two individual tests and both individual null hypotheses are not true ( $\mu_{t1} = 0.5$ ,  $\mu_{t2} = 0, 0.5, \dots, 4.5, 5.0$ ). RIUT demonstrated higher power than BIUT at small and moderate effect size. At large effect size, RIUT and BIUT show essentially the same power.

**Table 1: Monte-Carlo estimates of type I error rate  $\alpha'$ (%)**

$\mu_{i2}$	RIUT	BIUT	Fisher	Stouffer
0.0	0.047	0.003	0.050	0.050
0.5	0.044	0.006	0.090	0.085
1.0	0.035	0.015	0.210	0.188
1.5	0.035	0.027	0.405	0.327
2.0	0.040	0.039	0.632	0.476
2.5	0.049	0.048	0.821	0.621
3.0	0.048	0.048	0.931	0.729
3.5	0.047	0.047	0.982	0.808
4.0	0.051	0.051	0.995	0.875
4.5	0.048	0.048	0.999	0.907
5.0	0.046	0.046	1.000	0.930

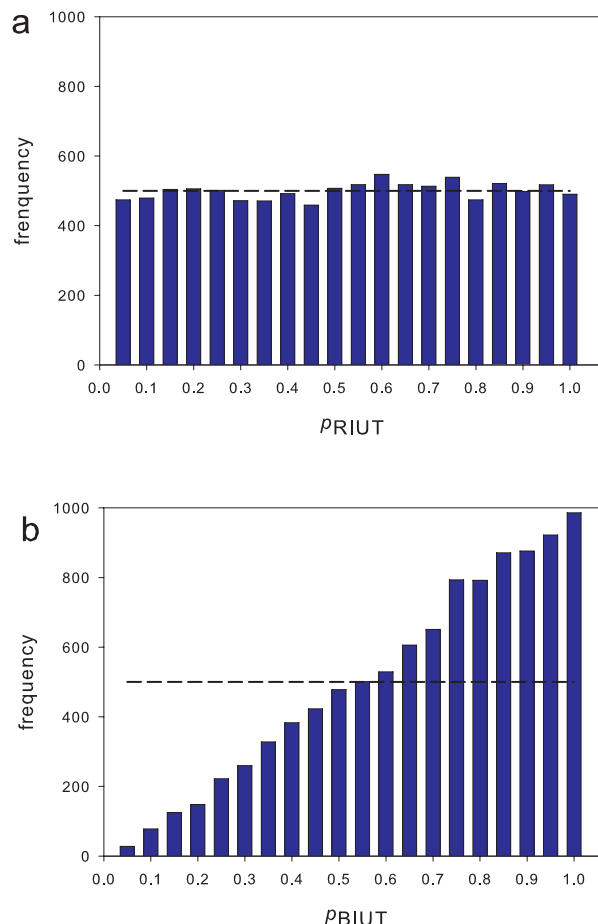
$m = 2, n = 5,$  and  $\mu_{t1} = 0.0$ ; individual tests were performed at  $\alpha = 0.05$  and each estimate were based on 10000 runs.

As a more realistic simulation, we pooled instances being positive ( $H_0$  not true) and negative ( $H_0$  true). We use  $\gamma$  to denote the simulation Bernoulli probability that an individual hypothesis is not null such that  $\text{Prob}(H_{Ai}) = \gamma$ . The population means were set at  $\mu_{t1} = \mu_{t2} = 0.5$ . According to this procedure, we simulated 10000 instances under different  $\gamma$  and the resultant number of true negative and true positive were denoted as  $K_0$  and  $K_1$  respectively. The identified false positive and true positives were notated as  $V$  and  $S$  respectively. The results in Table 3 confirmed our previous observations that RIUT is a more powerful and less conservative. The same patterns were observed in simulations using other parameter values (results not shown).

