



(51) International Patent Classification:

C12Q 1/68 (2006.01) *G06F 17/18* (2006.01)
G01N 33/48 (2006.01) *G06F 19/18* (2011.01)

(21) International Application Number:

PCT/CA2017/000095

(22) International Filing Date:

25 April 2017 (25.04.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/327,131 25 April 2016 (25.04.2016) US

(71) Applicants: **ONTARIO INSTITUTE FOR CANCER RESEARCH (OICR)** [CA/CA]; MaRS Centre, 661 University Ave., Suite 510, Toronto, Ontario M5G 1L7 (CA). **UNIVERSITY HEALTH NETWORK** [CA/CA]; 190 Elizabeth Street, R. Fraser Elliott Building - Room 1S-417, Toronto, Ontario M5G 2C4 (CA).

(72) Inventors: **BOUTROS, Paul**; 400 Adelaide Street East, Penthouse #9, Toronto, Ontario M5A 4S3 (CA). **BRISTOW, Robert G.**; 610 University Ave., 5th Floor, Toron-

to, Ontario M5G 2M9 (CA). **LALONDE, Emilie**; c/o Ontario Institute for Cancer Research (OICR), MaRS Centre, 661 University Ave., Suite 510, Toronto, Ontario M5G 0A3 (CA).

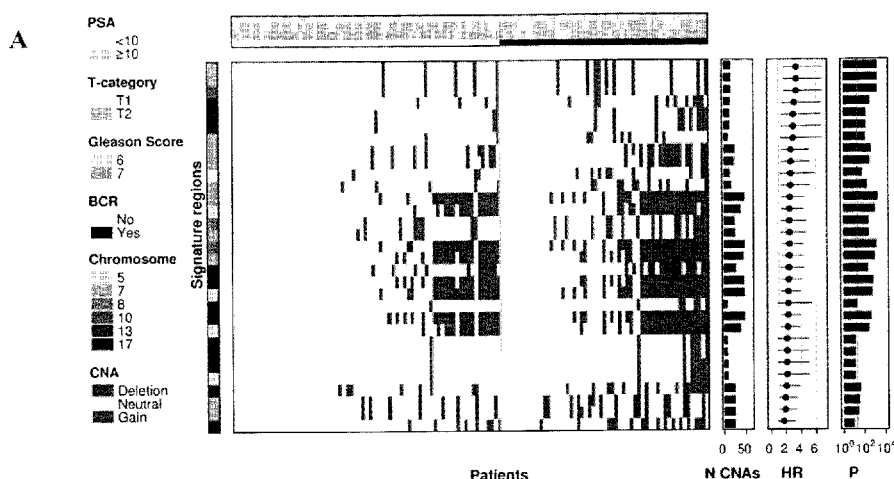
(74) Agent: **CHIU, Jung-Kay**; Norton Rose Fulbright Canada LLP/S.E.N.C.R.L., s.r.l., 1 Place Ville Marie, Suite 2500, Montreal, Quebec H3B 1R1 (CA).

(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, 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,

(54) Title: GENE SIGNATURE FOR PROSTATE CANCER PROGNOSIS

Figure 1



(57) Abstract: There is described herein a method for determining a risk of recurrence of cancer following a cancer therapy of a patient, comprising determining genomic instability of a tumour of the patient from a biopsy by identifying genome regions of the biopsy wherein the regions are at least loci rankings 1-15 of specific 31-loci and using copy number calls and calculating a plurality of statistical distances between the CNA tumour profile and a reference profile of recurring cancer patients to determine the risk of cancer recurrence following the cancer therapy of the patient.



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))*

Gene Signature for Prostate Cancer Prognosis

Field of the Invention

The present invention relates to methods for prostate cancer patient prognosis. Specifically, certain embodiments of the present invention relate to a method for
5 determining a risk of recurrence of cancer following a cancer therapy of a patient, comprising determining genomic instability of a tumour of the patient.

Background of the Invention

Inaccurate prognostication for patients with localized prostate cancer results in significant rates of both over-treatment with associated co-morbidities^{1,2}, and under-
10 treatment resulting in disease progression and increased prostate cancer-specific mortality^{3,4}. Currently, these patients are classified into low, intermediate or high risk groups based on three clinico-pathological variables: Gleason Score (GS), pre-treatment PSA serum concentration(s) and TNM T-category^{5,6}. Patients considered to be low-risk are offered active surveillance, whereas intermediate- and high-risk
15 patients are offered radiotherapy or surgery, with possible intensification using adjuvant hormone therapy treatment. Delineating more precise risk groups would spare patients with insignificant disease from unnecessary interventions while providing confidence to more aggressively treat those patients at high risk of recurrence^{7,8}.

20 Several biomarkers based on the abundances of multiple RNA species have been developed to guide treatment decisions, some of which are available to patients^{e.g.9-11}. We previously developed a DNA-based 100-locus copy number alteration (CNA) genomic classifier for localized disease which effectively stratifies patients into low and high risk of recurrence¹². In two independent cohorts, the genomic classifier identified
25 patients with localized disease (of all risk groups) that would experience a biochemical recurrence (BCR; rising PSA after treatment) within 18-months. This genomic classifier comprises 276 genes and was developed using a ~27,000 probe array comparative genomic hybridization (aCGH) platform. To more effectively translate this genomic classifier to the clinic, we needed to compress the prognostic signature into a smaller
30 feature size that could be processed from routine diagnostic biopsies and be assayed

using a more precise technology.

The NanoString mRNA platform is CLIA certified and is used to implement Prosigna, an FDA-approved assay measuring three breast cancer PAM50 intrinsic subtypes^{13,14}. Their copy number platform (<http://www.nanostring.com/products/CNV>) is used in the
5 OmniSeq Target™ assay to use amplifications in *ERBB2*, *FGFR1* and *MET* to help guide treatment decisions in lung cancer and melanoma patients.

Herein, we refine the 100-locus genomic classifier to 31 loci showing concordant RNA abundance changes, and illustrate the strength of this 31-locus genomic classifier in three ways. First, we show its generalizability by validating it in three new independent
10 cohorts ($n = 104$, $n = 86$, and $n = 102$, respectively), resulting in a total of 563 validation patients. Secondly, the robustness of the genomic classifier is shown by evaluating the classifier using new technologies, including the NanoString CNV targeted assay. Finally, we demonstrate clinical utility of the genomic classifier by evaluating strong clinical endpoints of 10-year biochemical relapse-free and
15 metastasis relapse-free rates.

Summary of the Invention

In an aspect, there is provided a method for determining a risk of recurrence of cancer following a cancer therapy of a patient, comprising determining genomic instability of a tumour of the patient by: (a) obtaining a biopsy of the tumour; (b) identifying genome
20 regions of the biopsy wherein the regions are at least loci rankings 1-15 of the 31-loci in Table A; (c) determining a plurality of copy number calls in the genome regions; (d) intersecting the plurality of copy number calls with a reference gene list, to obtain a plurality of Copy Number Alterations (CNA) calls for each gene; (e) generating a CNA tumour profile based on the plurality of CNA calls; (f) comparing the CNA tumour
25 profile to a reference profile of recurring cancer patients and a reference profile of nonrecurring cancer patients; (g) calculating a plurality of statistical distances between the CNA tumour profile and the reference profile of recurring cancer patients and the reference profile of nonrecurring cancer patients; wherein the statistical distance
30 between the CNA tumour profile and the reference profile of recurring cancer patients and the reference profile of nonrecurring cancer patients is associated with the risk of cancer recurrence following the cancer therapy of the patient.

In an aspect of the present invention, there is provided a method, performed by at least one computing device, for determining the risk of recurrence of cancer following a cancer therapy of a patient, comprising determining genomic instability of a tumour of the patient based on: (a) determining, at a processor, a genome of the tumour; (b) 5 determining, by the processor, genome regions of the biopsy wherein the regions are at least loci rankings 1-15 of the 31-loci in Table A; (c) determining, by the processor, a plurality of copy number calls in the genome regions; (d) determining, by the processor, a plurality of Copy Number Alterations (CNA) calls for each gene by intersecting the plurality of copy number calls with a reference gene list; (e) 10 determining, by the processor, a CNA tumour profile based on the plurality of CNA calls; (f) determining, by the processor, a plurality of statistical distances between the CNA tumour profile and a reference profile of recurring cancer patients and a reference profile of nonrecurring cancer patients; wherein the statistical distance between the CNA tumour profile and the reference profile of recurring cancer patients 15 and the reference profile of nonrecurring cancer patients is associated with a risk of cancer recurrence following the cancer therapy.

In yet another aspect of the present invention, a system for determining the risk of recurrence of cancer following a cancer therapy of a patient comprising determining genomic instability, the system comprising: a non-transitory computer readable storage 20 medium that stores computer-readable code; a processor operatively coupled to the non-transitory computer readable storage medium, the processor configured to implement the computer-readable code, the computer-readable code configured to:

determine a genome of the tumour; determine genome regions of the biopsy wherein the regions are at least loci rankings 1-15 of the 31-loci in Table A; determine a 25 plurality of Copy Number Alterations (CNA) calls for each gene based on intersecting the copy number calls with a reference gene and storing the plurality of CNA calls in the non-transitory computer readable storage medium; determine a CNA tumour profile based on the plurality of CNA calls and storing the CNA tumour profile in a non-transitory computer readable storage medium; determine a plurality of statistical 30 distances between the CNA tumour profile and a reference profile of recurring cancer patients and a reference profile of nonrecurring cancer patients;

wherein the statistical distance between the CNA tumour profile and the reference

profile of recurring cancer patients and the reference profile of nonrecurring cancer patients is associated with a risk of cancer recurrence following the cancer therapy.

Brief Description of the Drawings

- 5 A detailed description of the preferred embodiments is provided herein below by way of example only and with reference to the following drawings, in which:

Figure 1 shows genomic classifier reduction and signature-estimated percent genome alteration. (A) The 31-locus genomic classifier in the training cohort. Patients are sorted according to biochemical recurrence status, then by the number of CNAs in the 31 loci. A univariate Cox proportional hazard model was fit to each feature in the training cohort only. The hazard ratio and Wald p-value are displayed on the right. The red vertical line indicates $p = 0.05$. The 31 loci are sorted according to hazard ratio. (B) The correlation of global PGA to signature-estimated PGA using the reduced 31-locus genomic classifier. Signature-estimated PGA is calculated by the fraction of genomic base pairs involved in a region of CNA when considering the 109 genes in the 31-locus genomic classifier only. (C) The correlation of global PGA to signature-estimated PGA using the reduced 31-locus genomic classifier plus an additional 30 genes which were previously selected to maximize PGA estimation¹².

Figure 2 shows genomic classifier performance in the Combined-Arrays cohort. Cox models are adjusted for clinical variables as in Table 5 RFR: relapse-free rate. (A) The reduced 100-locus and the 31-locus genomic classifiers effectively stratify patients from the Taylor, Ross-Adams, Hieronymus, and Stockholm cohorts, including patients diagnosed with low, intermediate, and high risk disease. (B) The reduced 31-locus genomic classifier effectively stratifies patients from the Taylor, Ross-Adams, Hieronymus, and Stockholm cohorts, including only patients diagnosed with low and intermediate risk disease. (C) The reduced 31-locus genomic classifier effectively stratifies patients from the Taylor, Hieronymus, and Stockholm cohorts, including only patients diagnosed with high risk disease. (D) The reduced 31-locus genomic classifier effectively predicts which patients from the Taylor cohort will develop metastasis.

30 Figure 3 shows clinical utility of the reduced 31-loci genomic classifier. (A) Receiver

operator curve analysis for predicting BCR at 5 years with the 31-locus genomic classifier, clinical models, and clinico-genomic models. The "31 loci + risk." model includes the continuous 31-locus risk score and the NCCN risk groups whereas the "31 loci + GS/PSA/T" model includes the continuous 31-locus risk score, Gleason score, T-category, and PSA. The clinico-genomic classifiers have significantly higher AUCs than their matching clinical-only classifiers ($p = 0.00111$ and $p = 0.00745$ for the 31 loci + risk and 31 loci + GS/PSA/T, respectively, based on 5000 bootstrapped permutations). AUC = Area under the curve; GS = Gleason Score; T = T-category from TNM; PSA = Prostatic Specific Antigen. (B) The net reclassification index (NRI) based on using the full clinico-genomic model (31-loci + GS/PSA/T) in comparison to the clinical model with GS/PSA/T only for predicting BCR in the Combined-Arrays cohort. (C) Receiver operator curve analysis for predicting metastasis at 10 years with the 31-locus genomic classifier, clinical models, and clinico-genomic models in the Taylor cohort. The "31 loci + risk." model includes the continuous 31-locus risk score and the NCCN risk groups whereas the "31 loci + GS/PSA/T" model includes the continuous 31-locus risk score, Gleason score, T-category, and PSA. The clinico-genomic classifiers have significantly higher AUCs than their matching clinical-only classifiers ($p = 0.0980$ and $p = 0.159$ for the 31-loci + risk and 31-loci + GS/PSA/T, respectively, based on 5000 bootstrapped permutations). GS = Gleason Score; T = T-category from TNM; PSA = Prostatic Specific Antigen. (D) The net reclassification index (NRI) based on using the full clinico-genomic model (31 loci + GS/PSA/T) in comparison to the clinical model with GS/PSA/T only for predicting metastasis in the Taylor cohort. The overall NRI is indicated in the legend.

Figure 4 shows validation of reduced 31-locus genomic classifier in the CPC-GENE cohort using the NanoString platform. (A) The Kaplan-Meier curves for the CPC-GENE cohort stratified by the reduced 31-locus genomic signature. The Cox model is adjusted for clinical variables as shown in Table 5. (B) Receiver operator curve analysis for predicting BCR at 5 years with the 31-locus genomic classifier, clinical models, and clinico-genomic models. The "31 loci + Risk." model includes the continuous 31-locus risk score and the NCCN risk groups whereas the "31 loci + GS/PSA/T" model includes the continuous 31-locus risk score, Gleason score, T-category, and PSA. AUC = Area under the curve; GS = Gleason Score; T = T-category from TNM; PSA = Prostatic Specific Antigen.

Figure 5 shows signature reduction strategy. A flowchart of the workflow used in this study. The signature was refined by examining RNA ~ DNA associations in the Taylor dataset using logistic regression (see Methods for full details). Of the 276 genes, 36 genes had significant associations using a lenient threshold (false-discovery rate adjusted p-values < 0.1). We selected the 31 loci containing one of these 36 genes to build the 31-locus genomic classifier. This 31-locus (or 109 gene) classifier was trained on the Toronto aCGH cohort, which is the same cohort used to train the original classifier. To validate this signature, we collapsed the 109 genes contained within the 31 genomic classifier loci to get a CNA call for each of the 31 loci for each patient.

Figure 6 shows prognosis of clinical variables in the CPC-GENE cohort. Cox proportional hazard regression models were fit to each individual clinical covariate (A: Gleason score, B: T-category, C: PSA, D: NCCN risk-group). When there were more than two levels (i.e. NCCN risk group) a logrank test was used instead.

Figure 7 shows prognosis of clinical variables in the Taylor cohort. Logrank tests were used to quantify the prognosis of each individual clinical covariate (A: Gleason score, B: T-category, C: PSA, D: NCCN risk-group).

Figure 8 shows prognosis of clinical variables in the Ross-Adams cohort. Logrank tests were used to quantify the prognosis of each individual clinical covariate (A: Gleason score, B: T-category, C: PSA, D: NCCN risk-group). Since there were only two patients with PSA ≥ 20 , only two levels were used to evaluate the prognosis of PSA, and a Cox proportional hazard regression model was used to quantify this effect.

Figure 9 shows prognosis of clinical variables in the Hieronymus cohort. Cox proportional hazard regression models were fit to each individual clinical covariate (A: Gleason score, B: T-category, C: PSA, D: NCCN risk-group). When there were more than two levels (i.e. NCCN risk group) a logrank test was used instead.

