

Binary Contingency Table Method for Analyzing Gene Mutation in Cancer Genome

Emi Ayada¹, Atsushi Niida¹, Takanori Hasegawa², Satoru Miyano¹,
and Seiya Imoto¹(✉),

¹ Human Genome Center, The Institute of Medical Science,
The University of Tokyo 4-6-1 Shirokanedai, Minato-ku, Tokyo, Japan
emi.ayada@gmail.com, {aniida,miyano,imoto}@ims.u-tokyo.ac.jp

² Bioinformatics Center, Institute for Chemical Research,
Kyoto University Gokasyo Uji, Kyoto, Japan
t-hasegw@kuicr.kyoto-u.ac.jp

Abstract. Gene mutations are responsible for a large proportion of genetic diseases such as cancer. Hence, a number of computational methods have been developed to find loci subject to frequent mutations in cancer cells. Since normal cells turn into cancer cells through the accumulation of gene mutations, the elucidation of interactive relationships among loci has great potential to reveal the cause of cancer progression; however, only a few methods have been proposed for measuring statistical significance of pairs of loci that are co-mutated or exclusively mutated. In this study, we proposed a novel statistical method to find such significantly interactive pairs of loci by employing the framework of binary contingency tables. Using Markov chain Monte Carlo procedure, the statistical significance is evaluated by sampling null matrices whose marginal sums are equal to those of the input matrix. We applied the proposed method to mutation data of colon cancer patients and successfully obtained significant pairs of loci.

Keywords: Cancer · Gene mutation · Binary contingency tables · Markov chain Monte Carlo

1 Introduction

Gene mutations can change the normal function of proteins, leading to genetic diseases such as cancer. For example, the mutation can result in loss of function that helps to repair damaged DNA. In cancer research, mutations observed only in tumor cells have been intensively investigated, however, these mutations are analyzed independently and combinations of mutations are not well studied. Since the accumulation of gene mutations causes cancer, it is important to detect pairs of genes, which contribute to the accumulation interactively, *e.g.*, genes with mutations that tend to occur together among a lot of samples. The elucidation of these relationships allows us to identify how genes interact with other genes. To achieve it, some studies[3,14,9] concerned the identification of

the relationships between genes in cancer cells based on the fact that oncogenesis is a process with multiple stages, in which normal cells transform into cancer cells via multiple genetic mutations. However, almost these researches have not measured the statistical significance of the relationships. Although Dees et al. [3] considered statistical significance for identifying the relationships based on permutation test-based method, they did not consider the varieties of the numbers of the mutations in each of samples and genes.

Therefore, to provide an accurate statistical assessment, we propose a novel method to measure the statistical significance of relationships between genes using a statistical framework of binary contingency tables (BCTs), which are defined as binary tables with fixed column and row sums. BCTs are utilized as tables that are composed of entries with binary values indicating absence (0) or presence (1) of mutations. In this study, we analyze the following two types of gene relationships: co-mutated relationships, which represent pairs of genes getting mutated together and exclusive relationships, which represent pairs of genes including both a mutated gene and a gene without a mutation. Since the numbers of mutations vary among samples and the numbers of mutated samples vary among genes, we propose random sampling of BCTs keeping the sums of mutations on each samples and genes using Markov chain Monte Carlo procedure. By using gene mutation data as a BCT, the proposed method measures the statistical significance of an observed state of each gene pair by the algorithm developed by Bezáková [1]. This algorithm samples random BCTs and non-BCTs satisfying almost marginal sums of the original data and obtains p -values for all combinations of gene pairs; however, an accurate statistical test cannot be performed when sampling non-BCTs. In order to sample BCTs only, we further proposed a novel algorithm, termed Perfect BCT(PBCT)-sampling, that samples BCTs under the restrictions of the number of mutations occurring on each sample and gene, and measures the statistical significance of the relationships between genes in the mutation data.

To show the effectiveness of the proposed method, we compared the performance of our proposed method, the proposed method using BCT-sampling and an existing method (Fisher’s exact test) through a simulation study. As a result, our method outperformed other methods and we confirmed the advantage of using BCTs. Next, we analyzed gene mutation data downloaded from The Broad Institute (<http://gdac.broadinstitute.org/>) using our proposed method and obtained significantly co-mutated and exclusively mutated gene pairs. We confirmed that the result of the analysis by our proposed method contains pairs of genes, which have been thought as genes related to cancer.