**Identifying overlapping genes that respond to multiple drug treatment**

This example illustrates how RIUT algorithm can be used in real DNA microarray-based multiple-testing problems. Originally generated from the MAQC project [2], this data set consists of rat RNA samples that came from six treatment/tissue groups. The treatment/tissue groups were aristolochic acid/liver, aristolochic acid/kidney, riddelline/liver, comfrey/liver, control/liver and control/kidney. There were six biological replicates in each treatment/tissue group. mRNA expression profiles were obtained using four commercial platforms including Affymetrix (Rat Genome 230 2.0), Agilent (Whole Rat Genome Oligo Microarray, G4131A), Applied Biosystems (Rat Genome Survey Microarray) and GE Healthcare (RatWhole Genome Bioarray, 300031) in five different labs with two labs using the Affymetrix microarray platform. Totally, 180 chips were obtained and the cross-platform probe-mapping gave rise to 4609 genes commonly detected across four platforms.

In this example, our goal is to identify the common genes responding to different drug treatments. These genes may



**Figure 1**  
**Distributions of the p-values of 10000 simulations with  $\mu_{t1} = \mu_{t2} = 0.0$  using RIUT and BIUT.** a. The distribution of RIUT p-value  $p'$ . b. The distribution of unadjusted BIUT p-value  $p$ . The dashed line indicates the hypothetical uniform distribution of the p-values of the 10000 runs.

shed lights on common cytotoxicity mechanisms of these drugs. We used the liver/control (L\_CTL) group as control and the 3 drugs treatment at rat liver were referred to as L\_AA, L\_CFY, and L\_RDL respectively. The IUT consists of two individual tests, (i.e.,  $t_1$ : L\_AA vs. L\_CTL and  $t_2$ : L\_CFY vs. L\_CTL), each testing whether a gene differentially expressed in response to one specific drug treatment versus control. The combined IUT was used to identify the genes that differentially expressed in combined each pair of the three treatments. Table 4 shows the number of overlapping differentially expressed genes identified using RIUT compared with the traditional BIUT at different labs/platforms. As expected, the number of significant genes obtained using RIUT is consistently greater than that obtained using BIUT. The magnitude of increase ranges from 13% to 184%. For many IUTs, the number of

**Table 2: Monte-Carlo estimates of power (%)**

$\mu_2$	$\alpha = 0.05$		$\alpha = 0.01$	
	RIUT	BIUT	RIUT	BIUT
0.5	0.057	0.014	0.007	0.001
1.0	0.059	0.029	0.006	0.003
1.5	0.077	0.064	0.008	0.006
2.0	0.085	0.081	0.014	0.012
2.5	0.101	0.101	0.020	0.020
3.0	0.103	0.103	0.023	0.023
3.5	0.109	0.109	0.030	0.030
4.0	0.105	0.105	0.024	0.024
4.5	0.112	0.112	0.027	0.027
5.0	0.104	0.104	0.024	0.024

$m = 2, n = 5$ , and  $\mu_{t_1} = 0.5$ ; individual tests were performed at  $\alpha = 0.05$  or  $\alpha = 0.01$  and each estimate based on 10000 runs.

identified overlapping genes using traditional BIUT is close to the number of nominal false positives (230), suggesting that BIUT lacks power to identify true overlapping differentially expressed genes. The testing results using real microarray data confirm our analytical results of Theorem 1 and Theorem 2 (shown in Methods). In addition our Monte-Carlo simulation demonstrates that RIUT is a more powerful and less conservative approach than BIUT.

**Detecting genes with time-course and dose-response effect to chemical treatment**

To further demonstrate the applicability of RIUT in microarray studies, we focus on the rat cadmium toxicogenomic data set [3,18]. This study employed a more complex study design, in which both gene expression and cytotoxicity changes were profiled in a multi-dose multi-time-point setting. Briefly, primary rat hepatocytes were isolated and were exposed to three different doses of cadmium acetate (0, 1.25 and 2.0  $\mu\text{M}$ ) for 2 h. Cells were collected at 0, 3, 6, 12 and 24 h in all three groups (0, 1.25