Figure 10 shows prognosis of clinical variables in the Stockholm cohort. Logrank tests were used to quantify the prognosis of each individual clinical covariate (A: Gleason score, B: T-category, C: PSA, D: NCCN risk-group).

Figure 11 shows prognosis of clinical variables in the Combined-Arrays cohort. Logrank tests were used to quantify the prognosis of each individual clinical covariate

(A: Gleason score, B: T-category, C: PSA, D: NCCN risk-group).

Figure 12 shows association of mRNA abundance with copy number status. The signature was refined by examining RNA ~ DNA associations in the Taylor dataset using logistic regression (see Methods for full details). Genes that had both deletions and gains were fit as a three-level factor. The mRNA abundance of the gene was used as a continuous variable and modeled as a function of the CNA state of the same gene. Of the 276 genes from the original full genomic classifier, 36 genes had significant associations for deletions and/or gains using a lenient threshold (false-discovery rate adjusted p-values < 0.1).

Figure 13 shows Gini scores for full and 31-locus genomic classifiers. Gini score (represents the relative importance of each feature in the genomic classifier) of the 100-loci and reduced 31-locus genomic classifiers. The 31-locus genomic classifier illustrates the loci which were selected due to associations between gene copy number and mRNA abundance.

Figure 14 shows correlation of genomic classifier scores in Taylor and Ross-Adams cohorts using the 31-locus genomic classifier compared to the full genomic classifier. A comparison of the genomic classifier scores produced by the full and reduced classifiers. "Yes votes" represents the genomic classifier score produced by the random forest models.

Figure 15 shows validation of the 31-locus genomic classifier per cohort. Cox proportional hazard regression models are adjusted for clinical variables as defined in Table 5. The HR and p-value shown are for the genomic classifier term. A- Taylor BCR. B- Ross-Adams BCR. C- Hieronymus BCR. D- Stockholm BCR.

Figure 16 shows CNA rate per cohort. A- The proportion of CNAs per gene per cohort. B- The proportion of CNAs per patient per cohort.

Figure 17 shows validation of the 31-locus genomic classifier in low-risk (A) and intermediate-risk (B) patients from the Combined-arrays cohort. Cox proportional hazard regression models are adjusted for clinical variables as defined in Table 5. The HR and p-value shown are for the genomic classifier term.

Figure 18 shows AUC comparison for the 31-locus genomic classifier in each cohort.

Receiver operator characteristic (ROC) comparison of various clinico-genomic models. Each model was fit with a Cox proportional hazard regression model and the predicted risk scores were evaluated using in a ROC analysis for each cohort separately: A- Taylor. B- Ross-Adams. C- Hieronymus. D- Stockholm.

- 5 Figure 19 shows AUC comparison for 31-locus genomic classifier in the Combined-Arrays cohort. A- All patients. B- Low-risk patients. C- Intermediate-risk patients. D- High-risk patients.

Figure 20 shows clinical utility of the reduced 31-loci genomic classifier. The probability of BCR increases as the full clinico-genomic risk score increases. The full clinico-
10 genomic risk score is the predicted risk from the multivariate Cox model predicting 5-year BCR.

Figure 21 shows concordance of NanoString replicates. Concordance of the six sets of replicate samples processed on the NanoString platform. A- CNAs identified in each replicate sample, with genes and samples clustered. All replicates pair with each other,
15 except for one of the three CPG0462 replicate samples. Only endogenous and housekeeping probes were used. B- The concordance rate per sample, where the concordance is measured as the number of unique CNA calls (deletion, neutral or gain) divided by the number of replicates.

Figure 22 shows suitable configured computer device, and associated communications
20 networks, devices, software and firmware to provide a platform for enabling one or more embodiments as described herein.

In the drawings, preferred embodiments of the invention are illustrated by way of example. It is to be expressly understood that the description and drawings are only
25 for the purpose of illustration and as an aid to understanding, and are not intended as a definition of the limits of the invention.

Detailed Description of the Invention

Prostate cancer is a clinically heterogeneous disease, despite tightly-defined, clinical risk-groups that represent relative prostate cancer-specific mortality. Patients with localized disease are burdened with high rates of both over-treatment and under-
30 treatment, suggesting that current patient stratification schemes are inefficient to triage

patients to less or more intensive treatment protocols. Previously, we developed a 100-locus DNA genomic classifier capable of sub-stratifying patients at risk of biochemical relapse within NCCN risk groups. The genomic classifier contained 276 genes which were enriched for lipid metabolism genes, and associated with global genomic instability. The 100-locus DNA genomic classifier is also described in WO 2015/106341, the disclosure of which is hereby incorporated by reference.

Herein, we reduced this classifier to 109 genes from 31 DNA loci, and evaluated this 31-locus genomic classifier in 563 men with localized disease, including a cohort of 102 men where the genomic classifier was measured with the NanoString CNV platform. We find that this reduced 31-locus genomic classifier identified a subset of patients with localized disease with failure rates more than 2-times faster than remaining patients (hazard ratio (HR) = 2.42; 95% confidence interval (CI) 1.39-3.20; Wald $P = 4.77 \times 10^{-5}$). This difference is clinically meaningful as combining the genomic classifier with standard clinical prognostic variables (i.e. clinico-genomic classifiers) outperforms prognostic models using the clinical variables only (e.g. area under the survival curve (AUC) 0.75 vs. 0.67, $p = 0.00910$) and shows increases in clinical benefit as measured with net reclassification analyses. Furthermore, in one cohort where metastatic information is available, the signature is highly effective at identifying patients at risk of developing metastasis (HR = 5.94; 95% CI: 1.69-20.8; $p = 0.0054$). This finding will be validated in the remaining cohorts as follow-up time increases and metastatic events are documented. Finally, we developed probe sets to measure the reduced 31-locus genomic classifier with the NanoString CNV platform. This assay will be useful for prospective clinical trials where all patients can be evaluated using this assay mitigating any potential biases with retrospective studies. We propose novel clinical trials for patient management based on combined clinico-genomic classifiers where patients are selected for de-escalation or escalation trials on the basis of the 31-locus clinico-genomic classifier.

In an aspect, there is provided a method for determining a risk of recurrence of cancer following a cancer therapy of a patient, comprising determining genomic instability of a tumour of the patient by: (a) obtaining a biopsy of the tumour; (b) identifying genome regions of the biopsy wherein the regions are at least loci rankings 1-15 of the 31-loci in Table A; (c) determining a plurality of copy number calls in the genome regions; (d) intersecting the plurality of copy number calls with a reference gene list, to obtain a

plurality of Copy Number Alterations (CNA) calls for each gene; (e) generating a CNA tumour profile based on the plurality of CNA calls; (f) comparing the CNA tumour profile to a reference profile of recurring cancer patients and a reference profile of nonrecurring cancer patients; (g) calculating a plurality of statistical distances between the CNA tumour profile and the reference profile of recurring cancer patients and the reference profile of nonrecurring cancer patients; wherein the statistical distance between the CNA tumour profile and the reference profile of recurring cancer patients and the reference profile of nonrecurring cancer patients is associated with the risk of cancer recurrence following the cancer therapy of the patient.

10 In further embodiments, the present 31-locus genomic classifier may be used and implemented in embodiments similar to those described in WO 2015/106341, previously incorporated by reference.

As used herein, "genomic instability" is the degree of genetic differences that exist between a reference genetic baseline and a genetic sample. The genetic differences that exist may be expressed by proxy with specific reference to the number of copy number calls made between the reference genetic baseline and the genetic sample.

As used herein, "locus" is a specific genetic region of variable length and identity. A ranking of a selection of relevant loci is found in Table A.

As used herein, "copy number call" is the quantity of a genetic unit obtained from a genetic sample subjected to a genetic assay. Copy number calls may be assessed thorough the use of an amplified fragment pool assay, as described more fully below.

As used herein, "copy number alteration", or CNA, is the value representing a comparison of the copy number call of a given genetic unit to that of a reference genome that may give rise to a determination as to whether there is a loss or gain of genetic material for that given genetic unit.

As used herein, "CNA tumour profile" is the plurality of CNAs associated with a given genetic tumour sample.

As used herein, "reference profile of recurring cancer patients" is the plurality of CNAs associated with a given set of genetic tumour samples of a population of patients wherein it is known that cancer reoccurred after a given cancer treatment.

As used herein, “reference profile of nonrecurring cancer patients” is the plurality of CNAs associated with a given set of genetic tumour samples of a population of patients wherein it is known that cancer did not reoccur after a given cancer treatment.

As used herein, “statistical distance” is a value representing the comparison of sets of data that gives rise to a determination of the degree of association, or lack thereof, between said sets of data. A specific embodiment of a statistical distance may be the use of a Jaccard distance (Jaccard, 1901), as described more fully below.

In an embodiment, the genome regions are at least loci rankings 1-20, 1-25, or 1-31 in Table A.

10 In an embodiment, the genome regions are a whole tumour genome.

In some embodiments, the patient has been diagnosed with prostate cancer. In some instances, the patient has been diagnosed with localized prostate cancer. Preferably, the patient has one of a low or intermediate risk for prostate cancer. For example, the patient has one of a low or intermediate risk for prostate cancer as determined by at least one of T-category, Gleason score or pre-treatment prostate-specific antigen blood concentration.

Classifying a patient as being at low, intermediate or high risk for prostate cancer mortality is well understood by a person skilled in the art. For example, there are five common classification systems used to clinically stratify prostate cancer patients into low, intermediate or high risk groups: NCCN, D'Amico, GUROC, CAPSURE and ESMO (see Table 10). Each of these will stratify prostate cancer patients as low, intermediate or high risk based on Gleason score, pre-treatment PSA and T-category. The Gleason score is obtained from the diagnostic biopsy, and determined by a pathologist. The T-category is related to the size and spread of the tumour within the prostate and surrounding area, as determined by a digital rectum exam and imaging tests. PSA is a blood-based biomarker, measured in ng/mL.

In some embodiments, the low risk for prostate cancer is determined by at least one of the following: (a) a T-category of T1-T2a, a Gleason score less than or equal to 6, and a pre-treatment prostate-specific antigen blood concentration less than or equal to 10 ng/mL; (b) a T-category of T1-T2a, a Gleason score greater than or equal to 2 and less

than or equal to 6, and a pre-treatment prostate-specific antigen blood concentration less than or equal to 10 ng/mL; and (c) a T-category of T1c, a Gleason score less than or equal to 6, a pre-treatment prostate-specific antigen blood concentration less than or equal to 10 ng/mL, and fewer than 3 biopsy cores of a tumour that are positive
5 for cancer and having less than or equal to 50% cancer in each.

In some embodiments, the intermediate risk for prostate cancer is determined by at least one of the following: (a) at least one of a T-category of T2b, a Gleason score equal to 7, and a pre-treatment prostate-specific antigen blood concentration greater than 10 ng/mL; (b) at least one of a T-category of T1-T2, a Gleason score equal to or
10 less than 7, and a pre-treatment prostate-specific antigen blood concentration less than or equal to 20 ng/mL; (c) at least one of a T-category of T2b, a Gleason score equal to 7 and a pre-treatment prostate-specific antigen blood concentration greater than 10 ng/ml and equal to or less than 20 ng/mL; and (d) at least one of a T-category
15 of T2b, a T-category of T2c, a Gleason score equal to 7 and a pre-treatment prostate-specific antigen blood concentration greater than 10 ng/ml and equal to or less than 20 ng/mL.

The present system and method may be practiced in various embodiments. A suitably configured computer device, and associated communications networks, devices, software and firmware may provide a platform for enabling one or more embodiments
20 as described above. By way of example, Figure 22 shows a generic computer device 100 that may include a central processing unit ("CPU") 102 connected to a storage unit 104 and to a random access memory 106. The CPU 102 may process an operating system 101, application program 103, and data 123. The operating system 101, application program 103, and data 123 may be stored in storage unit 104 and loaded
25 into memory 106, as may be required. Computer device 100 may further include a graphics processing unit (GPU) 122 which is operatively connected to CPU 102 and to memory 106 to offload intensive image processing calculations from CPU 102 and run these calculations in parallel with CPU 102. An operator 107 may interact with the computer device 100 using a video display 108 connected by a video interface 105,
30 and various input/output devices such as a keyboard 115, mouse 112, and disk drive or solid state drive 114 connected by an I/O interface 109. In known manner, the mouse 112 may be configured to control movement of a cursor in the video display 108, and to operate various graphical user interface (GUI) controls appearing in the

video display 108 with a mouse button. The disk drive or solid state drive 114 may be configured to accept computer readable media 116. The computer device 100 may form part of a network via a network interface 111, allowing the computer device 100 to communicate with other suitably configured data processing systems (not shown).

- 5 One or more different types of sensors 135 may be used to receive input from various sources.

The present system and method may be practiced on virtually any manner of computer device including a desktop computer, laptop computer, tablet computer or wireless handheld. The present system and method may also be implemented as a computer-
10 readable/useable medium that includes computer program code to enable one or more computer devices to implement each of the various process steps in a method in accordance with the present invention. In case of more than computer devices performing the entire operation, the computer devices are networked to distribute the various steps of the operation. It is understood that the terms computer-readable
15 medium or computer useable medium comprises one or more of any type of physical embodiment of the program code. In particular, the computer-readable/useable medium can comprise program code embodied on one or more portable storage articles of manufacture (e.g. an optical disc, a magnetic disk, a tape, etc.), on one or more data storage portioned of a computing device, such as memory associated with a
20 computer and/or a storage system.

In an aspect of the present invention, there is provided a method, performed by at least one computing device, for determining the risk of recurrence of cancer following a cancer therapy of a patient, comprising determining genomic instability of a tumour of the patient based on: (a) determining, at a processor, a genome of the tumour; (b)
25 determining, by the processor, genome regions of the biopsy wherein the regions are at least loci rankings 1-15 of the 31-loci in Table A; (c) determining, by the processor, a plurality of copy number calls in the genome regions; (d) determining, by the processor, a plurality of Copy Number Alternations (CNA) calls for each gene by intersecting the plurality of copy number calls with a reference gene list; (e)
30 determining, by the processor, a CNA tumour profile based on the plurality of CNA calls; (f) determining, by the processor, a plurality of statistical distances between the CNA tumour profile and a reference profile of recurring cancer patients and a reference profile of nonrecurring cancer patients; wherein the statistical distance

between the CNA tumour profile and the reference profile of recurring cancer patients and the reference profile of nonrecurring cancer patients is associated with a risk of cancer recurrence following the cancer therapy.

In yet another aspect of the present invention, a system for determining the risk of recurrence of cancer following a cancer therapy of a patient comprising determining genomic instability, the system comprising: a non-transitory computer readable storage medium that stores computer-readable code; a processor operatively coupled to the non-transitory computer readable storage medium, the processor configured to implement the computer-readable code, the computer-readable code configured to:

- 10 determine a genome of the tumour; determine genome regions of the biopsy wherein the regions are at least loci rankings 1-15 of the 31-loci in Table A; determine a plurality of Copy Number Alterations (CNA) calls for each gene based on intersecting the copy number calls with a reference gene and storing the plurality of CNA calls in the non-transitory computer readable storage medium; determine a CNA tumour
- 15 profile based on the plurality of CNA calls and storing the CNA tumour profile in a non-transitory computer readable storage medium; determine a plurality of statistical distances between the CNA tumour profile and a reference profile of recurring cancer patients and a reference profile of nonrecurring cancer patients;

wherein the statistical distance between the CNA tumour profile and the reference profile of recurring cancer patients and the reference profile of nonrecurring cancer patients is associated with a risk of cancer recurrence following the cancer therapy.