2 Method

2.1 Binary Contingency Tables

BCTs are typically used to represent two exclusive events, such as “absence” or “presence” by 0 and 1, respectively. We apply the framework of BCTs to a $m \times n$ binary table B containing the presence of mutations for pairs of genes and

samples in order to clarify the relationships between genes in cancer cells using gene mutation data. In the table B , rows and columns respectively correspond to genes and samples, and the entry at the i th row and j th column is set to 1 if the i th gene of the j th sample is mutated, and 0 otherwise. From the framework of BCTs, we analyze the two types of gene relationships, co-mutated relationships and exclusive relationships.

2.2 BCT-Sampling

From the number of mutations that occur on each column and row, we measure the statistical significance of gene pairs such as co-mutated gene pairs or gene pairs with exclusive mutations. For statistical testing, we first build the null distribution by sampling null matrices, and then measure the statistical significance in terms of the number of samples for each co-mutated gene pair or for gene pair with exclusive mutations in the original binary table B . For sampling null matrices, we adopted BCT-sampling proposed by Bezáková [1], which samples binary tables keeping the given marginal sums. Let define perfect and near-perfect tables as tables satisfying completely the marginal sums of B and tables with one row and one column sum decreased by 1, respectively. Starting from a perfect table, the above method recursively samples a table from the previously sampled table as follows:

1. Let (i, j) be a pair of indices for indicating the i th row and j th column of a table or matrix. If the current $m \times n$ table T is a perfect table, randomly select a pair of indices from $\{(i, j) | T_{i,j} = 1\}$, where $T_{i,j}$ is the i th row and j th column of T , and set $T_{i,j}$ to 0. A near-perfect table with marginal sums of rows $r_1, \dots, r_i - 1, \dots, r_m$ and columns $c_1, \dots, c_j - 1, \dots, c_n$ and a deficiency at (i, j) is obtained.
2. If the current $m \times n$ table T is a near-perfect table with a deficiency at (i, j) , randomly select a pair of indices in $\{(i, j)\} \cap \{(k, l) | T_{k,l} = 1\}$. If $(k, l) = (i, j)$ holds, then set $T_{i,j}$ to 1 and a perfect table is obtained. Otherwise, select one of the following two procedures randomly:
 - (a) If $T_{k,j} = 0$ holds, then set $T_{k,j}$ to 1 and $T_{k,l}$ to 0 (the deficiency moves from (i, j) to (i, l)).
 - (b) If $T_{i,l} = 0$ holds, then set $T_{i,l}$ to 1 and $T_{k,l}$ to 0 (the deficiency moves from (i, l) to (k, l)).

These sampled tables from the above steps contain both near-perfect and perfect tables.

2.3 PBCT-Sampling

In order to keep precisely the condition of the original binary table B , we developed PBCT-sampling by extracting only perfect tables from all the samples tables with BCT-sampling. Sampling different perfect tables enables us to calculate precisely the frequencies of that gene pairs with the number of mutated

samples, which happen together or exclusively, under the given conditions. We design PBCT-sampling to check if a newly generated perfect table is different from the previously sampled tables in order to obtain various different perfect tables. Algorithm 1 shows the detail of PBCT-sampling.

Algorithm 1. PBCT-sampling

M : the number of perfect tables to be sampled
 f : the function of sampling BCTs proposed by Bezáková
 X_i : i th sampled binary table from f
 \mathcal{X} : a set of sampled binary tables
Set $\mathcal{X} = \emptyset$
Set $n = 1$
Set $count = 0$
while $count < M$ **do**
 $X_n = f(X_{n-1})$
 if X_n is a Perfect Table and $X_n \neq X_{count}$ **then**
 Put X_n to \mathcal{X}
 $count = count + 1$
 $n = n + 1$

2.4 Computation of p -value

For the pair of i th and j th genes in B , the test statistic for the detection of co-mutated gene pairs is defined by

$$\sum_{k=1}^{n_s} B_{ik} \times B_{jk}, \quad (1)$$

and that for the exclusively mutated gene pairs is defined by

$$\sum_{k=1}^{n_s} I(k), \quad I(k) = \begin{cases} 1 & \text{if } B_{ik} + B_{jk} = 1 \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

where n_s is the number of samples. Here, we calculate p -value referring to the way adopted by CDCOCA [10]. Let T_{ij} be the test statistic for the i th and j th gene pair in a sampled table, O_{ij} be the test statistic for that in B , C_{ij} be the total number of the test statistics that satisfy $O_{ij} > T_{ij}$ and M be the total number of sampled matrices. The algorithm for the calculation of p -value is shown in Algorithm 2.