**Table 3: Simulation results of pooled instances**

$\gamma$	$K_0$	$K_1$	RIUT		BIUT	
			V (type I)	S(power)	V (type I)	S(power)
0.1	9899	101	482(0.049)	12(0.119)	24(0.002)	1(0.010)
0.2	9607	393	447(0.047)	39(0.099)	32(0.003)	6(0.015)
0.3	9087	913	394(0.043)	98(0.107)	29(0.003)	8(0.009)
0.4	8427	1573	363(0.043)	126(0.080)	33(0.004)	21(0.013)
0.5	7560	2440	310(0.041)	202(0.083)	30(0.004)	28(0.011)
0.6	6470	3530	270(0.042)	264(0.075)	22(0.003)	35(0.010)
0.7	5023	4977	209(0.042)	317(0.064)	23(0.005)	59(0.012)
0.8	3534	6466	119(0.034)	410(0.063)	10(0.003)	83(0.013)
0.9	1893	8107	75(0.040)	483(0.060)	10(0.005)	84(0.010)

$m = 2, n = 5$ , and  $\mu_{t_1} = \mu_{t_2} = 0.5$ ; individual tests were performed at  $\alpha = 0.05$  and each estimate were based on 10000 runs.

**Table 4: Number of differentially expressed genes in all drug treatment groups**

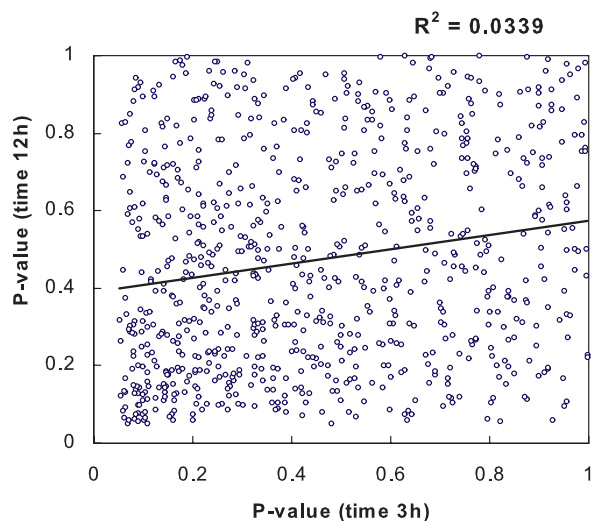
Platforms	IUT : A		IUT : B		IUT : C	
	RIUT	BIUT	RIUT	BIUT	RIUT	BIUT
Applied Biosystems	1160	1011	<b>1143</b>	<b>1008</b>	940	763
Agilent	362	262	525	413	452	159
GE Healthcare	697	528	862	723	639	415
Affymetrix (Site 1)	359	251	502	382	<b>422</b>	<b>175</b>
Affymetrix (Site 2)	524	375	724	556	521	289

$\alpha = 0.05$  (Nominal # false positive < 230)  
 IUT A:  $t_1$ : L\_AA vs. L\_CTL and  $t_2$ : L\_CFY vs. L\_CTL  
 IUT B:  $t_1$ : L\_RDL vs. L\_CTL and  $t_2$ : L\_CFY vs. L\_CTL  
 IUT C:  $t_1$ : L\_AA vs. L\_CTL and  $t_2$ : L\_RDL vs. L\_CTL

and 2.0  $\mu\text{M}$  Cd) for cytotoxicity evaluation by lactase dehydrogenase (LDH) leakage as well as for mRNA expression profiling by DNA microarray. Affymetrix GeneChip<sup>®</sup> oligonucleotide arrays (RatTox U34) were used for mRNA expression profiling. There are 972 probe sets representing ~800 important toxicology-related genes in the RT U34 array. The microarray experiment was repeated using primary hepatocytes from 3 animals, each with 2 replicates (independent cultures) for each dosage (3 dosage levels) at each time point (5 time points), resulting in a total of 90 chips (3 animals • 2 replicates • 3 doses • 5 time points). The 2 replicates were averaged in our analysis.

Firstly, to identify differentially expressed genes in responding to cadmium treatment at each time point, two-sample t-tests were performed between the treatment (1.25 and 2.0  $\mu\text{M}$  Cd) and control at each time point. Secondly, to identify the genes with persistent differentially expression due to cadmium exposure across different time points, the overlapping of the DEGs at both short term (3 h) and long term (12 h) were identified using our proposed method. The second research question was then formulated an IUT problem which could be solved using our proposed method.

