Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3038743 A1 2018/04/12

(21) 3 038 743

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) **Date de dépôt PCT/PCT Filing Date:** 2017/10/05
(87) **Date publication PCT/PCT Publication Date:** 2018/04/12
(85) **Entrée phase nationale/National Entry:** 2019/03/28
(86) **N° demande PCT/PCT Application No.:** EP 2017/075368
(87) **N° publication PCT/PCT Publication No.:** 2018/065525
(30) **Priorité/Priority:** 2016/10/05 (GB1616912.0)

(51) **Cl.Int./Int.Cl.** ***C12Q 1/68*** (2018.01)

(71) **Demandeur/Applicant:**
UNIVERSITY OF EAST ANGLIA, GB

(72) **Inventeurs/Inventors:**
LUCA, BOGDAN-ALEXANDRU, GB;
MOULTION, VINCENT, GB;
BREWER, DANIEL SIMON, GB;
COOPER, COLIN STEPHEN, GB

(74) **Agent:** BENOIT & COTE INC.

(54) **Titre : CLASSIFICATION ET PRONOSTIC DE CANCER**
(54) **Title: CLASSIFICATION AND PROGNOSIS OF CANCER**

(57) **Abrégé/Abstract:**
The present invention relates to the classification of cancers, in particular prostate cancers, using samples from patients. In particular, the invention provides methods for identifying potentially aggressive prostate cancers to determine which cancers are or will become aggressive (and hence require treatment) and which will remain indolent (and will therefore not require treatment). The present invention is therefore useful to identify patients with a poor prognosis. The specific population of cancer identified by the present invention is referred to herein as DESNT cancer. The invention also provides biomarker panels useful in the diagnosis and prognosis of cancer.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 April 2018 (12.04.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/065525 A1

(51) International Patent Classification:
C12Q 1/68 (2018.01)

(21) International Application Number:
PCT/EP2017/075368

(22) International Filing Date:
05 October 2017 (05.10.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1616912.0 05 October 2016 (05.10.2016) GB

(71) Applicant: **UNIVERSITY OF EAST ANGLIA** [GB/GB]; Norwich Norfolk NR4 7TJ (GB).

(72) Inventors: **LUCA, Bogdan-Alexandru**; c/o University of East Anglia, Norwich Norfolk NR4 7TJ (GB). **MOULTION, Vincent**; c/o University of East Anglia, Norwich Norfolk NR4 7TJ (GB). **BREWER, Daniel Simon**; c/o University of East Anglia, Norwich Norfolk NR4 7TJ (GB). **COOPER, Colin Stephen**; c/o University of East Anglia, Norwich Norfolk NR4 7TJ (GB).

(74) Agent: **LEONARD, Thomas**; Kilburn & Strode LLP, Lacon House, 84 Theobalds Road, LONDON WC1X 8NL (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

WO 2018/065525 A1

(54) Title: CLASSIFICATION AND PROGNOSIS OF CANCER

(57) Abstract: The present invention relates to the classification of cancers, in particular prostate cancers, using samples from patients. In particular, the invention provides methods for identifying potentially aggressive prostate cancers to determine which cancers are or will become aggressive (and hence require treatment) and which will remain indolent (and will therefore not require treatment). The present invention is therefore useful to identify patients with a poor prognosis. The specific population of cancer identified by the present invention is referred to herein as DESNT cancer. The invention also provides biomarker panels useful in the diagnosis and prognosis of cancer.

CLAIMS

1. A method of classifying cancer or predicting cancer progression, comprising:
   a. determining the expression status of a plurality of genes in a sample obtained from the patient to provide a patient expression profile;
   b. conducting a statistical Bayesian clustering analysis or other clustering analysis on the patient expression profile and a reference dataset for the same plurality of genes from different patients;
   c. optionally repeating the analysis step b) multiple times; and
   d. classifying the cancer or predicting cancer progression.