3 Result

3.1 Simulation Study

In the simulation study, we compared our proposed method using PBCT-sampling with the method using BCT-sampling and an alternative statistical method,

Algorithm 2. Computation of p -value

```

 $M$ : the number of perfect tables to be sampled
 $n_g$ : the number of genes
 $p_{ij}$ :  $p$ -value for gene pair,  $i$  and  $j$ 
Set  $count = 0$ 
Set  $C = 0$ 
while  $count < M$  do
  Sample binary tables  $X$ 
  for  $i < n_g$  do
    for  $j < n_g$  do
      calculate  $T_{ij}$  for  $X$ 
      if  $T_{i,j} \geq O_{ij}$  then
         $C_{ij} = C_{ij} + 1$ 
     $count = count + 1$ 
 $p_{ij} = C_{ij}/M$ 

```

Fisher's exact test [5,8]. The Fisher's exact test in this context calculates the probability of obtaining the observed state under the given conditions as follows. Let a be the number of samples without any mutations on both gene x and y , b be the number of samples with mutations on only gene y and not on gene x , c be the number of samples with mutations on only gene x and not on gene y and d be the number of samples with mutations on both genes x and y . Then the probability of occurring such event is calculated by

$$p_{xy} = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}}, \quad (3)$$

where n is the number of all samples, which means n is identical to the sum of a, b, c and d . Table 3.1 shows the relationship between these letters and genes.

Table 3.1. The summary of the relationships between two genes, x and y . Fisher's exact test calculates the probability of obtaining the tables using a, b, c and d .

		Gene y	
		Not mutated	Mutated
Gene x	Not mutated	a	b
	Mutated	c	d

We measure the performance of the methods by area under the precision recall curve (AUC) [2]. The performance is shown by the AUC score, which is the space under the curve plotted according to the precision and recall of p -values at each threshold. Fig. 3.1 shows an example of obtained p -values and AUC on our simulation study. The precision is defined as the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. The recall is defined as the ratio of the number of relevant records retrieved to the total number of relevant records.

To apply the method toward two problems, *i.e.*, co-mutated and exclusively mutated problems, we assume that simulated data have two pairs of co-mutated genes or exclusively mutated genes, as statistical significant gene pairs, respectively. We also suppose that the numbers of mutations vary among samples and genes. Then, we prepared simulation data as follows:

1. Generate a mutation rate for each column and row, r_i and r_j ($0 < r_i, r_j < 1$), which controls how often mutations occur on entities.
2. Set the mutation rate for each entry $r_{i,j} = r_i r_j$ based on the mutation rate of columns and rows.
3. Set noise rate that regulates the amount of mutations in the simulated mutation matrix. If the noise rate is 4, for example, we control the amount of entries with mutations to be approximately 40% of all entries. The rate is from 0 to 10.
4. Insert two pairs that have interactive relationships, which are co-mutated or exclusively mutated gene pairs, and are named as true mutation pairs. We define the number of co-mutated and exclusively mutated samples as “signal”. In the co-mutated problem, when the signal is 4, two true mutation pairs have four samples, where mutations happen in both of paired genes. In the exclusively mutated problems, when the signal is 4, two true mutation pairs have four samples, where mutations happen in either of paired genes. For each true pair, we choose these samples at random.

Figs. 3.2(a)-3.2(c) show background data, simulated data with two true co-mutated pairs and simulated data with two true exclusively mutated pairs, respectively. We had a simulation using data, which is composed of 50 samples and 50 genes, and several kinds of signal and noise.

