

Claims

1. ~~A computerized method for identifying one or more pairs of genes, selected from a set of N genes, which are statistically associated with prognosis of a potentially fatal medical condition,~~

the method employing test data which, for each subject k of a set of K^* subjects suffering from the medical condition, indicates (i) a survival time of subject k , and (ii) for each gene i , a corresponding gene expression value $y_{i,k}$ of subject k ;

the method comprising:

(i) forming a plurality of pairs of the identified genes (i, j with $i \neq j$), and for each pair of genes:

(a) partitioning the K^* subjects into two subsets according to whether $\log(y_{i,k}) - \log(y_{j,k})$ is respectively above or below a cut-off value $c_{i,j}$;

(b) computationally fitting the corresponding survival times of one of the subsets of the subjects to the Cox proportional hazard regression model, said fitting using a respective regression parameter $\beta_{i,j}$; and

(c) obtaining a significance value indicative of prognostic significance of the gene pair i, j ; and

(ii) identifying one or more of the pairs of genes i, j for which the corresponding significance values have the highest prognostic significance.

12. A computerized method for discovering one or more pair of genes, selected from a set of N genes, which are statistically associated with prognosis of a potentially fatal medical condition, the method including:

(i) for each of a plurality of sets of subjects, identifying a respective set of gene pairs from the set of N genes by a method employing test data which, for each subject k of a set of K^* subjects suffering from the medical condition, indicates (i) a survival time of subject k , and (ii) for each gene i , a corresponding gene expression value $y_{i,k}$ of subject k ; the method comprising:

(A) forming a plurality of pairs of the identified genes (i, j with $i \neq j$), and for each pair of genes:

(a) partitioning the K^* subjects into two subsets according to whether $\log(y_{i,k}) - \log(y_{j,k})$ is respectively above or below a cut-off value $c_{i,j}$;

(b) computationally fitting the corresponding survival times of one of the subsets of the subjects to the Cox proportional hazard regression model, said fitting using a respective regression parameter $\beta_{i,j}$; and

(c) obtaining a significance value indicative of prognostic significance of the gene pair i,j ; and

(B) identifying one or more of the pairs of genes i,j for which the corresponding significance values have the highest prognostic significance~~according to claim 1~~; and

(ii) discovering one or more gene pairs which are in common to the plurality of sets of gene pairs.

32. A computerized method according to claim 21 further including verifying the prognostic significance of the discovered gene pairs using data describing

the survival times and corresponding gene expression levels of a further set of subjects.

43. A computerized method according to claim 21 further including applying
5 a boosting algorithm to select one or more pairs of gene pairs from the gene pairs which are in common to the plurality of sets of gene pairs.

54. A computerized method according to claim 1 wherein the medical
10 condition is cancer, and the expression levels are from samples of respective tumours in the K^* subjects.

65. A method according to claim 54 wherein the medical condition is breast
cancer.

76. A method according to claim 1 in which the survival time is survival time
15 until mortality.

87. A method according to claim 1 in which the survival time is a survival
time without metastasis of a cancer.

98. A computerized method of obtaining information about a first subject in
20 relation to a potentially-fatal medical condition, the method including:

- (i) identifying, ~~by a method according to claim 1, from among a set of~~
25 ~~N genes,~~ one or more pairs of genes which are statistically associated with prognosis of the medical condition, from among a set of N genes, by a method employing test data which, for each subject k of a set of K^* subjects suffering from the medical condition, indicates (i) a survival time of subject k , and (ii) for each gene i , a corresponding gene expression value $y_{i,k}$ of subject k , the method comprising:
- 30

(A) forming a plurality of pairs of the identified genes (i, j with $i \neq j$), and for each pair of genes:

5 (a) partitioning the K^* subjects into two subsets according to whether $\log(y_{i,k}) - \log(y_{j,k})$ is respectively above or below a cut-off value $c_{i,j}$;

10 (b) computationally fitting the corresponding survival times of one of the subsets of the subjects to the Cox proportional hazard regression model, said fitting using a respective regression parameter $\beta_{i,j}$; and

(c) obtaining a significance value indicative of prognostic significance of the gene pair i,j ; and

15 (B) identifying one or more of the pairs of genes i,j for which the corresponding significance values have the highest prognostic significance, the one or more pairs of genes being statistically associated with prognosis of the medical condition; and;

20 (ii) for each of the one or more identified pairs of genes i, j obtaining corresponding gene expression values y_i and y_j of the first subject; and

 (iii) obtaining said information using a ratio of the obtained gene expression values.

25

~~109~~9. A computerized method according to claim ~~98~~98 in which said information is a prognosis for the first subject who is suffering from the medical condition, a susceptibility of the first subject to the medical condition, a prediction of the recurrence of the medical condition, or a recommended treatment for the

30 medical condition.

~~11. A method for obtaining information about a first subject in relation to breast cancer, the method including:~~

~~5 (a) for each of one or more pairs of genes i, j obtaining corresponding gene expression values y_i and y_j of the first subject; and~~

~~(b) obtaining said information using a ratio of the obtained gene expression values of the pair of genes;~~

~~10~~

~~the pairs of genes being selected from the group consisting of:~~

~~(i) SEQ ID NO:1 and SEQ ID NO:2;~~

~~15~~

~~(ii) SEQ ID NO:3 and SEQ ID NO:4 and~~

~~(iii) SEQ ID NO:5 and SEQ ID NO:6.~~

~~12. A kit, such as a microarray, for obtaining data for performing prognosis of breast cancer, the microarray being for measuring the expression value of a set of no more than 100 genes, and more preferably no more than 20 genes, including at least one of the following pairs of genes:~~

~~20~~

~~(i) SEQ ID NO:1 and SEQ ID NO:2;~~

~~25~~

~~(ii) SEQ ID NO:3 and SEQ ID NO:4; and~~

~~(iii) SEQ ID NO:5 and SEQ ID NO:6.~~