2. The method of claim 1, wherein step b) is repeated at least 2, at least 3, at least 5, at least 20 times, at least 50 times or at least 100 times.

3. The method of claim 2, wherein a different random seed is used for each clustering analysis.

4. The method of any preceding claim, wherein determining the expression status of the plurality of genes comprises determining the level of expression of the plurality of genes.

5. The method of any preceding claim, wherein the expression status is determined for at least 50, at least 100, at least 200 or most preferably at least 500.

6. The method of any preceding claim, further comprising normalising the patent expression profile to the reference dataset prior to conducting the statistical analysis.

7. The method of any preceding claim, where the genes of step a) are selected from the genes listed in Table 1.

8. The method of any preceding claim, wherein step a) comprises determining the expression status of at least 20, at least 50, at least 100, at least 200, at least 500 genes or at least 1000 genes.

9. The method of any preceding claim, wherein step a) comprises determining the expression status of at least 100 genes selected from Table 1.

10. The method of any preceding claim, wherein step a) comprises determining the expression status of all 500 genes selected from Table 1.

11. The method of any preceding claim further comprising a step of selecting a sub-set of genes whose expression status has been determined for statistical analysis.

12. The method of claim 11, wherein the expression status of the each of the genes in the subset of genes is known to vary across cancer patient samples.

13. The method of any preceding claim, further comprising assigning a unique label to the patient expression profile prior to statistical analysis.

14. The method of any preceding claim, wherein step d) comprises determining the DESNT status of the patient cancer sample.

15. The method of any preceding claim, wherein step d) comprises determining the DESNT status of the patient cancer sample.

16. The method of any preceding claim, wherein the DESNT status of each of the expression profiles in the reference dataset is known.

17. The method of any preceding claim, wherein the patient expression profile is combined with at least 2 reference datasets prior to statistical analysis.

18. The method of any preceding claim, wherein the statistical analysis is LPD analysis.

19. The method of claim 18, wherein the LPD analysis organises individual patient expression profiles into groups.

20. The method of claim 19, wherein organising individual patient expression profiles into groups comprises, for each expression profile, using the LDP analysis to determine the contribution (p_i) of each group to the overall expression profile for each patient expression profile.

21. The method of claim 20, wherein a patient expression profile is assigned to an individual group according to the group that contributes the most to the overall expression profile.

22. The method of claim 20 or claim 21, wherein cancer progression in the patient is predicted according to the contribution (p_i value) of the DESNT process to the overall expression profile.

23. The method of claim 22, wherein DESNT cancer is predicted when the contribution of the DESNT process to the overall expression profile is greater than the contribution of any other single process to the overall expression profile.

24. The method of claim 22, wherein DESNT cancer is predicted according to the contribution of the DESNT process to the overall expression profile and according to the stage of the patient's tumour, the Gleason score of the patient and/or PSA score of the patient.

25. The method of claim 22, wherein DESNT cancer is predicted when the p_i value for the DESNT process for the patient cancer sample is at least 0.1, at least 0.2, at least 0.3, at least 0.4 or at least 0.5.

26. The method of claim 19, wherein the groups are assigned either DESNT or a non-DESNT status.

27. The method of claim 26, wherein only group is assigned DESNT status.

28. The method of claim 27, wherein the patient cancer is classified as aggressive or cancer progression is predicted when the patient sample groups with DESNT cancers from the reference dataset or datasets.

29. The method of claim 27, wherein the statistical analysis is carried out multiple times and the patient cancer is classified as aggressive or cancer progression is predicted when the patient sample groups with DESNT cancers from the reference dataset or datasets in at least 60% of runs of the statistical analysis.