The present invention will be understood by reference to the following non-limiting examples:

Examples

25 **Materials and Methods**

Patient cohorts

Four previously published cohorts of patients treated by radical prostatectomy (RadP) were available for validation: Taylor (154 patients¹⁵), Ross-Adams (117 patients¹⁶), Hieronymus (104 patients¹⁷) and Stockholm (86 patients¹⁸); Table 1). We previously

used the Taylor and Ross-Adams cohorts to validate our 100-locus genomic classifier, and processing of those datasets were described in that report¹². To assess patient prognosis, cohorts were considered both individually and combined ("Combined-arrays cohort"). Finally, we measured the signature using the NanoString CNV assay in our
5 own cohort of 102 patients treated by RadP (CPC-GENE cohort, below).

Taylor cohort

We obtained updated clinical information from the Taylor cohort from a recent publication¹⁹, increasing median time to BCR from 4.6 to 7.8 years, and median survival time from 4.8 to 8.6 years. Importantly, time to metastasis is now available for
10 this cohort (median time to metastasis is 7.9 years), providing evaluation of a second clinical endpoint to further corroborate BCR analyses.

Ross-Adams cohort

CNAs were called for Ross-Adams cohort with OncoSNP which ranks CNA calls from 1-5. A rank of 1 indicates high confidence calls, and a rank of 5 represents the least
15 confident calls. We used CNA with ranks 1-3 as previously described¹².

Hieronymus cohort

The previously normalized $\log_2$ tumour-normal ratio values for probes was obtained from GEO (GSE54691). DNACopy was used to generate continuous segments with smoothing and segmentation with default parameters except the 'undo.splits' option
20 which was set to 'none'. CNA states (deletion, neutral or gain) were assigned by using a kernel density approach. Segments with $\log_2$ ratios greater than 0.17 are assigned as gains, and segments with $\log_2$ tumour-normal ratios less than -0.2 are assigned as deletions. The remaining segments are considered neutral. These thresholds resulted in PGA distribution similar to other published cohorts¹². Clinical data was obtained from
25 a previous publication¹⁷.

Table 1. Clinical characteristics of cohorts used in the study. *Combined-arrays cohort refers to the microarray cohorts only (Taylor, Hieronymus, Ross-Adams and Stockholm cohorts). *See methods for full details of NanoString preprocessing and CNA calling algorithm.

	Taylor	Hieronymus	Ross-Adams	Stockholm	Combined-arrays*	CPC-GENE (NanoString)
CNA platform	Agilent 244k aCGH	Agilent 1M aCGH	Illumina Human Omni 2.5-8 bead chip	Affymetrix SNP6	NA	NanoString CNV counter
CNA calling algorithm	RAE	RAE	OncoSNP	ASCAT	NA	NanoString linear model [†]
Primary treatment						
RadP	154	104	117	86	461	102
Gleason score						
5	2	0	0	6	8	0
6	82	48	28	35	193	11
7	53	43	73	32	201	91
8	11	10	11	10	42	0
9	6	3	1	2	12	0
NA	0	0	4	1	5	0
T category						
T1	79	61	67	37	244	48
T2	66	41	35	37	179	54
T3	9	2	15	9	35	0
NA	0	0	0	3	3	0
PSA						
<10	115	87	83	50	335	80
10 – 20	24	12	31	24	91	21
≥20	15	5	2	9	22	1
NA	0	0	1	3	4	0

Risk group						
Low	58	38	16	29	141	6
Intermediate	66	51	70	33	220	95
High	30	15	26	21	92	1
NA	0	0	5	3	8	0
BCR						
No	107	80	97	44	328	54
Yes	47	24	20	42	133	48
Median time to BCR	7.8 years	5.1 years	3.0 years	6.9 years	5.8 years	6.6 years
Median follow-up time	8.6 years	6.4 years	2.9 years	6.9 years	7.0 years	8.5 years

Table A: 31-loci location

Feature	Chromosome	Start	End	Genes_in_loci
1	8	8559665 9057965	8890849	CLDN23,MFHAS1,ERI1
2	10	8 9003362	90611732	ANKRD22
3	10	0	90343082	RNLS
4	8	8993763	9009152	PPP1R3B
5	5	892968	918164	TRIP13
6	5	271735 1286977	443258	PDCD6,AHRR,C5orf55
7	8	2 1257940	12887284	C8orf79
8	8	5	12612992	LONRF1
9	8	9413444	9639856	TNKS
10	13	1.12E+08 1170003	1.12E+08	ANKRD10
11	8	3 4096214	11853760	CTSB,DEFB137,DEFB136,DEFB134
12	17	9	41065386	BECN1,PSME3,AOC2,AOC3,G6PC
13	7	1.5E+08	1.5E+08	ZNF862
14	17	7623038	7853237	DNAH2,KDM6B,TMEM88,LSMD1,CYB5D1,CHD3,KCNAB3,TRAPPC1,CNTROB
15	17	7942357	7952451	ALOX15B
16	17	7999217 4239693	8151413	ALOXE3,HES7,PER1,VAMP2,TMEM107,C17orf59,AURKB,C17orf68
17	8	8	42408140	C8orf40
18	8	4224927	42397068	VDAC3,SLC20A2

19	8	4201046	3	42065194	AP3M2, PLAT	
20	8	1114199	9	11189695	MTMR9, AMAC1L2	
21	11	1.17E+08	1.17E+08	BUD13, ZNF259, APOA5, APOA4, APOC3, APOA1		
22	11	1.17E+08	1.18E+08	KIAA0999, PAFAH1B2, SIDT2, TAGLN, PCSK7, RNF214, BACE1, CEP164, DSCAML1, FXVD		
23	8	1156171	6	11696818	1	GATA4, NEIL2, FDFT
24	20	4512970	6	45985474	ZNF334, SLC13A3, TP53RK, SLC2A10, EYA2, ZMYND8	
25	20	4628614	9	46415360	SULF2	
26	20	3965746	1	40247133	TOP1, PLCG1, ZHX3, LPIN3, EMILIN3, CHD6	
27	20	4208650	3	42170535	SFRS6, L3MBTL	
28	20	4221957	8	42345122	IFT52, MYBL2	
29	22	5062435	9	50883518	TRABD, RP3-	
30	20	6079001	6	61303647	O4A1	
31	16	8863678	8	88729495	ZC3H18, IL17C, CYBA, MVD	

Stockholm cohort

The normalized genomic data was downloaded from GEO (accession GSE73076). ASCAT (version 2.1) was used to segment the BAF and LRR using default parameters²⁰. One of three copy number states (neutral, loss, or gain) was assigned to each segment relative to the average genome ploidy (AGP) per tumour sample as calculated by ASCAT. The following thresholds were tested to assign copy number states: 0.6, 0.7, 0.8, 0.9 and 1.0. A threshold of 0.9 was selected, such that the median percent genome alteration of the cohort is between 2-4%, which is consistent with the median PGA of other published prostate cancer cohorts¹². Segments whose copy number was greater than $AGP + 0.9$ were annotated as gain and whose copy number was greater than $AGP - 0.9$ were annotated as loss. All other segments were annotated as neutral. The clinical data was obtained from a previous publication¹⁶.

Combined-arrays cohort

The three cohorts processed using CNA microarrays were combined to assess patient prognosis in a larger cohort of 461 patients, allowing for examination of effect within NCCN risk groups.

CPC-GENE Cohort

The CPC-GENE cohort consists of 102 prostate cancer patients treated by RadP, a subset of the Canadian Prostate Cancer Genome Network (CPC-GENE). The patients are primarily intermediate risk, as defined by the National Comprehensive Cancer Network (NCCN; GS = 6-7, 20 > PSA > 10, cT1 or T2; Table 1). Fresh frozen RadP specimens were obtained mostly (69/100) from the University Health Network Pathology BioBank, and the remaining 33 from the Genito-Urinary BioBank of the Centre Hospitalier Universitaire de Québec (CHUQ). Whole blood and informed consent were collected during follow-up clinical appointments. Tumour tissues had been collected according to protocols approved by the University Health Network Research Ethics Board (UHN 06-0822-CE, UHN 11-0024-CE, CHUQ 2012-913:H12-03-192). GS and tumour cellularity were independently evaluated by two genitourinary pathologists (TvdK, BT). Importantly, none of the patients included in this study were used in the initial genomic classifier discovery¹².

Samples were cut into 60 x 10 µm sections, and one in every 10 cuts 4 µm section was H&E-stained. Areas with at least 70% tumour cellularity (as assessed by two pathologists, TvdK and BT) were macro-dissected with sterile scalpel blades, and DNA was extracted by phenol:chloroform, as previously reported²¹. ArchivePure DNA Blood
 5 Kit (5 PRIME, Inc., Gaithersburg, MD) was used to extract DNA from whole blood using at the Applied Molecular Profiling Laboratory at the Princess Margaret Cancer Centre. A Qubit 2.0 Fluorometer (Life Technologies, Burlington, ON) a Nanodrop ND-1000 spectrophotometer were used to quantify and assess purity of the DNA samples, respectively.

10 NanoString assay

Feature selection

CPC-GENE samples were processed on the NanoString CNV platform in 350ng/7µL concentrations at the Princess Margaret Genomics Center in Toronto, Ontario. To design the NanoString probes, we limited each of the original 100 genomic classifier
 15 regions up to 3 probes, each representing by a gene. Where regions consisted of 1-3 genes (n=49, n=16, n=10, respectively), every gene was included in the NanoString gene panel. Where a genomic classifier region consisted of more than 3 genes (n = 25), genes in the region were examined for evidence of involvement in prostate cancer or cancer. We used the prostate data from the Taylor *et al.* study¹⁵, and the 81
 20 samples from The Cancer Genome Atlas available at the time (publicly available on the cBioPortal^{22,23}). We checked genes for evidence of 1) somatic SNVs 2) prognosis of mRNA abundance 3) mRNA abundance changes reflecting underlying CNA status 4) recurrent CNAs and 5) literature-based evidence of involvement in prostate cancer or the hallmarks of cancer. In total, 251 genes were included in the panel which
 25 covered the 100 signature regions as well as included 26 candidate prostate cancer driver genes (Table 4), 9 housekeeping genes and 30 genes to maximize percent genome alteration (PGA) estimation, as previously described¹².

Table 4: Reduced genomic classifier

The 31 loci included in the reduced genomic classifier and the number of genes
 30 contained within each locus. There is one NanoString probe per gene per signature locus, except when there are more than three genes per region, where three

NanoString probes were selected.

Signature Row	Genes in locus	NanoString probes in locus
1	3	3
2	1	1
3	1	1
4	1	1
5	1	1
6	3	3
7	1	1
8	1	1
9	1	1
10	1	1
11	4	3
12	5	3
13	1	1
14	9	3
15	1	1
16	8	3
17	1	1
18	2	2
19	2	2
20	2	2
21	6	3
22	10	3
23	3	3
24	6	3
25	1	1
26	6	3
27	2	2
28	2	2
29	9	3
30	11	3
31	4	3

NanoString normalization

Using the NanoString platform, we processed 102 RadP specimens and blood
 5 samples from 30 patients were used to generate a pooled-normal reference. The
 NanoString RCC files were loaded into the R statistical language with the

NanoStringNorm package (v1.1.20)²⁴. All 132 samples passed quality control as evaluated by coverage and variance of control genes, and no observable batch effect (Adjusted Rand Index 0.209).

- To normalize the data and call CNAs, we applied a linear model to the raw data.
- 5 Briefly, the measured counts of each gene were divided by the median invariant probe count per sample. The observed counts are modeled as a function of the sample, region, and a sample-region interaction term:

$$\log_2(\text{observed counts}) = B_0 + \text{sample} + \text{region} + \text{sample:region}$$

- The sample:region term estimates the log copy number and NanoString defined
- 10 thresholds are used to convert these continuous results to ternary copy number calls. We have updated NanoStringNorm²⁴ to perform additional quality control and normalization techniques specific to CNA analysis (Lalonde *et al.*, in preparation).

Signature reduction

- To reduce the number of features in the genomic classifier (100 genomic regions,
- 15 corresponding to 276 genes), we focused on biologically-relevant loci by selecting those where the RNA levels reflect the underlying copy number state (Figure 5). The 110 Taylor patients with both CNA and mRNA information were used in a multinomial logistic regression model, where the copy number was fit as a 3-level factor (-1, 0, 1). The smallest p-value per model was used for feature selection, and loci with genes
- 20 that have Benjamini-Hochberg adjusted p-values less than 0.1 and the expected RNA change in terms of the CNA were selected (n = 36 genes from 31 signature loci). These 31 loci (containing 109 genes) were used to re-train a random forest²⁵ in the original aCGH data, consisting of 126 low- and intermediate-risk patients treated with IGRT¹². The randomForest package (v4.6-10) with 100,000 trees and otherwise
- 25 default parameters was used to train the 31-locus genomic classifier, which was then applied to the remaining cohorts alone ("31-locus genomic classifier") or in combination with the clinical variables NCCN risk group ("31-loci + risk clinico-genomic classifier"), or GS, PSA and T-category ("31-loci + GS/PSA/T clinico-genomic classifier").

- 30 *Statistical analysis*

- The primary endpoints were biochemical recurrence at 5 and 10 years. Biochemical recurrence is defined as two consecutive PSA readings above 0.2 ng/mL, or salvage treatment. At 10 years, 48 CPC-GENE patients (48%), 46 Taylor patients (30%), 24 Hieronymus patients (23%), and 42 Stockholm patients (49%) have had a biochemical failure. For the UK cohort with shorter follow-up time, 19 patients (16%) have had a biochemical recurrence by 18 months. Median time to event was estimated using the Kaplan-Meier approach²⁶. In each cohort, the univariate prognosis of GS, T-category and PSA were assessed with Cox proportional hazard regression models when the variable had only two levels, and otherwise with a logrank test. (Figures 6-11).
- 10 To assess prognostic ability of the 31-locus genomic classifier, Cox proportional hazard regression models were used for creating both continuous and discretized risk scores. The survMisc package (v0.4.6) was used to determine the optimal threshold for dichotomizing patient risk scores (0.01207 for the microarrays cohorts and 0.44101 for the NanoString cohort). The risk score was fit in both the univariate setting and in
- 15 the multivariate setting where appropriate clinical covariates were included, depending on the cohort (Table 5). Proportional hazard assumptions were tested with the R function cox.zph which assesses the correlation of survival time with the scaled Schoenfeld residuals of each variable; a p-value threshold of 0.05 was used to identify variables which failed the assumption. Finally, to assess clinical utility, receiver
- 20 operator characteristic curves were fit and the area under the curve (AUC) was measured with the survivalROC package (v1.0.3)²⁷. The survIDINRI package (v1.1-1) was used to calculate the net reclassification index with 5000 permutations and alpha of 0.01²⁸. Data visualization was performed using the BPG package (v5.3.2; P'ng *et al.* submitted).
- 25 Table 5: Clinical covariates and stratification used in multivariate models per cohort.

	Taylor	Hieronimus	Ross-Adams	Stockholm	Arrays-Combined	CPC-GENE
Gleason score	5-6 vs. 7 vs. 8-9	5-6 vs. 7 vs. 8-9	5-6 vs. 7 vs. 8-9	5-6 vs. 7 vs. 8-9	5-6 vs. 7-9	5-6 vs. 7-8
T category	T1-2 vs. T3	T1 vs. T2-3	T1-2 vs. T3	T1 vs. T2-3	T1 vs. T2-3	T1 vs. T2
PSA	Stratified at 10ng/mL	<10 ng/mL vs. ≥ 10ng/mL	<10 ng/mL vs. ≥ 10ng/mL	<10 ng/mL vs. ≥ 10ng/mL	<10 ng/mL vs. ≥ 10ng/mL	<10 ng/mL vs. ≥ 10ng/mL

Results and Discussion

Signature reduction

We obtained four cohorts of men diagnosed with localized disease with complete clinical annotation including long-term follow-up information (Table 1). To reduce the original 100-locus genomic classifier to a size amenable to routine clinical testing, we used the Taylor cohort to find loci where the mRNA abundance differs between patients with different copy number states. A recent study has shown that this feature selection strategy is effective in identifying prognostic signatures¹⁶, and thus, focusing on CNAs which have a cis-regulatory effect on mRNA abundance can select for functional CNAs. Using an FDR threshold of 10%, we identified 37 genes where the mRNA abundance reflected the CNA state (Figure 12). All genes except for *CTP1B* showed changes in mRNA abundance concordant with the change in copy-number, thus *CTP1B* was excluded from downstream analyses. Of the original 100 genomic classifier loci, we selected those that contained at least one of the 36 genes (n = 31 loci), resulting in a reduced genomic classifier consisting of 109 genes (Figure 1A).