We constructed an IUT consisting of two individual tests ( $t_1$ : 2.0  $\mu\text{M}$  Cd vs control at 3 h, and  $t_2$ : 2.0  $\mu\text{M}$  Cd vs control at 12 h). The joint p-value distribution from  $t_1$  and  $t_2$  for all genes is illustrated in Figure 2. It is shown that  $t_1$  and  $t_2$  were approximately independent ( $R^2 = 0.03$ ) and their p-values were approximately uniformly distributed under null hypothesis ( $p_1 > 0.05$  and  $p_2 > 0.05$ ). Therefore the assumptions in Theorem 1 and 2 were not violated. In fact for most microarray-based studies, these assumptions need to be checked by scatter plot.



**Figure 2**  
**The joint p-value distribution from  $t_1$  and  $t_2$  for all genes.** The IUT consists of two individual two-sample t tests ( $t_1$ : 2.0  $\mu$ M Cd vs. control at 3 h,  $t_2$ : 2.0  $\mu$ M Cd vs. control at 12 h).

We performed RIUT and BIUT ( $m = 2, n = 3$ ) for all 972 probe sets. RIUT identified 80 overlapping probe sets and traditional BIUT identified 19 at  $\alpha = 0.05$ . The power of IUT methods directly affected the identified gene set on which the follow-up pathway interpretation was based. To demonstrate this effect, a pathway enrichment search was performed by comparing the IUT-identified gene sets with the specific KEGG pathways [19-21] using a Fisher's Exact Test on DAVID 2006 [22], resulting in an enrichment p-value for each pathway which were listed in Table 5. Using the RIUT gene set, we identified 3 significantly over-enriched pathways (Fisher Exact p-value < 0.05). However, using the BIUT gene set, no significant pathways in KEGG was identified. It has been reported that the *metabolism of xenobiotics by cytochrome p450* pathway is significantly affected by cadmium exposure [23,24]. The activity of *MAPK signaling* pathway and *porphyrin and chlorophyll* metabolism pathway have also been affected by environmental cadmium exposure [25,26]. Again, Our analysis results using Cd time-course data set indicate that RIUT is a more powerful method than BIUT and the significance can be seen in the biological pathway identification.

**Discussion**

RIUT can be improved in three ways to be applicable in more scenarios: (1) estimating unknown nuisance parameters; (2) dealing with more than 2 individual tests; and (3) combining non-independent tests. For the scenario (1), the current approach for estimating  $\pi$  with an arbitrary  $\lambda$  was originally proposed by Storey [14] for estimat-

ing false discovery rate. The authors proposed a bootstrap method for finding optimal  $\lambda$  which can also be used here. It should be noted that Theorem 2 is valid no matter how the  $\lambda$  is chosen and how good  $\pi$  is estimated. For IUTs consisting of more than 2 individual tests, it is difficult to obtain the analytical solution as Theorem 1. However, we can apply a step-up procedure which agglomeratively applies RIUT on its least significant individual test. It is more difficult to extend our procedure for non-independent tests. It may need a resampling-based algorithm to incorporating correlation structure of multiple tests and dealing with non-normality issues. Resampling needs intensive computation which can be largely offset by today's powerful and inexpensive computing facility. Resampling of IUT is based on resampling of individual tests which can be conveniently performed by either bootstrap or permutation. Bootstrap and permutation were discussed in many literatures [12,27,28].

**Conclusion**

Our study demonstrated that the current unadjusted IUT approaches were overly conservative, which resulted in loss of power in finding overlapping genes in microarray-based gene expression studies. Our proposed RIUT was analytically proved to be a more powerful and less conservative approach than the current unadjusted IUT. The power improvement is more apparent in tests with weak and moderate effect sizes. This is also demonstrated in Monte-Carlo simulations and real case studies. In addition, certain known biologically relevant pathways were identified using the RIUT-derived overlapping genes which were not detected by using the traditional BIUT.

**Appendix**

Let  $X$  denote the random vector of data values. Suppose the probability distribution of  $X$  depends on an unknown parameter  $\theta$ . The set of possible values for  $\theta$  will be denoted by  $\Theta$ . Suppose we have  $m$  individual tests and let  $R_i$  denote a rejection region for a level- $\alpha$  test of  $H_{0i}: \theta \in \Theta_i$  versus  $H_{Ai}: \theta \in \Theta_i^c, 1 \leq i \leq m$ , where  $\Theta_i$  is a specified subset of  $\Theta$  and  $\Theta_i^c$  is its complement. Then IUT tests the union of sets against an intersection of sets.  $H_0: \mathcal{Q} \in \Theta_0 = \bigcup_{i=1}^m \Theta_i$  versus  $H_A: \mathcal{Q} \in \Theta_0^c = \bigcap_{i=1}^m \Theta_i^c$ , with the rejection region  $R = \bigcap_{i=1}^m R_i$ . In other words, the IUT rejects only if all of the tests reject.