Fig. 3.3 shows the result of the co-mutated problem. Consequently, BCT-sampling has a higher AUC score than other methods in case of small noise and large signal; however, PBCT-sampling demonstrates stably superior performance in any parameter compared to other two methods. We also confirmed the advantage on employing the BCT framework to the exclusively mutated problem as concluded in Table 3.4. The result shows that our proposed method has higher AUC scores than other methods and detects true mutation pairs on both co-mutated and exclusively mutated problems.

We can consider that the simulation results of our proposed method are better than others because it takes the number of mutations on each gene and sample of the mutation data into account. In contrast, Fisher’s exact test utilizes only the number of mutations happening on each gene. Also, we can see that PBCT-sampling had better performance compared with BCT-sampling since PBCT-sampling keeps completely the marginal sums and it utilizes more accurate conditions of the original mutation data.

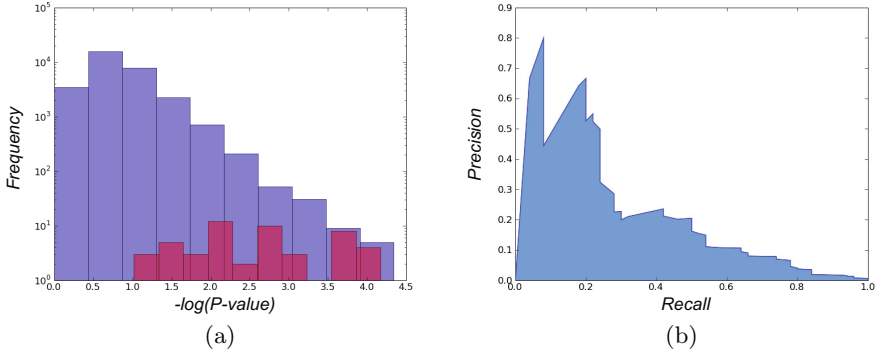


Fig. 3.1. The example for the calculation of AUC. We use a dataset of twenty simulated data to measure the performance of methods. Fig. 3.1(a) plots the number of obtained p -values through a simulation, and the red and blue bars are the number of p -values for gene pairs with true mutations and other gene pairs, respectively. Blue area on Fig. 3.1(b), which consists precision and recall axes, shows AUC of data plotted on Fig. 3.1(a).

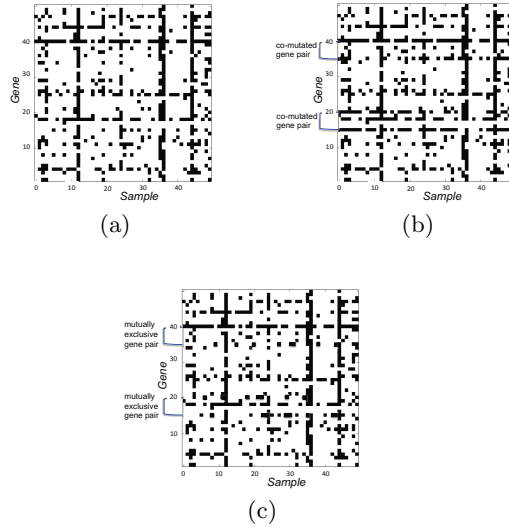


Fig. 3.2. A left figure shows background data with 20% noise and no signal. Figs. 3.2(b) and 3.2(c) are simulated data containing two true co-mutated and exclusively pairs, respectively. Black in the figure represents a gene mutation. These data are 50×50 size matrices, whose vertical and horizontal axes are genes and samples, respectively.

	Fisher test	BCT-sampling			PBCT-sampling		
		30000	300000	3000000	30000	300000	3000000
Signal15 noise3	0.349	0.048	0.148	0.223	0.178	0.250	0.471
Signal10 noise3	0.070	0.030	0.207	0.192	0.211	0.211	0.222
Signal7 noise3	0.028	0.010	0.019	0.020	0.020	0.019	0.079
Signal15 noise2	0.733	0.197	0.514	0.915	0.671	0.835	0.877
Signal10 noise2	0.537	0.288	0.376	0.553	0.460	0.545	0.561
Signal7 noise2	0.222	0.130	0.225	0.179	0.187	0.263	0.299

Fig. 3.3. The simulation results on 50×50 size matrices with two true co-mutated pairs. Shown values are AUC scores, and methods with high scores are thought to have high performance of assessing the relationships between genes. We tested Fisher's exact test, BCT-sampling and PBCT-sampling with several kinds of signal and noise.