After re-training a random forest with the 31 genomic regions (*i.e.* the 31-locus genomic classifier) in the original Toronto-aCGH cohort¹², we find that the 96th loci from the original signature, and the 31st loci from the reduced signature, has the highest Gini score in the 31-locus genomic classifier (Figure 13). This locus contains the genes *MVD*, *ZC3H18*, *IL17C*, and *CYBA*. The Gini score represents the relative importance of each feature in accurately predicting patient outcome in the context of the overall model. The next most important loci in the reduced genomic classifier are *ZNF862*

from chromosome 7 and *RNLS* from chromosome 10. Chromosome 8 still has the most features in the genomic classifier with 11/31 features; however ten chromosome 8 loci were not selected for the reduced signature (Table 6).

Table 6: Signature loci per chromosome in full and reduced genomic classifiers.

Chromosome	Full genomic classifier (n = 100)	Reduced genomic classifier (n = 31)
1	1	0
3	1	0
5	13	2
7	3	1
8	21	11
9	1	0
10	13	2
11	4	2
13	3	1
16	4	1
17	7	4
19	1	0
20	19	6
22	9	1

5

We next sought to see whether the reduced genomic classifier maintained its ability to act as a proxy for global genomic instability (as measured with PGA). Indeed, the Spearman correlation of signature estimated PGA with global PGA is 0.64 ($p < 2.2 \times 10^{-16}$), and when the previously selected 30 PGA genes were added, the correlation increases to 0.82 ($p < 2.2 \times 10^{-16}$; Figure 1B). This is despite the fact that the original

10

signature contained information from 14 chromosomes compared to only 10 chromosomes in the reduced signature.

Validation of the 31-Locus Genomic Classifier

We first tested the 31-locus genomic classifier in the Taylor and Ross-Adams cohorts; these were used for validating the original 100-locus genomic classifier. We also applied both genomic classifiers to the Hieronymus and Stockholm cohorts, and finally, to all four cohorts combined ("Combined-arrays cohort"). The median time to BCR for the four cohorts ranges from 3.0-7.8 years, allowing for evaluation at 10 years and later (Table 1). The performance of the 31-locus genomic classifier for predicting outcome at 18-months and 5-year bRFR are in Table 7. In the Taylor and Ross-Adams cohorts, we see similar classification scores to the original 100-locus genomic classifier (Pearson correlations 0.85 – 0.95, $p < 0.001$; Figure 14). The hazard ratio (HR) for the continuous signature score adjusted for clinical variables ranges from 2.71 to 8.23 (Table 8). Threshold analysis indicated that a threshold of 0.01027 is the optimal choice for discretizing patients into two groups based on the signature risk score; this resulted in adjusted HRs from 1.75 – 4.30 (Table 7, Figure 15). Of note, the dichotomized genomic classifier score was not prognostic in the Ross-Adams cohort but this is likely at least partially due to the short follow-up time for this cohort and the choice of dichotomization threshold. Indeed, when predicting outcome at 18 months, the continuous genomic classifier score is predictive of BCR (Figure 15D and Tables 7-8). Different thresholds may be optimal for different cohorts due to the difference in CNA rates observed per platform (Figure 16).

Table 7: Cox proportional hazard models for the 31-locus genomic classifier using 18-month and 5-year BCR endpoints. A model was fit for the dichotomized genomic classifier risk score for each individual cohort and for the Combined-Arrays cohort. Models were also fit for each NCCN risk group using the patients from the Combined-Arrays cohort only. Each model was fit twice, one predicting for 18-month relapse and another for 5-year relapse.

* Adjusted for Gleason score, T-category and PSA as per Table 5

Cohort	18-month bRFR				5-year bRFR			
	HR	Lower 95% CI	Upper 95% CI	Wald P	HR	Lower 95% CI	Upper 95% CI	Wald P
Taylor*	1.27	0.416	3.87	0.676	3.41	1.47	7.91	0.00423
Ross-Adams*	2.44	0.651	9.15	0.185	1.75	0.674	4.47	0.253
Hieronymus*	4.79	1.41	16.3	0.0122	4.30	1.69	10.9	0.0840
Stockholm*	2.19	0.684	7.00	0.187	2.01	0.910	4.45	0.0264
Combined Arrays* (n = 461)	2.41	1.37	4.22	0.00215	2.68	1.78	4.04	2.38 x 10 ⁻⁶
Low-risk patients	6.79	0.794	58.2	0.0803	7.18	2.08	24.8	0.00183
Intermediate- risk patients	1.72	0.772	3.82	0.185	2.21	1.21	4.05	0.00974
High-risk patients	2.57	1.11	5.95	0.0276	2.04	1.10	3.80	0.0236
CPC-GENE NanoString*	1.59	0.540	4.67	0.400	2.78	1.47	5.27	0.00173

Table 8: Cox proportional hazard models for the reduced genomic classifier using a continuous risk-score and 18-month and 5-year BCR endpoints. A model was fit for the continuous genomic classifier risk score for each individual cohort and for the Combined-Arrays cohort. Models were also fit for each NCCN risk group using the patients from the Combined-Arrays cohort only. Each model was fit twice, one predicting for 18-month relapse and another for 5-year relapse.

* Adjusted for Gleason score, T-category and PSA as per Table 5.

Cohort	18-month bRFR				5-year bRFR			
	HR	Lower 95% CI	Upper 95% CI	Wald P	HR	Lower 95% CI	Upper 95% CI	Wald P
Taylor*	4.00	0.678	23.4	0.126	7.08	2.24	22.4	0.000876
Ross- Adams*	11.9	1.42	99.2	0.0224	5.62	0.909	34.8	0.0633
Hieronymus*	2.65	0.478	14.7	0.265	3.35	0.874	12.8	0.0779
Stockholm*	2.03	0.471	8.75	0.342	2.71	0.867	8.48	0.0863
CPC-GENE NanoString*	3.22	0.271	38.3	0.354	8.23	1.74	38.9	0.00777

Finally, we combined the Taylor, Hieronymus, Ross-Adams, and Stockholm cohorts (termed "Combined-arrays cohort") and evaluated genomic classifier performance for all 461 patients and within each NCCN risk group separately. A multivariate Cox proportional hazard model estimates that the 31-locus genomic classifier has a HR adjusted for GS, pre-treatment PSA and clinical T-category of 2.6 (95% confidence interval (CI) 1.8-3.8, $p = 1.61 \times 10^{-6}$, Wald test; Figure 2A; Table 2) in comparison to the 100-locus genomic classifier (adjusted HR of 2.7; 95% CI 1.8 – 3.9, $p = 2.85 \times 10^{-7}$, Wald test). Based on the HR and Akaike information criterion (AIC 1441 and 14373 for the 100-locus and 31-loci genomic classifiers, respectively), both models fit the data well. The HR for the 31-locus genomic classifier is consistent when only considering only patients with low to intermediate risk disease (Figure 2B). Note the 31-locus genomic classifier is particularly effective for low-risk patients, where the 57% (80/141) of patients classified as lower risk by the genomic classifier have a 5-year bRFR of 96.0% in comparison to 73.5% for the low-risk patients predicted to relapse (HR = 5.48 (95% CI 2.01-15.0), $p = 8.92 \times 10^{-4}$, Wald test; Table 7 Figure 17). Similarly, the genomic classifier detects a subset of high-risk patients that fail primary treatment within 18 months (HR 2.57 (95% CI 1.11-5.95), $p = 0.0276$, Wald test; Table 7 Figure 2C); this PSA-based outcome endpoint is a surrogate for prostate-cancer specific mortality^{29,30}.

Our 31-locus genomic signature identified patients at risk of metastasis (HR 13.9 (95% CI 1.77-208), $p = 0.0120$, Wald test; Figure 2D; Table 9. This is consistent with our previous demonstration that the 100-locus genomic classifier identified patients failing treatment within 18 months (a surrogate for prostate-cancer specific mortality)^{12,29,30}.

Although the confidence interval on the HR is large, we also note that the estimated 10-year metastasis-free rate for the 62 patients classified negatively by the 31-locus genomic classifier is 98.0% compared to 68.2% for the 92 patients classified as positive for the 31-locus genomic classifier. The 31-locus genomic classifier correctly
 5 predicted which patients would develop metastasis across all clinical risk categories. These results will be verified in the remaining cohorts as their follow-up time is increased and more metastasis events are noted.

Table 2. The effect of including the 31-locus genomic classifier to the clinical model.

Patient prognosis is evaluated by 10-year bRFR in the Combined-arrays cohort.

- 10 A) Clinical model only; B) Multivariate clinico-genomic model using the 31-locus genomic classifier as a continuous variable ("31-locus score"); C) Multivariate clinico-genomic model using the 31-locus genomic classifier as a dichotomized risk group ("31 loci + vs. -").

A)

	HR	95% CI Low	95% CI high	Wald P
GS 7 vs. 5-6.	2.54	1.52	4.24	0.000364
GS 8-9 vs. 5-6.	3.74	1.96	7.15	6.70×10^{-5}
T3 vs. T1-2	2.03	1.08	3.80	0.0277
PSA ≥ 10 vs. <10	1.62	1.03	2.54	0.0368
Overall Model	Wald P	6.80×10^{-8}	logrank P	9.67×10^{-8}

15

B)

HR	95% CI Low	95% CI high	Wald P
----	------------	-------------	--------

31-locus score	3.53	1.69	7.35	0.000761
GS 7 vs. 5-6.	2.30	1.40	3.79	0.00103
GS 8-9 vs. 5-6.	2.44	1.27	4.71	0.00780
T3 vs. T1-2	2.61	1.42	4.80	0.00207
PSA $\geq$ 10 vs. <10	1.41	0.888	2.23	0.145
Overall Model	Wald P	5.90×10^{-11}	logrank P	1.46×10^{-9}

C)

	HR	95% CI Low	95% CI high	Wald P
31 loci + vs. -	2.09	1.37	3.17	0.000602
GS 7 vs. 5-6.	2.36	1.44	3.88	0.000704
GS 8-9 vs. 5-6.	2.36	1.22	4.58	0.01073
T3 vs. T1-2	2.60	1.41	4.78	0.00217
PSA $\geq$ 10 vs. <10	1.52	0.964	2.41	0.0714
Overall Model	Wald P	2.35×10^{-9}	logrank P	1.51×10^{-8}

Table 9: Multivariate models for 10-year metastasis-free survival in the Taylor cohort.

- 5 A- A Cox proportional hazard regression for patients predicted to develop metastasis by the reduced 31-loci genomic classifier ("Signature +"). CI: confidence interval, HR:

hazard ratio

	HR	95% CI lower	95% CI upper	Wald P
Signature +	13.9	1.77	108	0.0122
GS7	1.72	0.544	5.45	0.355
GS8-9	3.47	0.929	12.9	0.0644
T3	3.68	1.14	11.8	0.0289
Full Model	Wald P	0.0111	logrank P	0.000439

Assessment of clinical utility for a 31-locus clinico-genomic classifier

To assess clinical utility within clinical prognostic groups, we evaluated area under the survival receiver operator curve (AUC). The 31-locus clinico-genomic classifiers applied to the Combined-arrays cohort have 5-year AUCs 0.73 and 0.72 for the classifiers adjusted for GS/PSA/T and NCCN risk group, respectively (Figure 3A). In the four individual cohorts (predicted at 18 months for the Ross-Adams cohort), these values range from 0.68 – 0.84 (Figure 18). Indeed, the clinico-genomic models are better at predicting outcome than the clinical models (AUC increase of 0.0756 ($p = 0.0111$) and 0.0717 ($p = 0.00745$) for the 31-locus + risk clinico-genomic classifier and the 31-locus + GS/PSA/T clinico-genomic classifier, respectively). These increases in AUC also hold within each NCCN risk group, illustrating that even within tight clinical cohorts, the reduced genomic classifier is useful in addition to standard clinical covariates (Figure 19).

In some cases AUC has limited utility within homogeneous clinical risk groups and it is difficult to assess whether a statistically significant increase in AUC is clinically meaningful²⁸. To further understand the clinical utility of the 31-locus genomic classifier, we calculated the net reclassification index for the 31-locus + GS/PSA/T clinico-genomic classifier in the Combined-arrays cohort. This revealed that 53.1% of patients with BCR, and 32.2% of patients without BCR had their risk score increased by this clinico-genomic classifier in comparison to using a clinical-only model with GS/PSA/T (overall net-reclassification index = 0.201; 95% CI: 0.122-0.303, $p = 0.00200$; Figure 3B). Finally, we see an increase in the likelihood of BCR with

increasing risk score from the 31-locus + GS/PSA/T clinico-genomic classifier, ranging from <5% for the patients with very low risk scores to >50% for the highest risk scores (Figure 20).

In the metastatic setting, the clinical utility of the clinico-genomic classifiers is even more pronounced. The AUC for the 31-locus clinico-genomic classifiers are 0.89 and 0.85 compared to 0.79 and 0.75 for the NCCN risk group and the GS/PSA/T classifiers, respectively, indicating increased clinical utility in guiding patient management compared to clinical standards (Figure 3C). Indeed, the net reclassification index is 0.32 (95% CI -0.08-0.55; $p = 0.16$), with the majority (60%) of patients who go on to develop metastasis having their risk score increased (Figure 3D).

Validation of 31-locus genomic classifier using a NanoString platform

Finally, we applied the 31-locus genomic classifier to the CPC-GENE cohort, a tightly defined clinical cohort, comprised primarily of intermediate-risk patients (95/102) where all but 11 patients were diagnosed with GS 7 disease (Table 1). Of the standard clinical variables, clinical T and pre-treatment PSA were statistically significant (Figure 8), and only T-category remains prognostic a multivariate Cox proportional hazard model (Table 10).

Table 10: Multivariate Cox proportional hazard regression model for the clinical variables in the CPC-GENE cohort for 10-year BCR-free progression.

	HR	CI95low	CI95high	Wald P
GS7 vs GS6	1.01	0.352	2.90	0.985
T2 vs T1	1.95	1.06	3.59	0.0314
PSA < 10 vs ≥ 10	1.36	0.708	2.66	0.361
Full Model	Wald P	0.119	logrank P	0.101

The genomic classifier was measured in this cohort using the NanoString CNA platform with 1-3 custom probes per region. This is the first time this genomic classifier

has been tested using a clinically-feasible approach. Six patients have technical replicates, and we observed high replicate concordance (range 0.72 – 0.94 for discretized CNAs, Figure 21). We find that the genomic classifier is able to stratify patients within this clinically-homogeneous cohort. Five years after surgery, 70.4% of patients predicted by the signature to fail have indeed failed compared to 28.5% of patients predicted to have good prognosis (HR = 2.46 (1.36-4.43), P = 0.0028, Wald test; Figure 4; Table 3). Furthermore, in multivariate analysis with clinical variables, the 31-locus classifier is the strongest predictor and only significant variable within the model, with an adjusted HR of 6.41 per unit increase of the continuous risk score. Indeed, the survival AUC for the clinical model is only 0.62 but increases to 0.72 after addition of the 31 loci, indicating a clear benefit to considering the genomic classifier along with standard clinical variables.

Table 3. The continuous risk score from the 31-locus genomic classifier applied to the CPC-GENE cohort in a multivariate Cox proportional hazard models for 10-year biochemical recurrence-free survival.