30. The method of any preceding claim, wherein the cancer is prostate cancer.

31. A method of classifying cancer or predicting cancer progression, comprising:
    a. providing a reference dataset wherein the cancer progression status of each patient sample in the dataset is known;
    b. selecting from this dataset a plurality of genes, wherein the plurality of genes comprises at least 5, at least 10, at least 20, at least 30, at least 40 or at least 45 genes selected from the group listed in Table 2 or at least 5, at least 10, at least 15 or at least 20 genes selected from the group listed in Table 3;
    c. optionally:
        i. determining the expression status of at least 1 further, different, gene in the patient sample as a control, wherein the control gene is not a gene listed in Table 2 or Table 3; and
        ii. determining the relative levels of expression of the plurality of genes and of the control gene(s);
    d. using the expression status of those selected genes to apply a supervised machine learning algorithm on the dataset to obtain a predictor for cancer progression;

e. determining the expression status of the same plurality of genes in a sample obtained from the patient to provide a patient expression profile;
f. optionally normalising the patient expression profile to the reference dataset; and
g. applying the predictor to the patient expression profile to classify the cancer, or to predict cancer progression.

32. The method of claim 31, wherein the plurality of genes comprises at least 5, at least 10, at least 20, at least 30, at least 40 or at least 45 genes selected from the group listed in Table 2.

33. The method of claim 31, wherein the plurality of genes comprises at least 5, at least 10, at least 15 or at least 20 genes selected from the group listed in Table 3.

34. The method of any one of claims 31 to 33, wherein determining the relative levels of expression comprises determining a ratio of expression for each pair of genes in the patient dataset and the reference dataset.

35. The method of any one of claims 31 to 34, wherein the machine learning algorithm is a random forest analysis.

36. The method of any one of claims 31 to 35, wherein the cancer progression status of each patient sample in the dataset is known according to the DESNT status of the cancer.

37. The method of any one of claims 31 to 36, wherein the DESNT status of the cancer has been previously determined using an analysis involving LPD.

38. The method of claim 37, the DESNT status of the cancer has been previously determined using a method according to any one of claims 1 to 23.

39. The method according to any one of claims 31 to 36, wherein classifying the cancer or predicting cancer progression comprises determining the DESNT status of the cancer.

40. The method of any one of claims 31 to 39, wherein the at least 1 control gene is a gene listed in Table 6 or Table 7.

41. The method of any one of claims 31 to 40, wherein expression status of at least 2 control genes is determined.

42. A method of classifying cancer or predicting cancer progression, comprising:

a. providing a reference dataset wherein the cancer progression status of each patient sample in the dataset is known (for example as determined by LPD analysis);
b. selecting from this dataset of a plurality of genes;
c. using the expression status of those selected genes to apply a supervised machine learning algorithm on the dataset to obtain a predictor for cancer progression;
d. determining the expression status of the same plurality of genes in a sample obtained from the patient to provide a patient expression profile;
e. optionally normalising the patient expression profile to the reference dataset; and
f. applying the predictor to the patient expression profile to classify the cancer, or to predict cancer progression.

43. The method of claim 42, wherein the cancer progression status of each patient sample in the dataset is known according to the DESNT status of the cancer.

44. The method of claim 42 or 43, wherein the DESNT status of the cancer has been previously determined using an analysis involving LPD.

45. The method of claim 44, the DESNT status of the cancer has been previously determined using a method according to any one of claims 1 to 23.

46. The method according to any one of claims 42 to 45, wherein the supervised machine learning algorithm is a random forest analysis.

47. The method according to any one of claims 42 to 46, wherein classifying the cancer or predicting cancer progression comprises determining the DESNT status of the cancer.

48. A method according to any one of claims 42 to 47, wherein at least 10, at least 20, at least 30, at least 40 or at least 50 genes are selected in step b).

49. A method according to any one of claims 42 to 48, wherein the genes selected in step b) are downregulated in cancers that will or have progressed.

50. A method according to any one of claims 42 to 49, wherein the genes selected in step b) comprise at least 5, at least 10, at least 15, at least 20, at least 30, at least 40, or all 45 genes listed in Table 2.