	Fisher test	BCT-sampling			PBCT-sampling		
		5000	50000	500000	5000	50000	500000
Signal15 noise3	0.006	0.009	0.072	0.151	0.098	0.171	0.199
Signal10 noise3	0.002	0.008	0.033	0.041	0.042	0.055	0.084
Signal7 noise3	0.002	0.008	0.008	0.015	0.011	0.015	0.028
Signal15 noise2	0.007	0.023	0.152	0.177	0.139	0.193	0.242
Signal10 noise2	0.003	0.018	0.037	0.041	0.036	0.045	0.042
Signal7 noise2	0.002	0.006	0.011	0.012	0.020	0.013	0.013

Fig. 3.4. The simulation results on 50×50 size matrices with two true exclusively mutated pairs. Shown values are AUC scores, and methods with high scores are thought to have high performance of assessing the relationships between genes. We tested Fisher's exact test, BCT-sampling and PBCT-sampling with several kinds of signal and noise.

Table 3.2. Statistically significant co-mutated gene pairs obtained by PBCT-sampling

Gene	Gene	p -value	q -value
PKHD1L1	RFC1	6.6×10^{-9}	4.5×10^{-6}
LAMA3	DOCK10	6.6×10^{-9}	4.5×10^{-6}
EPB41L3	DOCK5	6.6×10^{-9}	4.5×10^{-6}
EPB41L3	TRPM2	6.6×10^{-9}	4.5×10^{-6}
EPB41L3	SLIT1	6.6×10^{-9}	4.5×10^{-6}
ADAM7	DNAH1	6.6×10^{-9}	4.5×10^{-6}
ADAM7	CACNA1C	6.6×10^{-9}	4.5×10^{-6}
ADAM7	CATSPERB	6.6×10^{-9}	4.5×10^{-6}

3.2 Real Data Experiment

We used binary gene mutation data of colorectal adenocarcinoma, which is downloaded from The Broad Institute. We sorted 631 samples according to the number of mutations and retrieved top 155 samples. The data contains 699 genes and 155 samples, whose rows and columns correspond to genes and samples, respectively. We analyzed the data with PBCT-sampling to identify statistically significant gene pairs. Consequently, we obtained eight co-mutated gene pairs with 10,000,000 samplings as concluded in Tables 3.2. For the exclusively mutated

Table 3.3. Statistically significant exclusively mutated gene pairs obtained by PBCT-sampling

Gene	Gene	p -value	q -value
TP53	XIRP2	5.0×10^{-8}	3.3×10^{-6}
BRAF	DSCAM	5.0×10^{-8}	3.3×10^{-6}
ACVR1B	POSTN	5.0×10^{-9}	3.3×10^{-6}
SPHKAP	VAV1	5.0×10^{-9}	3.3×10^{-6}
TEX15	VAV1	5.0×10^{-9}	3.3×10^{-6}
LAMA3	CELSR1	5.0×10^{-9}	3.3×10^{-6}
KCNQ3	EPB41L3	5.0×10^{-9}	3.3×10^{-6}
SEMA4D	VAV1	5.0×10^{-9}	3.3×10^{-6}
PTPRT	VAV1	5.0×10^{-9}	3.3×10^{-6}
DOCK5	DSCAM	5.0×10^{-9}	3.3×10^{-6}
TMEM132B	VAV1	5.0×10^{-9}	3.3×10^{-6}
PXDNL	DSCAM	5.0×10^{-9}	3.3×10^{-6}

problem, we sampled 2,000,000 BCTs and obtained twelve gene pairs, which are listed as significant exclusively mutated pairs shown in Table 3.3. These results of both kinds, co-mutated and exclusively mutated gene pairs, show that gene pairs of each kind have extremely small and the same p -value and q -value [13]. This is because there was no sampled BCT with greater test statistics than those of the original mutation data at the significant gene pairs.

In the obtained results, we focus on one of significant co-mutated pairs, LAMA3 and DOCK10, as illustrated in Fig. 3.5. The figure shows the frequencies that mutations appear on each gene and sample. We can observe that DOCK10 and LAMA3 have twelve and eight mutations, respectively, and the pair has 7 co-mutated samples. Some of them are recurrently mutated, but others are not likely to be mutated. Our proposed method utilizes these frequencies of mutations on each sample and gene, and detects gene pairs with co-mutated samples including infrequently mutated sample, such as LAMA3 and DOCK10.