	HR	95% CI Low	95% CI High	Wald P
31-locus risk score	6.41	1.53	26.9	0.0111
GS7 vs. GS6	0.988	0.334	2.91	0.982
PSA $\geq$ 10 vs. >10	1.22	0.616	2.43	0.563
T2 vs. T1	1.83	0.987	3.38	0.0549
Overall Model	Wald P	0.0165	logrank P	0.0131

Discussion

Our 100-locus genomic classifier was the first DNA-based multi-locus gene signature proposed for stratification of localized prostate cancer patients¹². Here, we refine this

genomic classifier to 31 loci showing DNA-RNA associations, show its utility in three public datasets comprising 461 men with localized disease, and validate the genomic classifier in a tightly defined clinical risk group using a clinically-relevant technology. The NanoString CNV platform requires only 300 ng DNA, making it achievable for routine use after patient biopsy. To our knowledge, this is the first study using NanoString's CNV platform to measure a multi-locus gene signature.

The 31-locus genomic classifier is able to sub-stratify patients across all NCCN risk groups. Patient management could therefore integrate both clinical risk group and our genomic classifier score. Compared to using the GS/PSA/T clinico-pathological variables, the additional use of our 31-locus genomic classifier is more accurate. Based on Cox proportional hazard models, the smallest impact for the genomic classifier is in the intermediate-risk group which highlights the difficulty in further stratifying the risk for these patients; to improve predictions, information from other molecular or microenvironmental information may be required^{12,16}. Interestingly, in the primarily intermediate-risk CPC-GENE cohort measured with the NanoString CNV platform, the performance of the clinico-genomic classifier is considerably better than the clinical classifier (AUC 0.61 vs. 0.72). The 31-locus genomic classifier is particularly promising in the low-risk group, where patients identified as good prognosis by the signature have 10-year bRFR of 87% and are candidates for treatment de-escalation trials (Figure 17). The 31-locus genomic classifier is also highly effective in the high-risk group, where it identifies patients failing rapidly (within 18 months) who might benefit from more aggressive initial treatment targeting occult metastases and prevent disease progression. Future work will aim to further decrease the false negative rate (thereby increasing the sensitivity) to provide more assurance to patients opting for increased treatment. On the other hand, patients classified as good prognosis by the clinico-genomic classifier can be re-assured that decreased or standard treatment schemes are sufficient for management of their prostate cancer.

Finally, we also demonstrate that this 31-locus genomic classifier is able to predict which patients will go on to develop metastases. Remarkably, the 10-year metastasis-free rate for patients stratified as good prognosis by the genomic classifier is only 98% compared 68% for those predicted to be high-risk. The finding will need to be validated in larger cohorts including those used in this study once their follow-up time is sufficient for time to metastasis analysis. This is the first multi-gene DNA signature to

predict metastasis in prostate cancer. PGA has also been shown to predict for metastasis in this same cohort¹⁹, but is difficult to measure in the clinic as it requires CNA estimation for the entire genome. We previously showed that our 100-locus genomic classifier combined with 30 selected genes can estimate global PGA¹²,
5 providing two prognostic indices in a single genomic assay. Now, we have validated this finding using the CLIA-certified NanoString platform with the reduced 31-locus genomic classifier in combination with the same 30 PGA-associated genes. The 31-locus genomic classifier has comparable performance to the previously published 100-locus genomic classifier. The two signatures should be further evaluated and
10 contrasted in prospective trials using the NanoString assay for the 31-locus classifier and an array platform such as OncoScan for the 100-locus classifier.

Some prognostic molecular signatures are currently available to patients in the clinic such as Oncotype DX Genomic Prostate Score¹¹, ProMark³¹, Prolaris⁹ and Decipher¹⁰. The first two signatures cater to low-risk patients and aim to identify the patients least
15 likely to progress on active surveillance protocols. The Prolaris signature is most similar to ours; it is intended for all patients with localized disease and is meant to be used in conjunction with clinical variables to identify patients for increased or decreased treatment compared to standard protocols. Finally, the Decipher classifier identifies higher risk patients that are likely to fail therapy and develop metastasis. It is
20 also used to identify patients that would benefit from adjuvant androgen deprivation therapy in conjunction with RadP³². These signatures are CLIA certified, but the results of large retrospective trials have yet to be reported³³. Our clinical genomic signatures have AUCs of 0.72-0.73 in the 461 patient cohort of low to high risk disease. This range is comparable to the AUCs of the signatures mentioned above (range 0.65 –
25 0.78), although a direct comparison of AUCs is not meaningful unless the signatures are measured on the same patient cohort. Nonetheless, this naïve comparison indicates that our reduced 31-locus genomic classifier can provide additional information relevant to disease management similar to that of clinically-available prognostic signatures. Future work will evaluate whether this biomarker can identify
30 patients that will derive benefit from adjuvant treatment.

There were several limitations to this work. First, the study is retrospective and thus future work will involve prospective trials to evaluate whether the genomic classifier could be used as a predictive biomarker as well. Additionally, despite the large overall

sample size (n=563), each cohort has a relatively small sample size and the genomic classifier was measured using four different platforms and pre-processing techniques. Commercializing the NanoString probe set will be useful to standardize measurement of the 31-locus genomic classifier in future cohorts. Finally, our study did not address
5 intra-prostatic genetic heterogeneity. Previous studies have shown that distinct cancer foci can have independent genetic origins, including differences in mutation of prognostic and putative driver genes³⁴. Future studies should measure the genomic classifier from various foci and determine whether this heterogeneity is related to patient outcome; if the signature is identified in one focus only, does the patient have
10 poor outcome? If so, how can we ensure that the appreciate regions of a prostate are biopsied for classifier evaluation? To address this issue of intra-prostatic heterogeneity, in future studies we will also establish the reliability of measuring the genomic classifiers from circulating DNA.

The embodiments of the present disclosure described above are intended to be
15 examples only. Alterations, modifications and variations to the disclosure may be made without departing from the intended scope of the present disclosure. In particular, selected features from one or more of the above-described embodiments may be combined to create alternative embodiments not explicitly described. All values and sub-ranges within disclosed ranges are also disclosed. The subject matter
20 described herein intends to cover and embrace all suitable changes in technology. All references mentioned are hereby incorporated by reference in their entirety.

Reference List

- 1 Draisma, G. *et al.* Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *Journal of the National Cancer Institute* **101**, 374-383 (2009).
- 5 2 Hong, S. K., Vertosick, E., Sjoberg, D. D., Scardino, P. T. & Eastham, J. A. Insignificant disease among men with intermediate-risk prostate cancer. *World journal of urology* **32**, 1417-1421 (2014).
- 3 Shao, Y. H. *et al.* Contemporary risk profile of prostate cancer in the United States. *J Natl Cancer Inst* **101**, 1280-1283, doi:10.1093/jnci/djp262 (2009).
- 10 4 Nichol, A. M., Warde, P. & Bristow, R. G. Optimal treatment of intermediate-risk prostate carcinoma with radiotherapy: clinical and translational issues. *Cancer* **104**, 891-905, doi:10.1002/cncr.21257 (2005).
- 5 Mohler, J. L. The 2010 NCCN Clinical Practice Guidelines in Oncology on Prostate Cancer. *J Natl Compr Canc Netw* **8**, 145 (2010).
- 15 6 D'Amico, A. V. *et al.* Cancer-specific mortality after surgery or radiation for patients with clinically localized prostate cancer managed during the prostate-specific antigen era. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **21**, 2163-2172, doi:10.1200/jco.2003.01.075 (2003).
- 20 7 Prensner, J. R., Rubin, M. A., Wei, J. T. & Chinnaiyan, A. M. Beyond PSA: The Next Generation of Prostate Cancer Biomarkers. *Science Translational Medicine* **4**, 127rv123, doi:10.1126/scitranslmed.3003180 (2012).
- 8 Fraser, M., Berlin, A., Bristow, R. G. & van der Kwast, T. in *Urologic Oncology: Seminars and Original Investigations*. 85-94 (Elsevier).
- 25 9 Cuzick, J. *et al.* Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *The lancet oncology* **12**, 245-255, doi:10.1016/s1470-2045(10)70295-3 (2011).
- 10 Erho, N. *et al.* Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PloS one* **8**, e66855, doi:10.1371/journal.pone.0066855 (2013).
- 30 11 Klein, E. A. *et al.* A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. *European urology* **66**, 550-560 (2014).
- 35 12 Lalonde, E. *et al.* Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. *The lancet oncology* **15**, 1521-1532 (2014).
- 13 Nielsen, T. *et al.* Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. *BMC cancer* **14**, 177 (2014).
- 40 14 Martin, M. *et al.* Prospective study of the impact of the Prosigna assay on adjuvant clinical decision-making in unselected patients with estrogen receptor-positive, HER2-negative, node-negative early-stage breast cancer. *Current medical research and opinion*, 1-28 (2015).
- 45 15 Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer cell* **18**, 11-22, doi:10.1016/j.ccr.2010.05.026 (2010).
- 16 Ross-Adams, H. *et al.* Integration of copy number and transcriptomics provides risk stratification in prostate cancer: A discovery and validation cohort study.

- EBioMedicine* **2**, 1133-1144, doi:10.1016/j.ebiom.2015.07.017 (2015).
- 17 Hieronymus, H. *et al.* Copy number alteration burden predicts prostate cancer relapse. *Proceedings of the National Academy of Sciences* **111**, 11139-11144 (2014).
- 5 18 Liu, W. *et al.* Identification of novel CHD1-associated collaborative alterations of genomic structure and functional assessment of CHD1 in prostate cancer. *Oncogene* **31**, 3939-3948 (2012).
- 19 Hieronymus, H. *et al.* Copy number alteration burden predicts prostate cancer relapse. *Proc Natl Acad Sci U S A*, doi:10.1073/pnas.1411446111 (2014).
- 10 20 Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences* **107**, 16910-16915 (2010).
- 21 Boutros, P. C. *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet* **47**, 736-745, doi:10.1038/ng.3315 (2015).
- 22 Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* **2**, 401-404 (2012).
- 15 23 Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* **6**, pl1-pl1 (2013).
- 24 Waggott, D. *et al.* NanoStringNorm: an extensible R package for the pre-processing of NanoString mRNA and miRNA data. *Bioinformatics* **28**, 1546-1548 (2012).
- 20 25 Breiman, L. Random Forest. *Machine Learning* **45**, 5-32 (2001).
- 26 Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association* **53**, 457-481 (1958).
- 25 27 Heagerty, P. J. & Zheng, Y. Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92-105 (2005).
- 28 Uno, H., Tian, L., Cai, T., Kohane, I. S. & Wei, L. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Statistics in medicine* **32**, 2430-2442 (2013).
- 30 29 Freedland, S. J. *et al.* Risk of prostate cancer-specific mortality following biochemical recurrence after radical prostatectomy. *JAMA : the journal of the American Medical Association* **294**, 433-439, doi:10.1001/jama.294.4.433 (2005).
- 35 30 Buyyounouski, M. K., Pickles, T., Kestin, L. L., Allison, R. & Williams, S. G. Validating the interval to biochemical failure for the identification of potentially lethal prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **30**, 1857-1863, doi:10.1200/jco.2011.35.1924 (2012).
- 40 31 Shipitsin, M. *et al.* Identification of proteomic biomarkers predicting prostate cancer aggressiveness and lethality despite biopsy-sampling error. *British journal of cancer* (2014).
- 32 Den, R. B. *et al.* Genomic classifier identifies men with adverse pathology after radical prostatectomy who benefit from adjuvant radiation therapy. *Journal of Clinical Oncology* **33**, 944-951 (2015).
- 45 33 Murphy, L., Prencipe, M., Gallagher, W. M. & Watson, R. W. Commercialized biomarkers: new horizons in prostate cancer diagnostics. *Expert review of molecular diagnostics* **15**, 491-503 (2015).
- 34 34 Boutros, P. C. *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet* (2015).
- 50

CLAIMS

1. A method for determining a risk of recurrence of cancer following a cancer therapy of a patient, comprising determining genomic instability of a tumour of the patient by:
 - 5 (a) obtaining a biopsy of the tumour;
 - (b) identifying genome regions of the biopsy wherein the regions are at least loci rankings 1-15 of the 31-loci in Table A;
 - (c) determining a plurality of copy number calls in the genome regions;
 - 10 (d) intersecting the plurality of copy number calls with a reference gene list, to obtain a plurality of Copy Number Alterations (CNA) calls for each gene;
 - (e) generating a CNA tumour profile based on the plurality of CNA calls;
 - (f) comparing the CNA tumour profile to a reference profile of recurring cancer patients and a reference profile of nonrecurring cancer patients;
 - 15 (g) calculating a plurality of statistical distances between the CNA tumour profile and the reference profile of recurring cancer patients and the reference profile of nonrecurring cancer patients;

wherein the statistical distance between the CNA tumour profile and the reference profile of recurring cancer patients and the reference profile of nonrecurring cancer patients is associated with the risk of cancer recurrence following the cancer therapy of the patient.

- 20
- 2. The method of claim 1, wherein the genome regions are at least loci rankings 1-20, 1-25, 1-31 of the 31-loci in Table A.
- 3. The method of claim 1, wherein the genome regions are a whole tumour genome.
 - 25

4. The method according to any one of claims 1-3, wherein the patient has been diagnosed with prostate cancer.
5. The method according to any one of claims 1-3, wherein the patient has been diagnosed with localized prostate cancer.
- 5 6. The method according to any one of claims 4 or 5, wherein the patient has one of a low or intermediate risk for prostate cancer.
7. The method according to claim 6, wherein the patient has one of a low or intermediate risk for prostate cancer as determined by at least one of T-category, Gleason score or pre-treatment prostate-specific antigen blood concentration.
- 10 8. The method according to any one of claims 6 or 7, wherein the low risk for prostate cancer is determined by at least one of the following:
 - (a) a T-category of T1-T2a, a Gleason score less than or equal to 6, and a pre-treatment prostate-specific antigen blood concentration less than or equal to 10 ng/mL;
 - 15 (b) a T-category of T1-T2a, a Gleason score greater than or equal to 2 and less than or equal to 6, and a pre-treatment prostate-specific antigen blood concentration less than or equal to 10 ng/mL; and
 - (c) a T-category of T1c, a Gleason score less than or equal to 6, a pre-treatment prostate-specific antigen blood concentration less than or equal to 10 ng/mL, and fewer than 3 biopsy cores of a tumour that are positive for cancer and having less than or equal to 50% cancer in each.
- 20 9. The method according to any one of claims 6 or 7, wherein the intermediate risk for prostate cancer is determined by at least one of the following:
 - (a) at least one of a T-category of T2b, a Gleason score equal to 7, and a pre-treatment prostate-specific antigen blood concentration greater than 10 ng/mL;
- 25

- 5 (b) at least one of a T-category of T1-T2, a Gleason score equal to or less than 7, and a pre-treatment prostate-specific antigen blood concentration less than or equal to 20 ng/mL;
- (c) at least one of a T-category of T2b, a Gleason score equal to 7 and a pre-treatment prostate-specific antigen blood concentration greater than 10 ng/ml and equal to or less than 20 ng/mL; and
- (d) at least one of a T-category of T2b, a T-category of T2c, a Gleason score equal to 7 and a pre-treatment prostate-specific antigen blood concentration greater than 10 ng/ml and equal to or less than 20 ng/mL.
- 10 10. A method, performed by at least one computing device, for determining the risk of recurrence of cancer following a cancer therapy of a patient, comprising determining genomic instability of a tumour of the patient based on:
- (a) determining, at a processor, a genome of the tumour;
- 15 (b) determining, by the processor, genome regions of the biopsy wherein the regions are at least loci rankings 1-15 of the 31-loci in Table A;
- (c) determining, by the processor, a plurality of copy number calls in the genome regions;
- 20 (d) determining, by the processor, a plurality of Copy Number Alternations (CNA) calls for each gene by intersecting the plurality of copy number calls with a reference gene list;
- (e) determining, by the processor, a CNA tumour profile based on the plurality of CNA calls;
- 25 (f) determining, by the processor, a plurality of statistical distances between the CNA tumour profile and a reference profile of recurring cancer patients and a reference profile of nonrecurring cancer patients;

wherein the statistical distance between the CNA tumour profile and the reference profile of recurring cancer patients and the reference profile of

nonrecurring cancer patients is associated with a risk of cancer recurrence following the cancer therapy.