**Berger's Theorem:** IUT with rejection region  $R$  is a level- $\alpha$  test of  $H_0$  versus  $H_A$ .

**Table 5: KEGG pathways Identified and the enrichment p-values**

KEGG Pathways	Fisher's Exact Test P-Value
Metabolism of xenobiotics by cytochrome P450	0.004
MAPK signaling pathway	0.028
Porphyrin and chlorophyll metabolism	0.030

(Fisher's Exact Test) based on the gene sets derived from RIUT and BIUT

*Proof.* For any  $\theta \in \Theta_0$  and for any  $1 \leq l \leq m$ , we have  $P_{\alpha}(R) = P_{\alpha}(\bigcap_{i=1}^m R_i) \leq P_{\alpha}(R_l) \leq \alpha$ . Therefore, is IUT is a level- $\alpha$  test.  $\square$

**Methods**  
**IUT and UIT**

Suppose a gene on which a number of  $\alpha$ -level hypothesis tests were performed, represented as  $t_1, t_2, \dots, t_m$ , where  $m$  is the number of individual tests. Each test  $t_i$  tests the null hypothesis  $H_{0i}$  versus alternative hypothesis  $H_{Ai}$ . We can combine all tests into a Union-Intersection Test (UIT) which rejects if *any* of the  $t_i$  rejects. We can also combine all tests into an Intersection-Union Test (IUT) which rejects if *all* the  $t_i$  reject. The UIT tests the hypothesis  $H_0 = \{\text{all } H_{0i} \text{ are true}\}$  against  $H_A = \{\text{at least one } H_{0i} \text{ is false}\}$  and the IUT tests the hypothesis  $H_0 = \{\text{at least one } H_{0i} \text{ is true}\}$  against  $H_A = \{\text{all } H_{0i} \text{ are false}\}$ . IUT and UIT were named from the fact that their null and alternative hypothesis can be described by set intersections and unions. (see Appendix for details).

The FWER of UIT is defined as  $\alpha' = \Pr(\text{Reject at least one } H_{0i} \mid \text{all } H_{0i} \text{ are true})$ . It is well known that in general  $\alpha' \neq \alpha$ . For example, if  $\alpha = 0.05$  and  $m = 5$ ,  $\alpha'$  would be about 0.23 when all individual tests are independent. Therefore there is a need to adjust  $\alpha'$  for IUT and there exist many procedures to do so, ranging from simple Bonferroni correction to computer-intensive resampling-based correction [12,13]. The FWER IUT is  $\alpha' = \Pr(\text{Reject all } H_{0i} \mid \text{at least one } H_{0i} \text{ is true})$ . It is also obvious that  $\alpha' \neq \alpha$  for IUT in general. However unlike the well studied UIT, there is no known procedure for adjusting the FWER  $\alpha'$  for IUT. The unadjusted IUT, also known as the Berger's approach, denoted as BIUT, suggests that the overall unadjusted p-value for IUT is

$$p = \max p_i, 1 \leq i \leq m, \tag{1}$$

where  $p_i$  is the p-value for individual tests  $t_i$ . Berger proved [10] that the unadjusted IUT is a level- $\alpha$  test if all  $t_i$  are level- $\alpha$  tests. Berger's also showed that the above IUT is a size- $\alpha$  test under certain trivial case such as the case when exactly one  $H_{0i}$  is true while all the other  $H_{0i}$  are false. However, the unadjusted approach is not a size- $\alpha$  test in general. For example, when considering two independent

individual tests  $t_1$  and  $t_2$ , the chance of rejecting both hypotheses is  $\alpha^2$  rather than  $\alpha$  if both  $H_{01}$  and  $H_{02}$  are true. Nonetheless, due to its simplicity, the unadjusted IUT approach was implicitly adopted by current microarray studies when overlapping genes were taken from several significant gene lists. This BIUT is equivalent to the Venn diagram in obtaining overlapping genes from multiple significant gene lists.