In addition, Fig. 3.6 shows the frequencies and locations of mutations on gene-pair, TEX15 and VAV1, which is obtained as the statistically significant and exclusively mutated gene pair. We can confirm that their mutations happens in mutually exclusive way.

LAMA3 [12] and DOCK10 [6], which are obtained as a co-mutated gene pair, are considered to be involved in tumor cell invasion and progression. Additionally, among exclusively mutated gene pairs listed on Table 3.3, TP53 is widely know as tumor suppressor gene and XIRP2 was suggested as a potential driver gene in melanoma [7]. VAV1 [4] and TEX15 [11] are also identified as the candidates of genes related to cancer. VAV1 has been thought as a gene, which is associated with decreased survival and contributes to the tumorigenic properties of pancreatic cancer cells. Also, TEX15 was observed in a significant fraction of tumor samples different histological types.

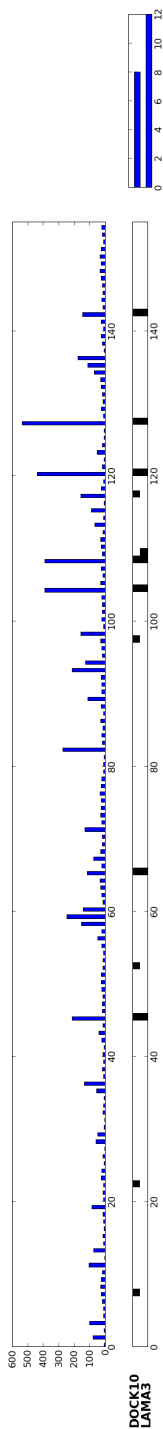


Fig. 3.5. The summary of gene mutations on LAMA3 and DOCK10. A horizontal axis represents samples. Black points and blue bars along the horizontal axis in the figure show gene mutations and the total number of gene mutations occurring on each sample, respectively. Two blue bars paralleled to the horizontal axis are the total number of samples with gene mutations on each gene. We can see these two genes have gene mutations at the same samples.

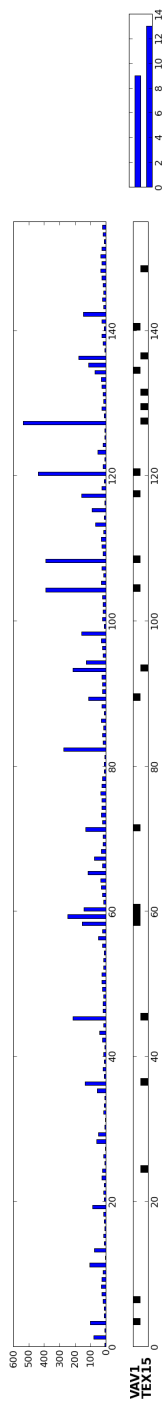


Fig. 3.6. The summary of gene mutations on TEX15 and VAV1. A horizontal axis represents samples. Black points and blue bars along the horizontal axis in the figure show gene mutations and the total number of gene mutations occurring on each sample, respectively. Two blue bars paralleled to the horizontal axis are the total number of samples with gene mutations on each gene. We can see that these two genes have mutation in mutually exclusive way.

4 Conclusion and Discussion

There exist many studies focusing on mutation genes and they play a key role in the field of cancer research. In this study, we aimed to elucidate the relationships of gene pairs using binary mutation data. The developed method, PBCT-sampling, enables us to utilize the frequencies that mutations occur on each gene and sample of the mutation data and assess the statistical significance of relationships between genes.