11. A system for determining the risk of recurrence of cancer following a cancer therapy of a patient comprising determining genomic instability, the system comprising:

a non-transitory computer readable storage medium that stores computer-readable code;

a processor operatively coupled to the non-transitory computer readable storage medium, the processor configured to implement the computer-readable code, the computer-readable code configured to:

determine a genome of the tumour;

determine genome regions of the biopsy wherein the regions are at least loci rankings 1-15 of the 31-loci in Table A;

determine a plurality of Copy Number Alterations (CNA) calls for each gene based on intersecting the copy number calls with a reference gene and storing the plurality of CNA calls in the non-transitory computer readable storage medium;

determine a CNA tumour profile based on the plurality of CNA calls and storing the CNA tumour profile in a non-transitory computer readable storage medium;

determine a plurality of statistical distances between the CNA tumour profile and a reference profile of recurring cancer patients and a reference profile of nonrecurring cancer patients;

wherein the statistical distance between the CNA tumour profile and the reference profile of recurring cancer patients and the reference profile of nonrecurring cancer patients is associated with a risk of cancer recurrence following the cancer therapy.

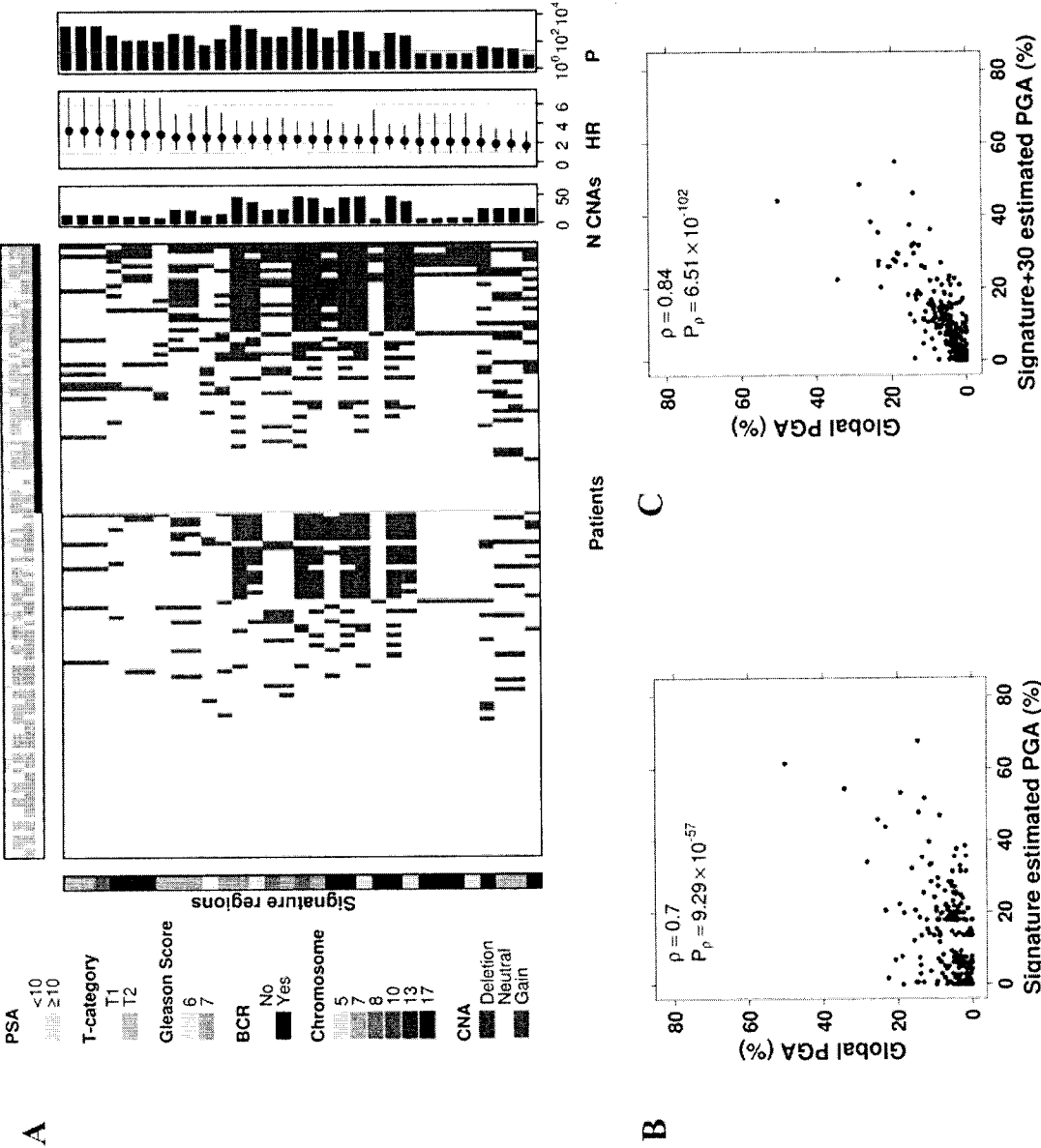