51. A method of classifying cancer or predicting cancer progression, comprising:

a. providing one or more reference datasets where the cancer progression status of each patient sample in the datasets is known (for example as determined by LPD analysis);
b. selecting from this dataset a plurality of genes whose expression statuses are known to vary between cancer that has or will progress and cancer that does not or will not progress;
c. applying a LASSO logistic regression model analysis on the selected genes to identify a subset of the selected genes that predict cancer progression;
d. using the expression status of this subset of selected genes to apply a supervised machine learning algorithm on the dataset to obtain a predictor for DESNT cancers;
e. determining the expression status of the subset of selected genes in a sample obtained from the patient to provide a patient expression profile;
f. optionally normalising the patient expression profile to the reference dataset(s); and
g. applying the predictor to the patient expression profile to classify the cancer or predict cancer progression.

52. The method of claim 51, wherein the cancer progression status of each patient sample in the dataset is known according to the DESNT status of the cancer.

53. The method of claim 51 or 52, wherein the DESNT status of the cancer has been previously determined using an analysis involving LPD.

54. The method of claim 53, wherein the DESNT status of the cancer has been previously determined using a method according to any one of claims 1 to 23.

55. The method of any one of claims 51 to 54, wherein the plurality of genes is at least 100, at least 200, at least 300, at least 400, at least 500 or at least 1000 genes listed in Table 4.

56. The method of any one of claims 51 to 55, wherein the genes predict cancer progression according to the DESNT status of the cancer.

57. The method of any one of claims 51 to 56, wherein the supervised machine learning algorithm is a random forest analysis.

58. The method according to any one of claims 51 to 57, wherein classifying the cancer or predicting cancer progression comprises determining the DESNT status of the cancer.

59. A method according to any preceding claim, wherein the cancer is prostate cancer.

60. A method according to any preceding claim, wherein the sample is a urine sample, a semen sample, a prostatic exudate sample, or any sample containing macromolecules or cells originating in the prostate, a whole blood sample, a serum sample, saliva, or a biopsy.

61. The method of claim 60, wherein the sample is a prostate biopsy, prostatectomy or TURP sample.

62. A method according to any preceding claim, further comprising obtaining a sample from a patient.

63. A method according to any preceding claim, wherein the method is carried out on at least 2, at least 3, at least 3 or at least 5 samples.

64. A method according to claim 63, wherein the method is conducted on the multiple patient samples concurrently.

65. A method according to claim 63, wherein the method is conducting on the multiple patient samples simultaneously.

66. A method according to any preceding claim wherein the dataset or datasets comprise a plurality or tumour or patient expression profiles.

67. The method of claim 66, wherein the datasets each comprise at least 20, at least 50, at least 100, at least 200, at least 300, at least 400 or at least 500 patient or tumour expression profiles.

68. The method of claim 66 or claim 67, wherein the patient or tumour expression profiles comprise information on the expression status of at least 10, at least 40, at least 100, at least 500, at least 1000, at least 1500, at least 2000, at least 5000 or at least 10000 genes.

69. The method of claim 66 or 67, wherein the patient or tumour expression profiles comprise information on the levels of expression of at least 10, at least 40, at least 100, at least 500, at least 1000, at least 1500, at least 2000, at least 5000 or at least 10000 genes.

70. A method of treating cancer, comprising administering a treatment to a patient that has undergone a diagnosis according to the method of any one of claims 1 to 69.

71. The method of claim 70, comprising:

a. providing a patient sample;
b. predicting cancer progression according to method as defined in any one of claims 1 to 69; and
c. administering to the patient a treatment for cancer if cancer progression is predicted, detected or suspected according to the results of the prediction in step b).

72. A method of diagnosing cancer, comprising predicting cancer progression according to a method as defined in any one of claims 1 to 69.

73. A computer apparatus configured to perform a method according to any one of claims 1 to 72.

74. A computer readable medium programmed to perform a method according to any one of claims 1 to 72.

75. A biomarker panel, comprising at least 5, at least 10, at least 20, at least 30, at least 40 or all 45 genes listed in Table 2.