Through the simulation study, we prepared two synthetic data representing co-mutated and exclusively mutated problems and demonstrated that PBCT-sampling outperformed BCT-sampling and Fisher's exact test, which is without the BCTs framework. These results indicated that the BCT framework is reasonable for assessing the statistical significance of gene mutation data and PBCT-sampling, which uses only BCTs, is capable of performing more accurate assessment and has superior performance compared to BCT-sampling using both of BCTs and non-BCTs, and also with Fisher's exact test. Therefore, we confirmed the advantage of the BCT-framework, which allows us to sample binary tables keeping the the varieties of the marginal sums of the mutation data. Furthermore, the analysis of real data with PBCT-sampling showed the statistically significant co-mutated and exclusively mutated gene pairs. Since they have been indicated as cancer related genes, the performance of detecting significant pairs may be suggested. In these gene pairs, we confirmed that some of obtained gene pairs as exclusively mutated gene pairs comprise of tumor suppressor gene and driver gene and some of obtained gene pairs as co-mutated gene pairs are the combination of genes, which contribute to the progression of cancer.

In this study, we focused on the detection of interactive gene pairs but we further expect that our proposed method can be applied to the problem of detecting more than three genes with interactive relationships. Also, since the BCT framework is practical for the detection of relationships in binary data, we consider that we can analyze other binary data, such as copy number data, with PBCT-sampling.

References

1. Bezáková, I.: Sampling binary contingency tables. *Computing in Science & Engineering* 10(2), 26–31 (2008)
2. Davis, J., Goadrich, M.: The relationship between Precision-Recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*, pp. 233–240. ACM (2006)
3. Dees, N.D., Zhang, Q., Kandath, C., Wendl, M.C., Schierding, W., Koboldt, D.C., Mooney, T.B., Callaway, M.B., Dooling, D., Mardis, E.R., et al.: MuSiC: identifying mutational significance in cancer genomes. *Genome Research* 22(8), 1589–1598 (2012)
4. Fernandez-Zapico, M.E., Gonzalez-Paz, N.C., Weiss, E., Savoy, D.N., Molina, J.R., Fonseca, R., Smyrk, T.C., Chari, S.T., Urrutia, R., Billadeau, D.D.: Ectopic expression of VAV1 reveals an unexpected role in pancreatic cancer tumorigenesis. *Cancer Cell* 7, 39–49 (2005)

5. Fisher, R.A.: Statistical methods for research workers (1934)
6. Gadea, G., Sanz-Moreno, V., Self, A., Godi, A., Marshall, C.J.: DOCK10-mediated Cdc42 activation is necessary for amoeboid invasion of melanoma cells. *Current Biology* 18(19), 1456–1465 (2008)
7. Hayward, N.K.: Mutation spectrum of the first melanoma genome points finger firmly at ultraviolet light as the primary carcinogen. *Pigment Cell & Melanoma Research* 23(2), 153–154 (2010)
8. Irwin, J.: Tests of significance for differences between percentages based on small numbers. *Metron* 12(2), 84–94 (1935)
9. Klijn, C., Bot, J., Adams, D.J., Reinders, M., Wessels, L., Jonkers, J.: Identification of networks of co-occurring, tumor-related DNA copy number changes using a genome-wide scoring approach. *PLoS Computational Biology* 6(1), 1000631 (2010)
10. Kumar, N., Rehrauer, H., Ca, H., Baudis, M.: CDCOCA: a statistical method to define complexity dependent co-occurring chromosomal aberrations. *Genome Biology* 11(1), 23 (2010)
11. Lorient, A., Boon, T., De Smet, C.: Five new human cancer-germline genes identified among 12 genes expressed in spermatogonia. *International Journal of Cancer* 105(3), 371–376 (2003)
12. Sathyanarayana, U.G., Padar, A., Suzuki, M., Maruyama, R., Shigematsu, H., Hsieh, J.-T., Frenkel, E.P., Gazdar, A.F.: Aberrant promoter methylation of laminin-5-encoding genes in prostate cancers and its relationship to clinicopathological features. *Clinical Cancer Research* 9(17), 6395–6400 (2003)
13. Storey, J.D., Tibshirani, R.: Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 100(16), 9440–9445 (2003)
14. Vandin, F., Upfal, E., Raphael, B.J.: De novo discovery of mutated driver pathways in cancer. *Genome Research* 22(2), 375–385 (2012)

Bioinformatics Research and Applications
11th International Symposium, ISBRA 2015 Norfolk,
USA, June 7-10, 2015 Proceedings
Harrison, R.; Li, Y.; Măndoiu, I. (Eds.)
2015, XXII, 446 p. 163 illus., Softcover
ISBN: 978-3-319-19047-1