**Exact solution for IUT consisting of independent tests**

Berger's approach can be very conservative and therefore substantial power could be lost for detecting overlapping genes. Given the observed p-values  $(p_1, p_2, \dots, p_m)$  for the  $m$  individual tests, we are interested in estimate p-value for the entire test  $H_0$  of IUT. Here, we define a Westfall-Young-style p-value [12] for IUT denoted as  $p'$ ,

$$p' = \Pr(\text{all } P_j \leq p \mid H_0) = \Pr(\max P_j \leq \max p_i \mid H_0), 1 \leq i, j \leq m, \tag{2}$$

where  $P_j$  denote the distribution for the p-value of the  $j$ th hypothesis under null hypothesis for IUT  $H_0 = \{\text{at least one } H_{0i} \text{ is true}\}$ . The least significant p-value  $p$  is an observed statistic and the random variable  $\max P_j$  is the test statistic under  $H_0$ . This definition is intuitive as p-value measures the probability of false positive under null hypothesis, where a false positive of IUT means that all the p-values under  $H_0$  are less than the observed  $p$ . Similar to the above equation, we have the following relationship between the IUT FWER  $\alpha'$ , and individual test type 1 error rate  $\alpha$ .

$$\alpha' = \Pr(\text{all } P_j \leq \alpha \mid H_0) 1 \leq i, j \leq m \tag{3}$$

Unlike UIT, the null hypothesis  $H_0$  of IUT is a composite hypothesis and contains nuisance parameters. However under certain conditions, it is possible to derive the analytical solution for  $\alpha'$ . We obtain the following theorems:

**Theorem 1**

For an IUT constructed from two independent tests  $t_1: H_{01}$  vs.  $H_{A1}$ , and  $t_2: H_{02}$  vs.  $H_{A2}$ , both at significance level  $\alpha$ , suppose the observed p-values for  $t_1$  and  $t_2$  are uniformly distributed under  $H_{01}$  and  $H_{02}$  respectively, the exact FWER  $\alpha'$  for the IUT is controlled at

$$a' = \frac{(1-p_1)p_2(1-b_2)a + (1-p_2)p_1(1-b_1)a + (1-p_1)(1-p_2)a^2}{1-p_1p_2} \tag{4}$$

where the true probabilities of alternative hypotheses are  $\Pr(H_{A1}) = \pi_1$  and  $\Pr(H_{A2}) = \pi_2$ ; the type 2 error rate for  $t_1$  and  $t_2$  are  $\beta_1$  and  $\beta_2$ .

*Proof:*

$$\begin{aligned} a' &= \Pr(\max(P_1, P_2) \leq a | H_0) \\ &= \Pr(P_1 \leq a, P_2 \leq a | H_0) \\ &= \frac{\Pr(P_1 \leq a, P_2 \leq a, H_0)}{\Pr(H_0)} \\ &= \frac{\Pr(P_1 \leq a, P_2 \leq a, (H_{01} \text{ or } H_{02}))}{\Pr(H_{01} \text{ or } H_{02})} \\ &= \frac{\Pr(P_1 \leq a, P_2 \leq a, H_{01}, H_{A2}) + \Pr(P_1 \leq a, P_2 \leq a, H_{A1}, H_{02})}{\Pr(H_{01} \text{ or } H_{02})} \quad \text{independence} \\ &= \frac{\Pr(P_1 \leq a, P_2 \leq a, H_{01}, H_{02})}{\Pr(H_{01} \text{ or } H_{02})} \\ &= \frac{(1-p_1)p_2(1-b_2)a + (1-p_2)p_1(1-b_1)a + (1-p_1)(1-p_2)a^2}{1-p_1p_2} \quad \text{uniform} \end{aligned}$$

é

Recall that the p-value is the lowest level of significance at which the null hypothesis could have been rejected. We can obtain the adjusted p-value based on the observed p-values of the individual tests

$$p' = \frac{(1-p_1)p_2(1-b_2)p + (1-p_2)p_1(1-b_1)p + (1-p_1)(1-p_2)p^2}{1-p_1p_2} \tag{5}$$

where  $p = \max(p_1, p_2)$ .