76. The biomarker panel of claim 75, comprising at least 40 genes listed in Table 2.

77. The biomarker panel of claim 75, comprising all 45 genes listed in Table 2.

78. A biomarker panel, comprising at least 5, at least 10, at least 15 or all 20 genes listed in Table 3.

79. The biomarker panel of claim 78, comprising at least 15 genes listed in Table 3.

80. The biomarker panel of claim 78, comprising all 20 genes listed in Table 3.

81. A biomarker panel, comprising at least 10, or at least 15, at least 20, at least 50, at least 100, at least 200, at least 300, at least 400 or all 500 genes listed in Table 1.

82. The biomarker panel of claim 81, comprising at least 400 genes listed in Table 1.

83. The biomarker panel of claim 81, comprising all 500 genes listed in Table 1.

84. A biomarker panel, comprising at least 5, at least 10, at least 20, at least 30 or all 35 genes listed in Table 5.

85. The biomarker panel of claim 84, comprising at least 30 genes listed in Table 5.

86. The biomarker panel of claim 84, comprising all 37 genes listed in Table 5.

87. A biomarker panel prepared by a method according to any one of claims 1 to 69.

88. A biomarker panel according to any one of claims 75 to 87 for use in diagnosing cancer or for use in predicting cancer progression.

89. Use of a biomarker panel according to any one of claims 75 to 87 in a method of diagnosing or prognosing cancer, a method of predicting cancer progression, or a method of classifying cancer.

90. A method of diagnosing or prognosing cancer, or a method of predicting cancer progression, or a method of classifying cancer, comprising determining the level of expression or expression status of the genes in any one of biomarker panels 75 to 87.

91. The method of claim 90, further comprising comparing the level of expression or expression status of the measured biomarkers with one or more reference genes.

92. The method of claim 91, wherein the one or more reference genes is/are a housekeeping gene(s).

93. The method of claim 92, wherein the housekeeping genes is/are selected from the genes in Table 6 or Table 7.

94. The method of claim 93, wherein comparing the level of expression or expressions status of the measured biomarkers.

95. The method of claim 94, wherein the biomarker panel comprises at least at least 5, at least 10, at least 20, at least 30, at least 40 or all 45 genes listed in Table 2, or comprises least 5, at least 10, at least 20, at least 30 or all 35 genes listed in Table 5, and comparing the levels of expression or expression status with a reference comprises comparing with the level of expression or expression status of the same gene or genes in a sample from a healthy patient or a patient that does not have cancer that has or will progress.

96. The method of any one of claims 90 to 95, wherein the level of expression or expression status of the genes is done using a sample obtained from the patient.

97. The method of claim 90, wherein the method comprises conducting a statistical analysis according to a method as defined in any one of claims 1 to 69.

98. A kit comprising means for detecting the level of expression or expression status of at least 5 genes from a biomarker panel as defined in any one of claims 75 to 87.

99. The kit of claim 98, comprising means for detecting the level of expression or expression status of all the genes in a biomarker panel as defined in any one of claims 75 to 87.

100. The kit of claim 98 or claim 99, further comprising means for detecting the level of expression or expression status of one or more control or reference genes.

101. A kit of any one of claims 98 to 100, further comprising instructions for use.

102. A kit of any one of claims 98 to 101, further comprising a computer readable medium as defined in claim 74.

103. A biosensor configured to detect the level of expression or expressions status of at least 5 genes from a biomarker panel as defined in any one of claims 75 to 87.

104. The biosensor of claim 103, wherein the biosensor is configured to detect the level of expression or expression status of all the genes in a biomarker panel as defined in any one of claims 75 to 87.

105. The biosensor of claim 103 or 104, wherein the biosensor is further configured to detect the level of expression or expression status of one or more control or reference genes.

106. The biosensor of any one of claims 103 to 105, wherein the biosensor is a microarray.

107. The kit of any one of claims 98 to 102 comprising a biosensor as defined in any one of claims 103 to 106.