**Theorem 2**

The RIUT procedure is universally at least as powerful as the unadjusted IUT, such that  $p' \leq p$ .

*Proof:*

$$\begin{aligned} p' &\leq \frac{(1-p_1)p_2p + (1-p_2)p_1p + (1-p_1)(1-p_2)p^2}{1-p_1p_2} \\ &\leq \frac{(1-p_1)p_2 + (1-p_2)p_1 + (1-p_1)(1-p_2)}{1-p_1p_2} p \\ &\leq p \end{aligned}$$

é

The above derivation can be visualized in Figure 3, which shows the partition of outcome space  $(p_1, p_2)$  of the two independent tests.

The relationship between  $\alpha'$  and  $\alpha$  is much more complicated in the IUT than that in the UIT. To apply the above theorem, the nuisance parameters  $(\pi_1, \pi_2, \beta_1, \beta_2)$  need to be estimated. Since there are usually thousands of genes available for each statistical test, we can obtain crude and conservative estimates of the parameters according to [14].

$$\hat{p}_i(l) = \frac{\#\{p_i(j) < l\}}{(l-1)n} \quad i = 1, 2, \quad j = 1, \dots, n, \quad \text{and} \quad \hat{b}_1 = \hat{b}_2 = 0, \tag{6}$$

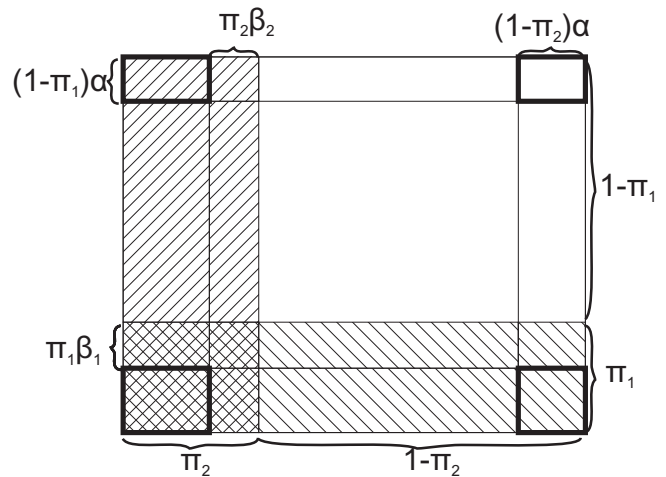
where  $n$  is the total number of genes,  $l$  is a chosen fixed value at 0.25 and  $p_i(j)$  represents the observed p-value for the  $i$ th test on the  $j$ th gene.

**Competing interests**

The authors declare that they have no competing interests.

**Authors' contributions**

XD and CW conceived of the study. XD drafted the manuscript and implemented the algorithm and performed most of the analysis. JX helped in data acquisition and results interpretation. All authors revised and approved the paper.



**Figure 3**  
**Partition of the sample space of the two dimensional outcome space defined by  $p_1$  and  $p_2$ .** The shaded two areas are the true non-null areas for the two individual tests respectively. The rejected regions by both tests are highlighted in bold rectangles at the four corners of the outcome space. Only the lower-left corner is the non-null region from both tests and therefore should be rejected by IUT. The other three corners are type I error region for IUT because at least one individual test is making type I error.

## Acknowledgements

This article has been published as part of *BMC Bioinformatics* Volume 9 Supplement 6, 2008: Symposium of Computations in Bioinformatics and Bio-science (SCBB07). The full contents of the supplement are available online at <http://www.biomedcentral.com/1471-2105/9?issue=S6>.

## References

- Wang C, Chelly MR, Chai N, Tan Y, Hui T, Li H, et al.: **Transcriptomic fingerprinting of bone marrow-derived hepatic beta2m-/Thy-1+ stem cells.** *Biochem Biophys Res Commun* 2005, **327**:252-260.
- Guo L, Lobenhofer EK, Wang C, Shippy R, Harris SC, Zhang L, et al.: **Rat toxicogenomic study reveals analytical consistency across microarray platforms.** *Nat Biotechnol* 2006, **24**:1162-1169.
- Tan Y, Shi L, Hussain SM, Xu J, Tong W, Frazier JM, et al.: **Integrating time-course microarray gene expression profiles with cytotoxicity for identification of biomarkers in primary rat hepatocytes exposed to cadmium.** *Bioinformatics* 2006, **22**:77-87.
- Lim DA, Suarez-Farinas M, Naef F, Hacker CR, Menn B, Takebayashi H, et al.: **In vivo transcriptional profile analysis reveals RNA splicing and chromatin remodeling as prominent processes for adult neurogenesis.** *Mol Cell Neurosci* 2006, **31**:131-148.
- Kooperberg C, Aragaki A, Strand AD, Olson JM: **Significance testing for small microarray experiments.** *Stat Med* 2005, **24**:2281-2298.
- Wang A, Gehan EA: **Gene selection for microarray data analysis using principal component analysis.** *Stat Med* 2005, **24**:2069-2087.
- Baldi P, Long AD: **A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes.** *Bioinformatics* 2001, **17**:509-519.
- Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
- Berger RL, Hsu JC: **Bioequivalence trials, intersection-union tests, and equivalence confidence sets.** *Statistical Science* 1996, **11**:283-319.
- Berger RL: **Multiparameter hypothesis testing and acceptance sampling.** *Technometrics* 1982, **24**:295-300.
- Allison DB, Cui X, Page GP, Sabripour M: **Microarray data analysis: from disarray to consolidation and consensus.** *Nat Rev Genet* 2006, **7**:55-65.
- Westfall PH, Young SS: *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* New York: Wiley-Interscience; 1993.
- Miller RGJ: *Simultaneous Statistical Inference* New York: Springer-Verlag; 1991.
- Storey J: **A direct approach to false discovery rates.** *J R Statist Soc B* 2002, **64**:479-498.
- Fisher RA: *Statistical Methods for Research Workers* Edinburgh: Oliver and Boyd; 2007.
- Hedges LV, Olkin I: *Statistical Methods for Meta-analysis* San Diego, California, USA: Academic Press Inc, Harcourt Brace Jovanovich Publishers; 1985.
- Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, et al.: **The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.** *Nat Biotechnol* 2006, **24**:1151-1161.
- Tan Y, Shi L, Tong W, Wang C: **Multi-class cancer classification by total principal component regression (TPCR) using microarray gene expression data 5.** *Nucleic Acids Res* 2005, **33**:56-65.
- Kanehisa M, Goto S, Hattori M, Iki-Kinoshita KF, Itoh M, Kawashima S, et al.: **From genomics to chemical genomics: new developments in KEGG.** *Nucleic Acids Res* 2006, **34**:D354-D357.
- Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**:27-30.
- Kanehisa M: **A database for post-genome analysis.** *Trends Genet* 1997, **13**:375-376.
- Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4**(5):P3.
- Bozcaarmutlu A, Arinc E: **Effect of mercury, cadmium, nickel, chromium and zinc on kinetic properties of NADPH-cytochrome P450 reductase purified from leaping mullet (*Liza saliens*).** *Toxicol In Vitro* 2007, **21**:408-416.
- Plewka A, Plewka D, Nowaczyk G, Brzoska MM, Kaminski M, Moniuszko-Jakoniuk J: **Effects of chronic exposure to cadmium on renal cytochrome P450-dependent monooxygenase system in rats.** *Arch Toxicol* 2004, **78**:194-200.
- Zaccaro MC, Salazar C, Zulpa dC, Storni dC, Stella AM: **Lead toxicity in cyanobacterial porphyrin metabolism.** *Environ Toxicol* 2001, **16**:61-67.
- Komatsu M, Furukawa T, Ikeda R, Takumi S, Nong Q, Aoyama K, et al.: **Involvement of mitogen-activated protein kinase signaling pathways in microcystin-LR-induced apoptosis after its selective uptake mediated by OATP1B1 and OATP1B3.** *Toxicol Sci* 2007, **97**:407-416.
- Efron B, Tibshirani R: *An introduction to the bootstrap* New York: Chapman & Hall; 1993.
- Dudoit S, Yang Y, Matthew J, Speed TP: **Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments.** 2000.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