Figure 1

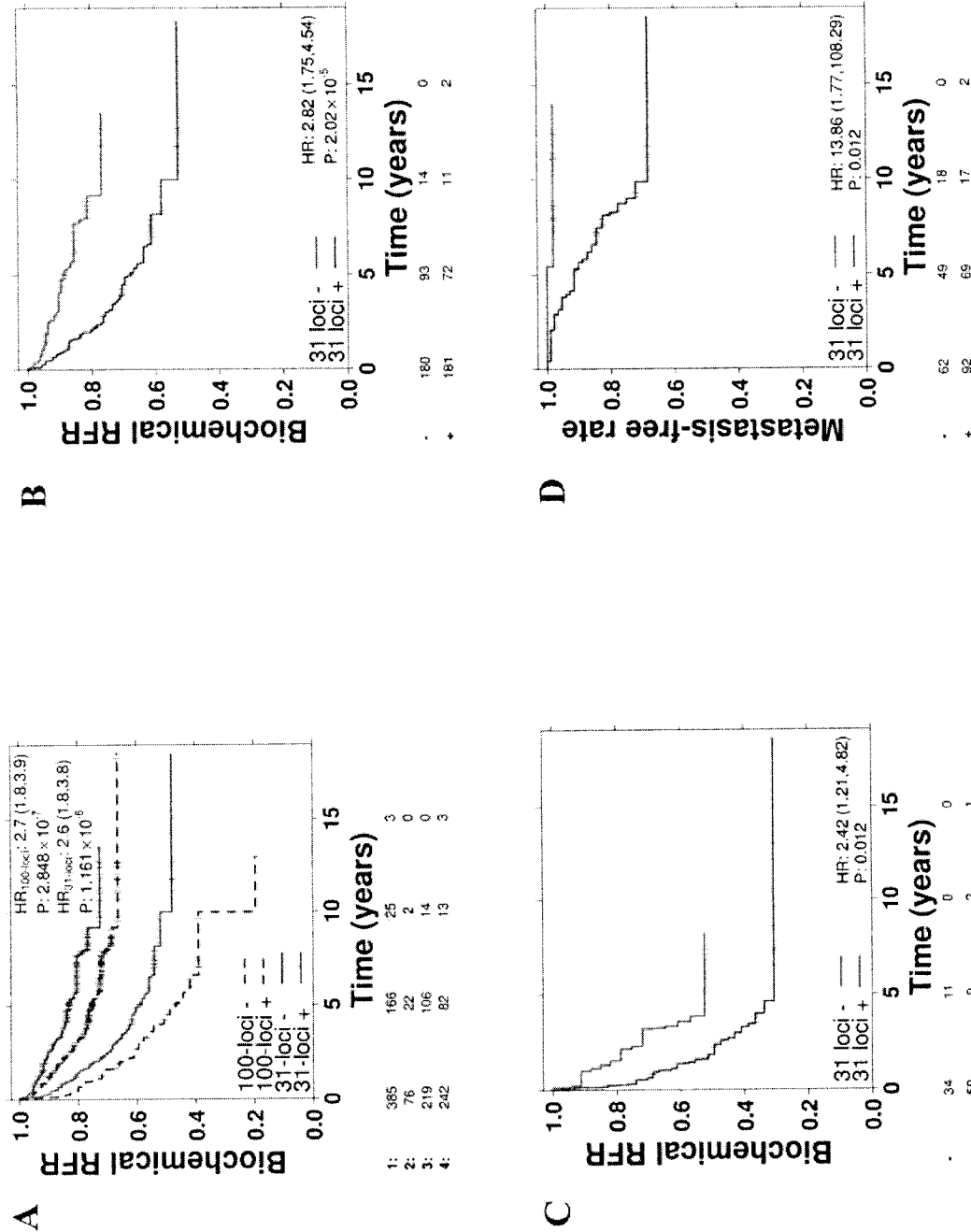


Figure 2

3/22

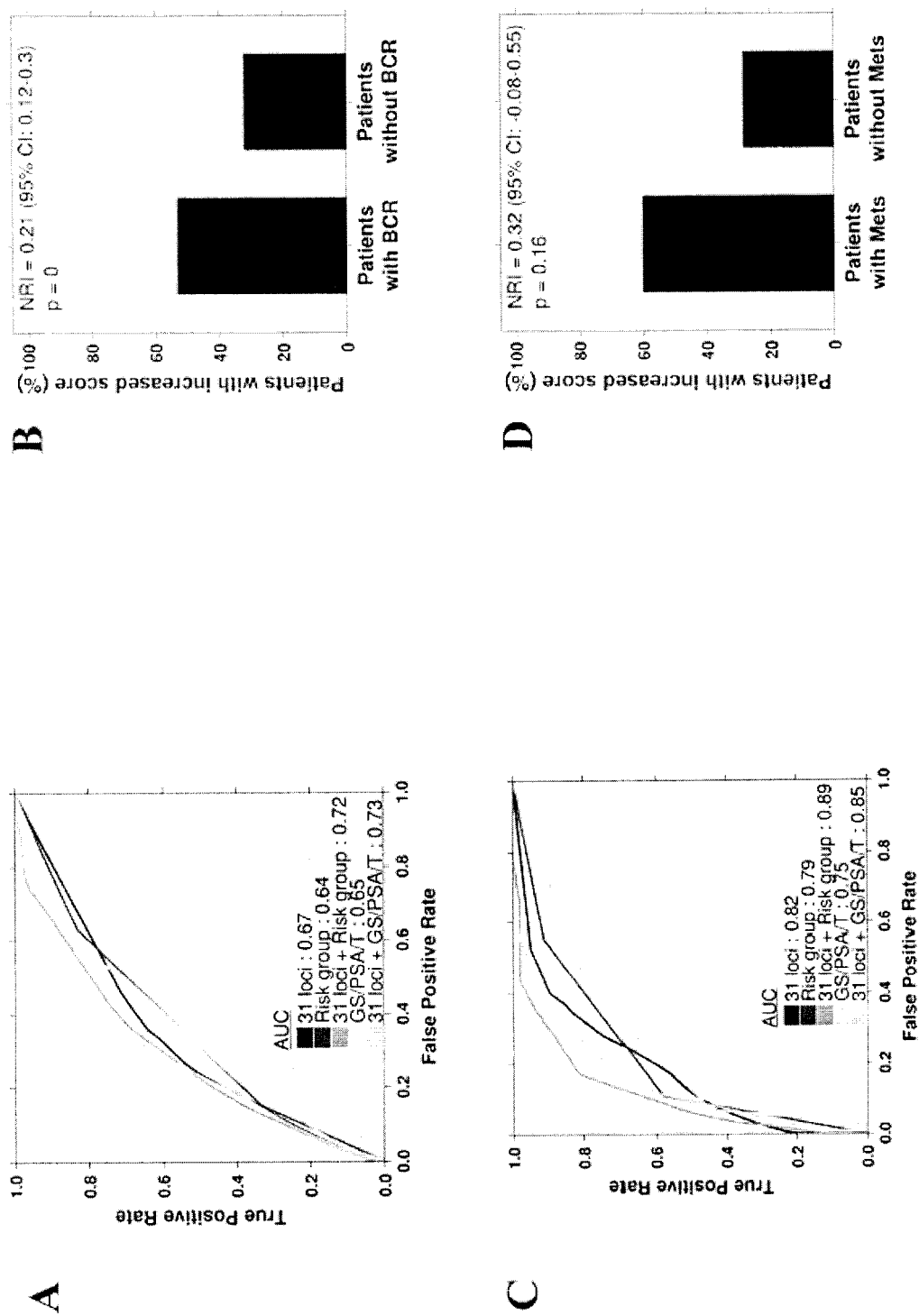


Figure 3

4/22

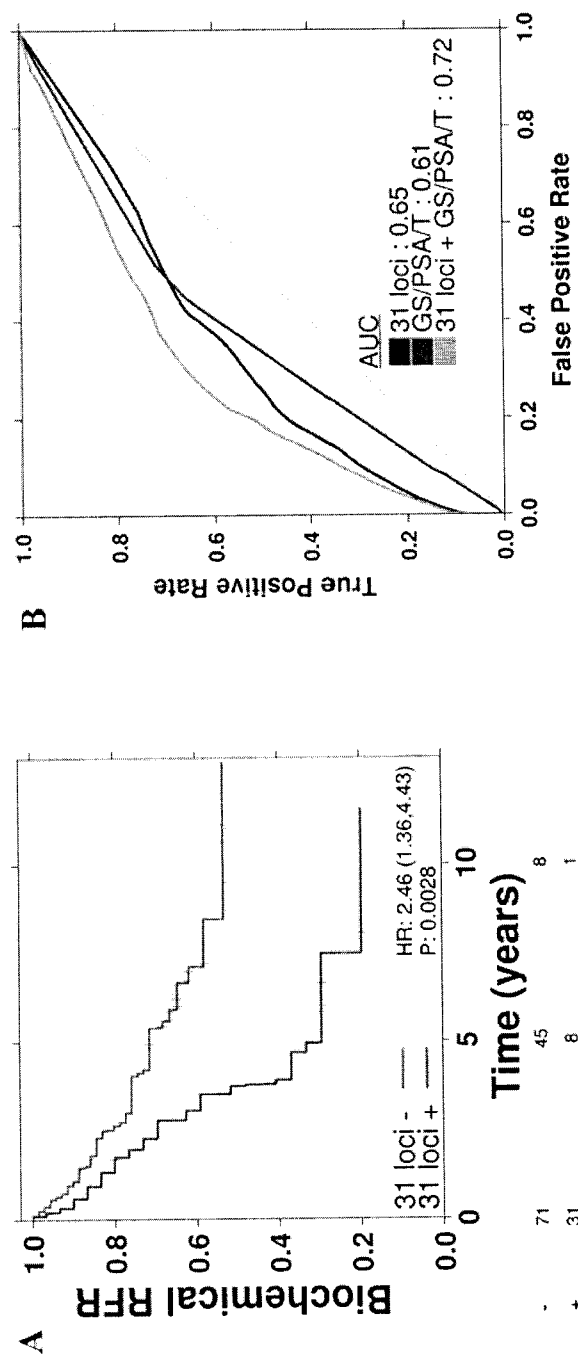


Figure 4

5/22

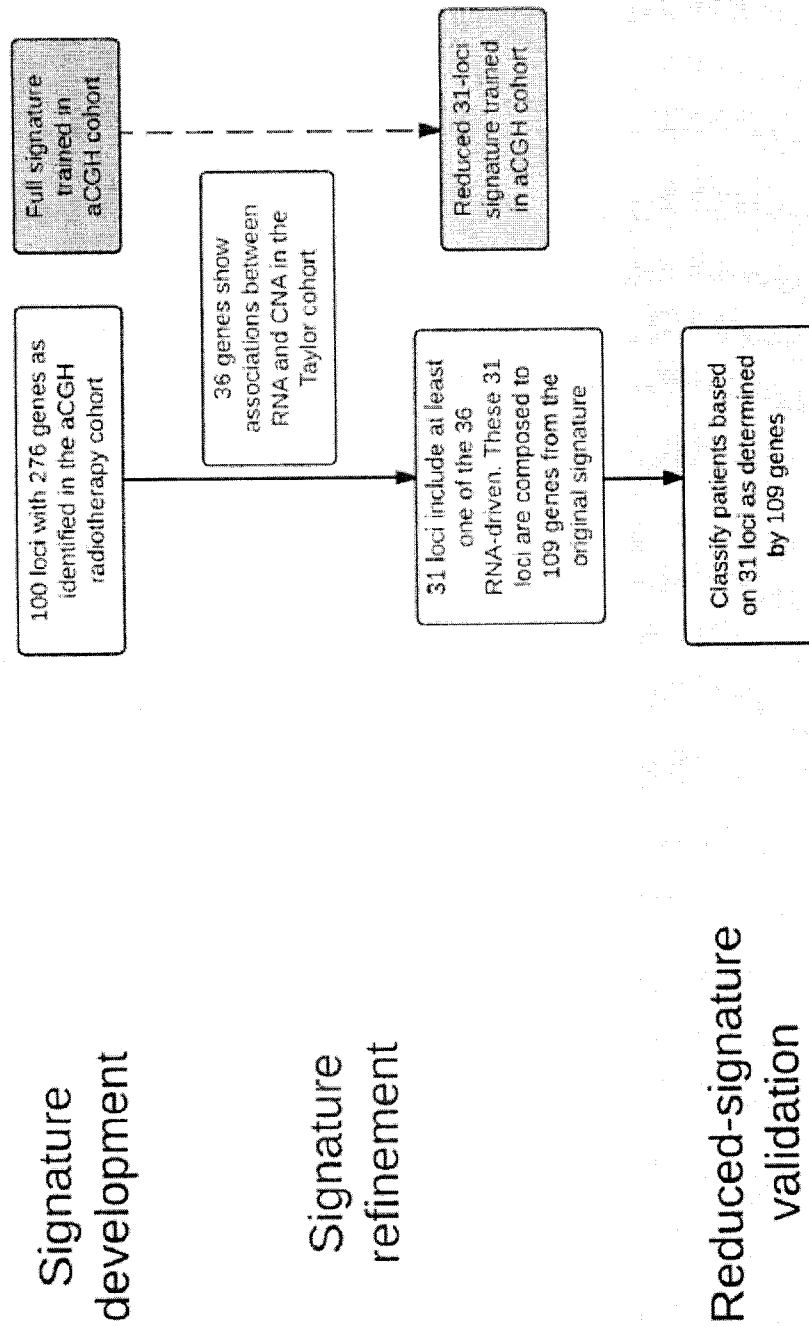


Figure 5

6/22

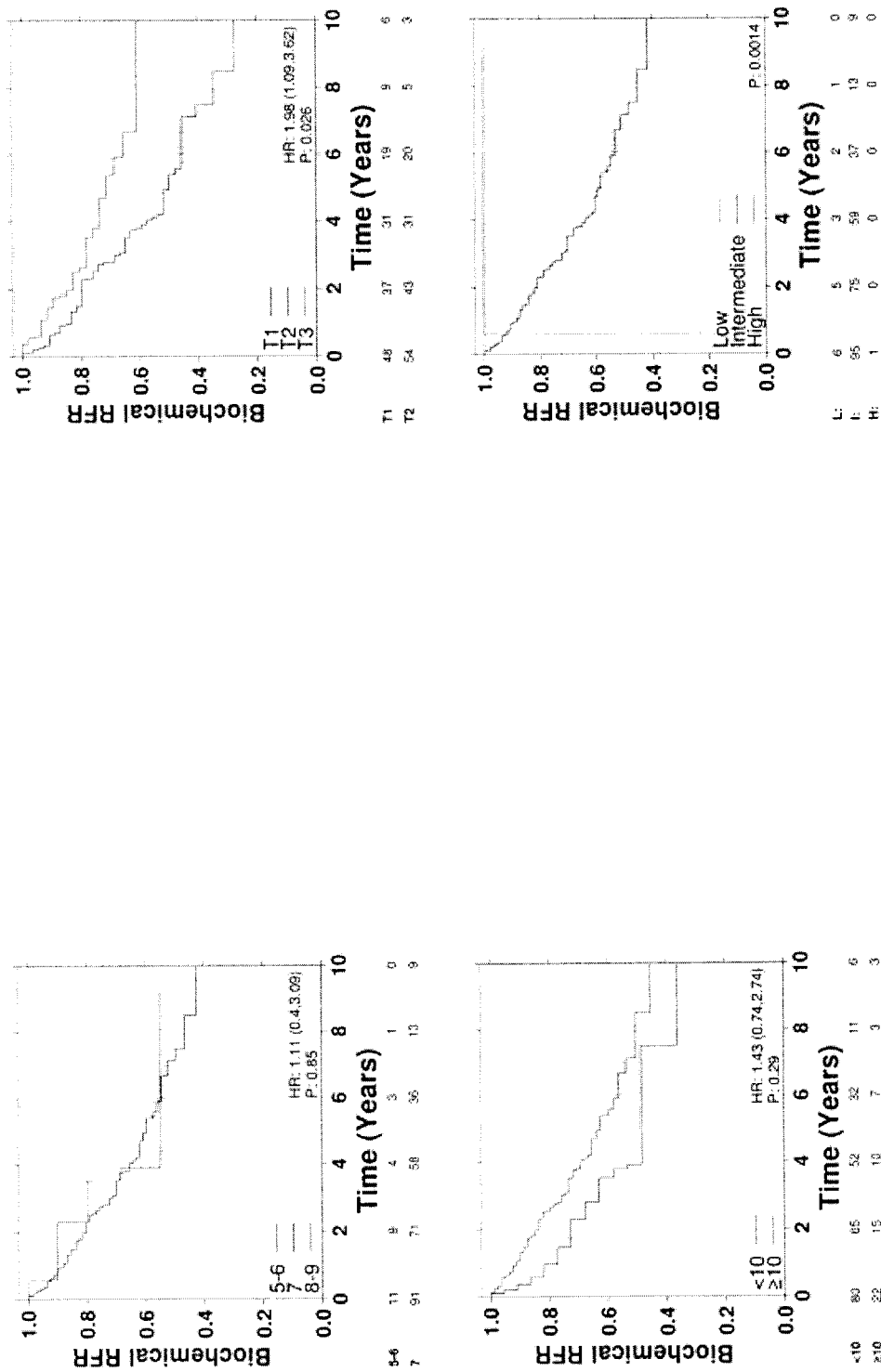


Figure 6

7/22

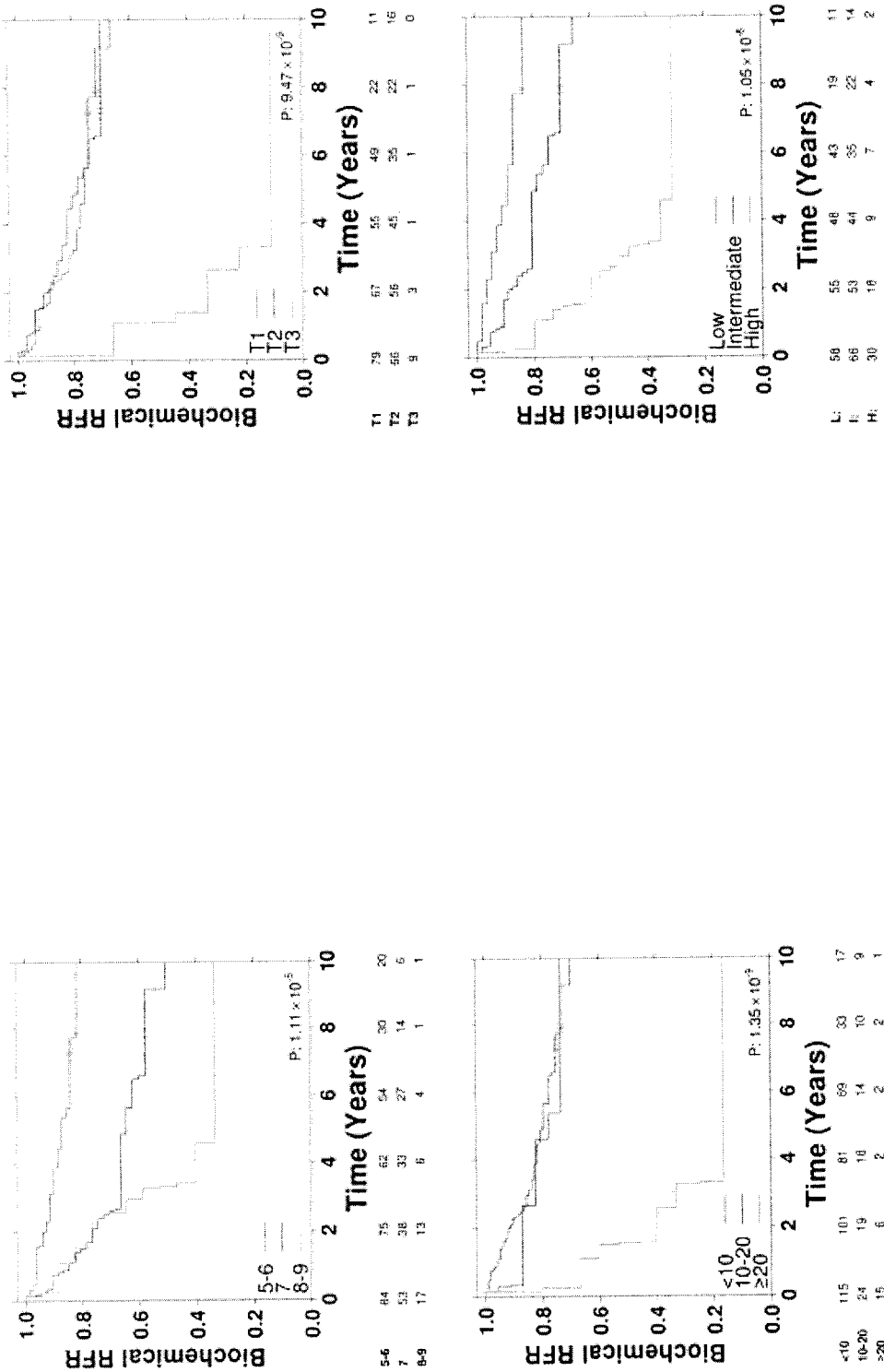


Figure 7

8/22

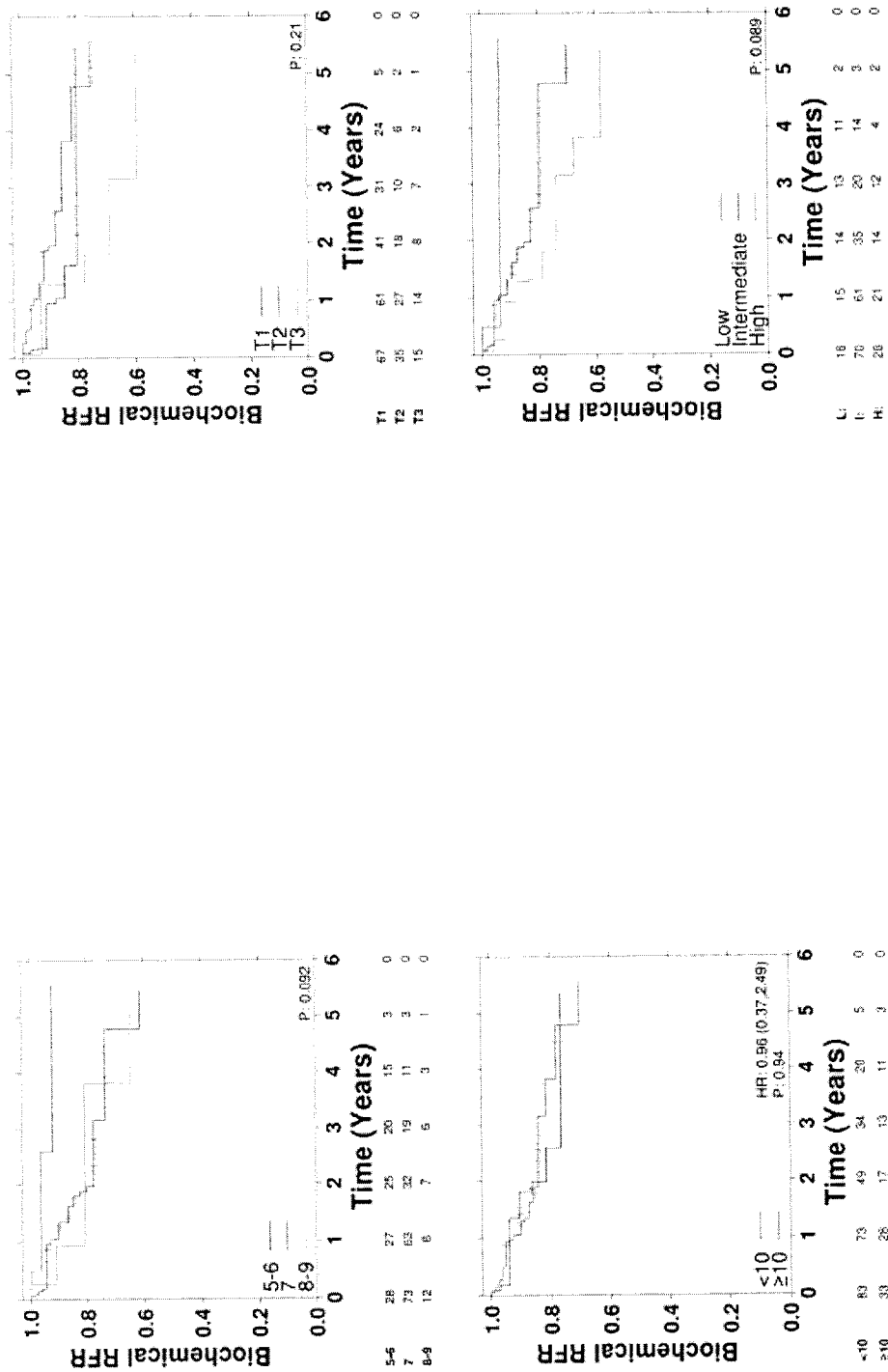


Figure 8

9/22

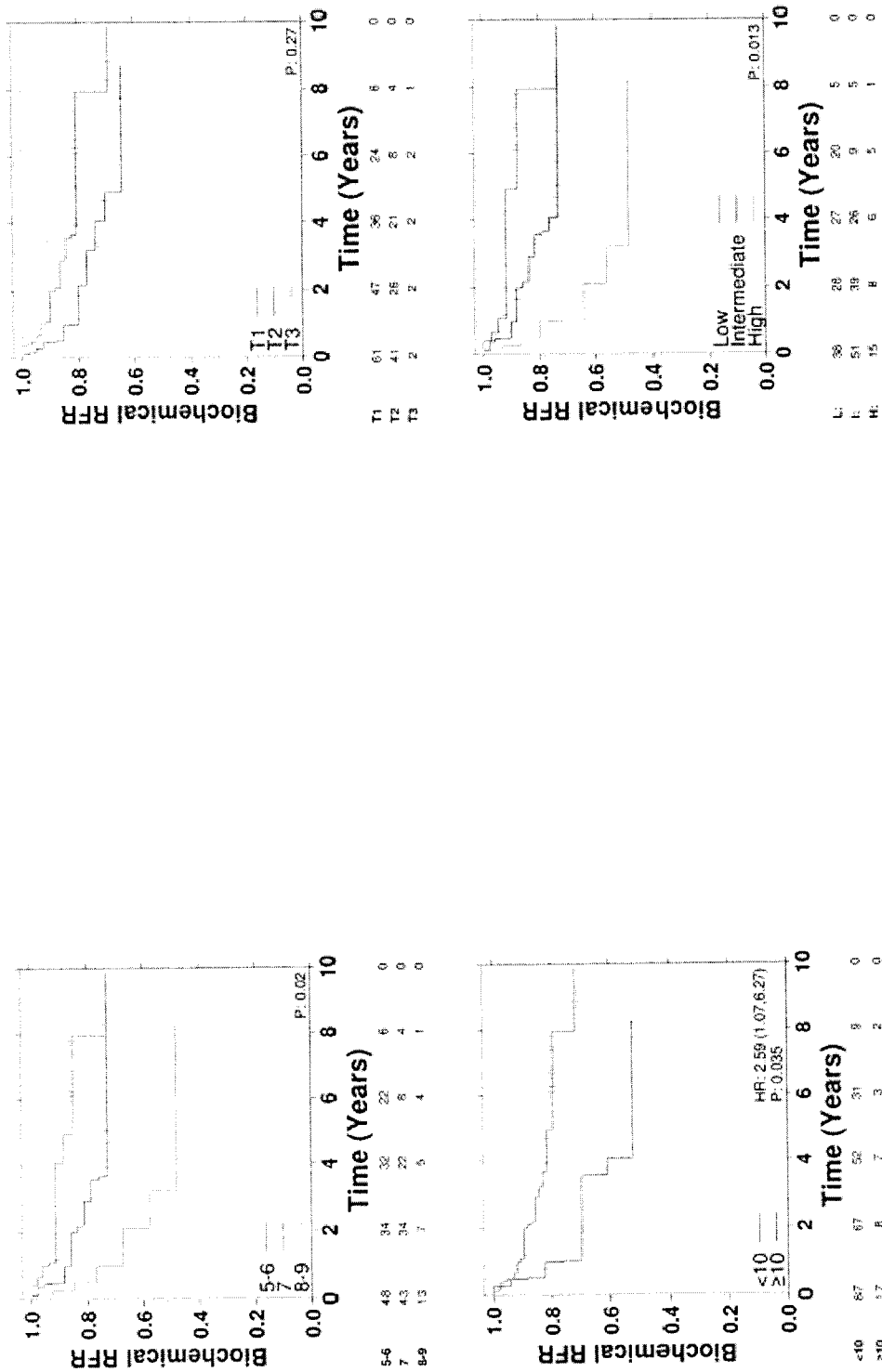


Figure 9

10/22

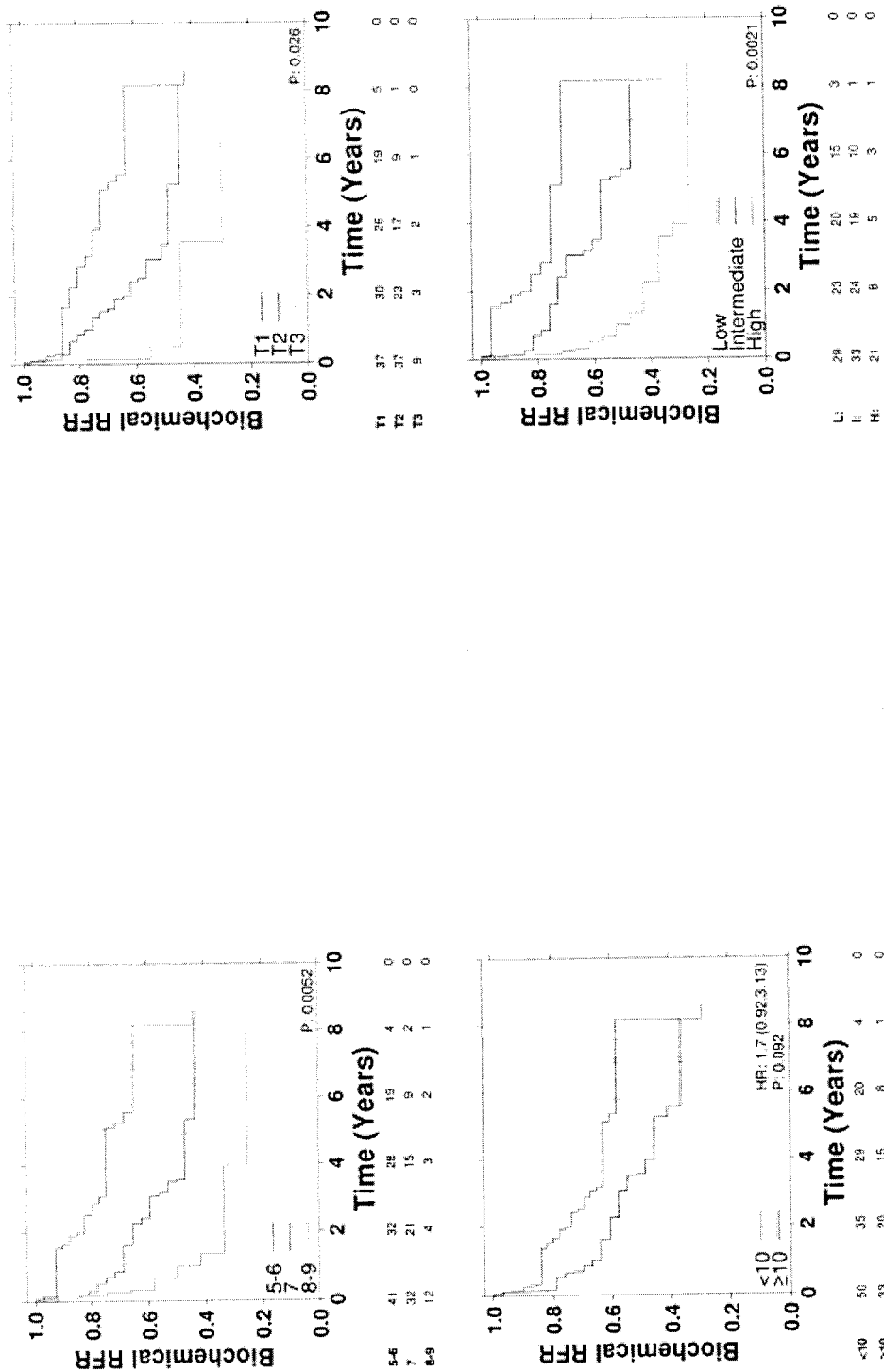


Figure 10

11/22

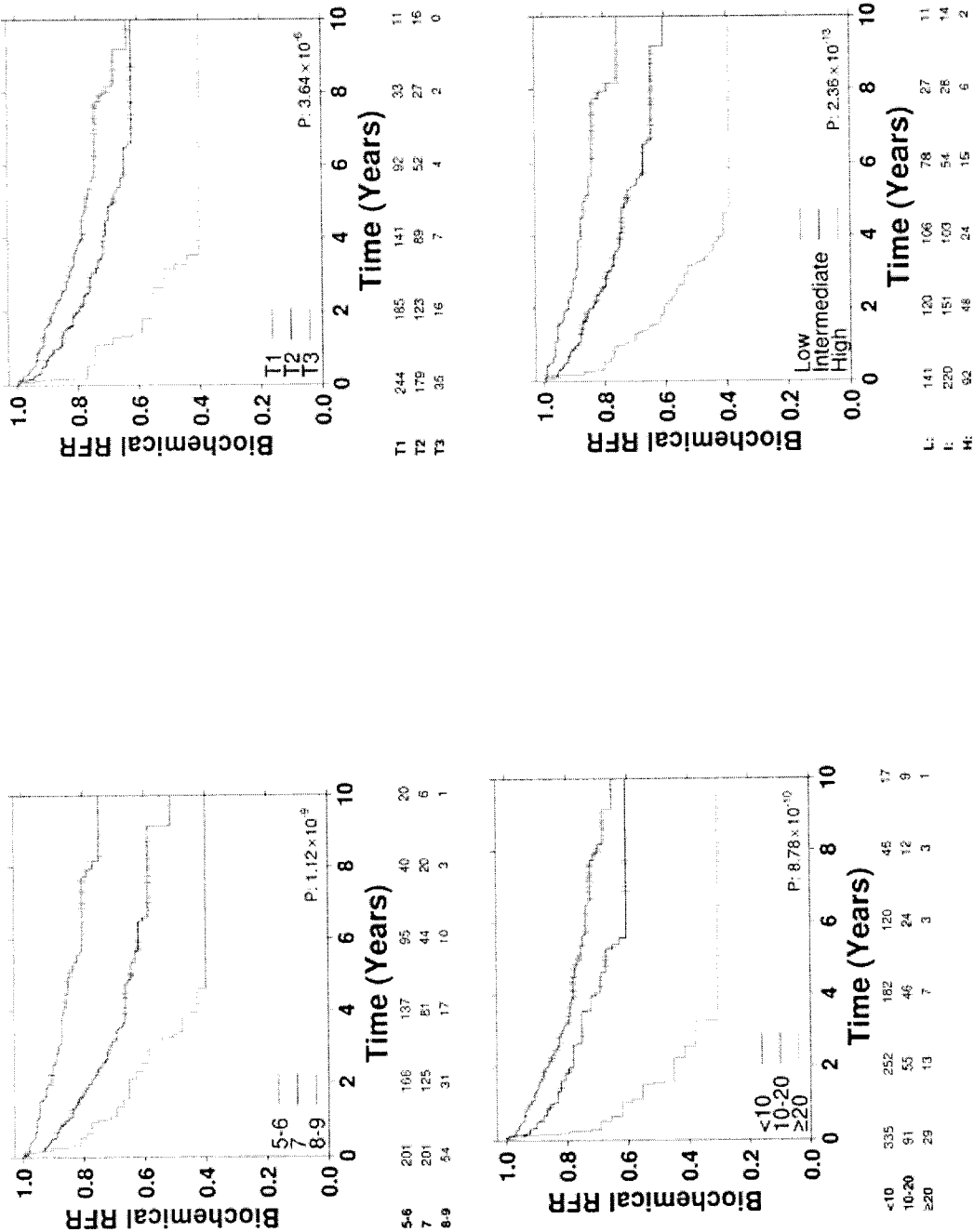


Figure 11

12/22

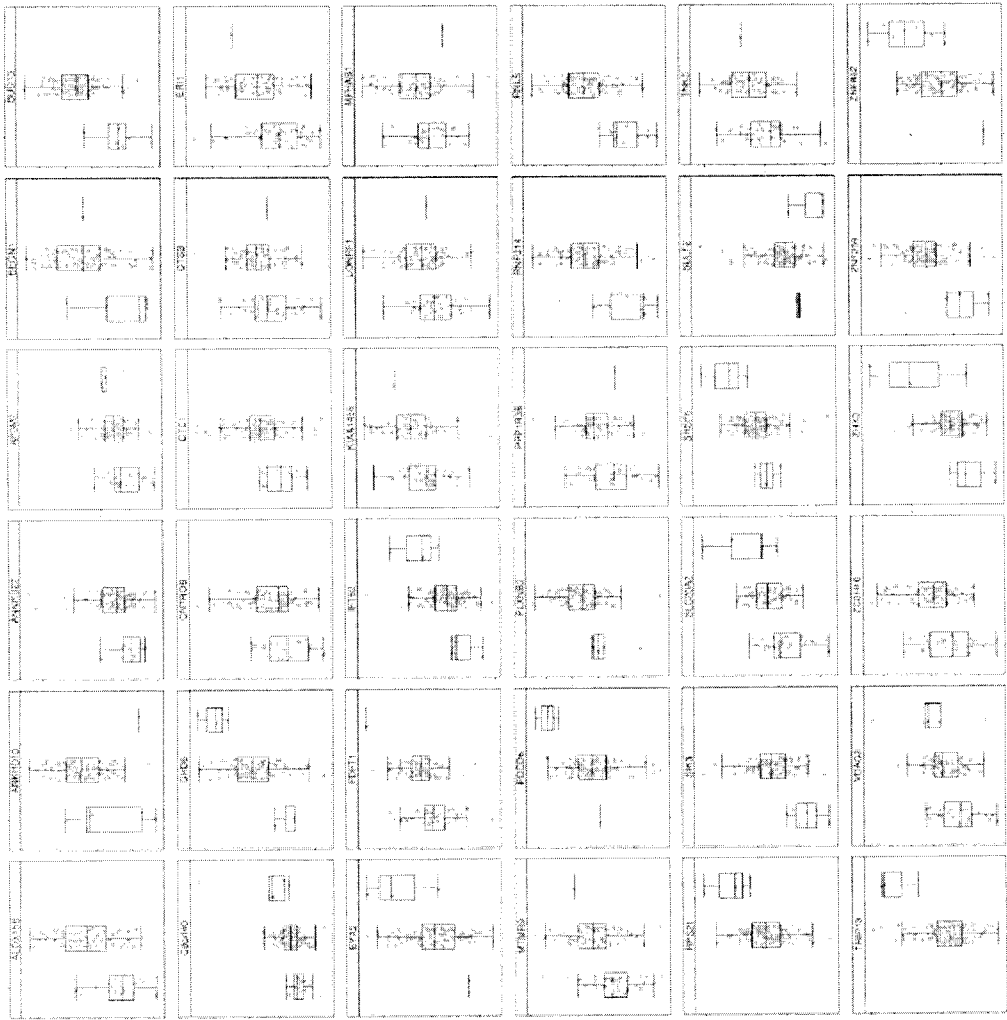


Figure 12

13/22

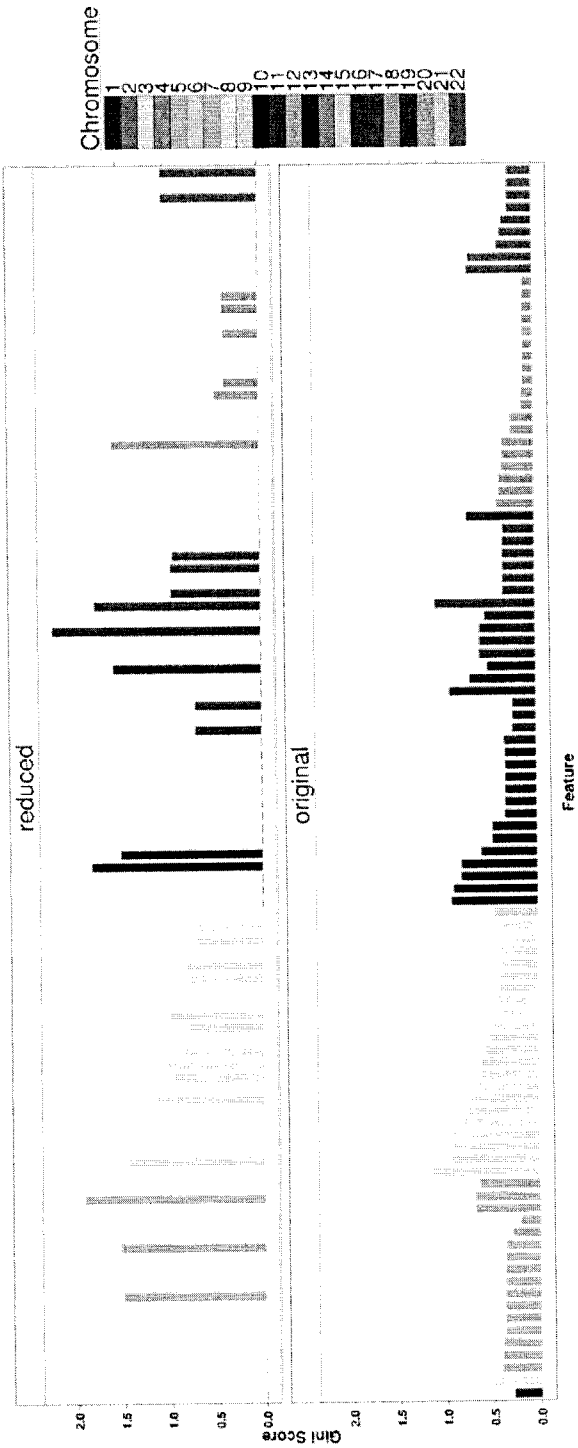


Figure 13

14/22

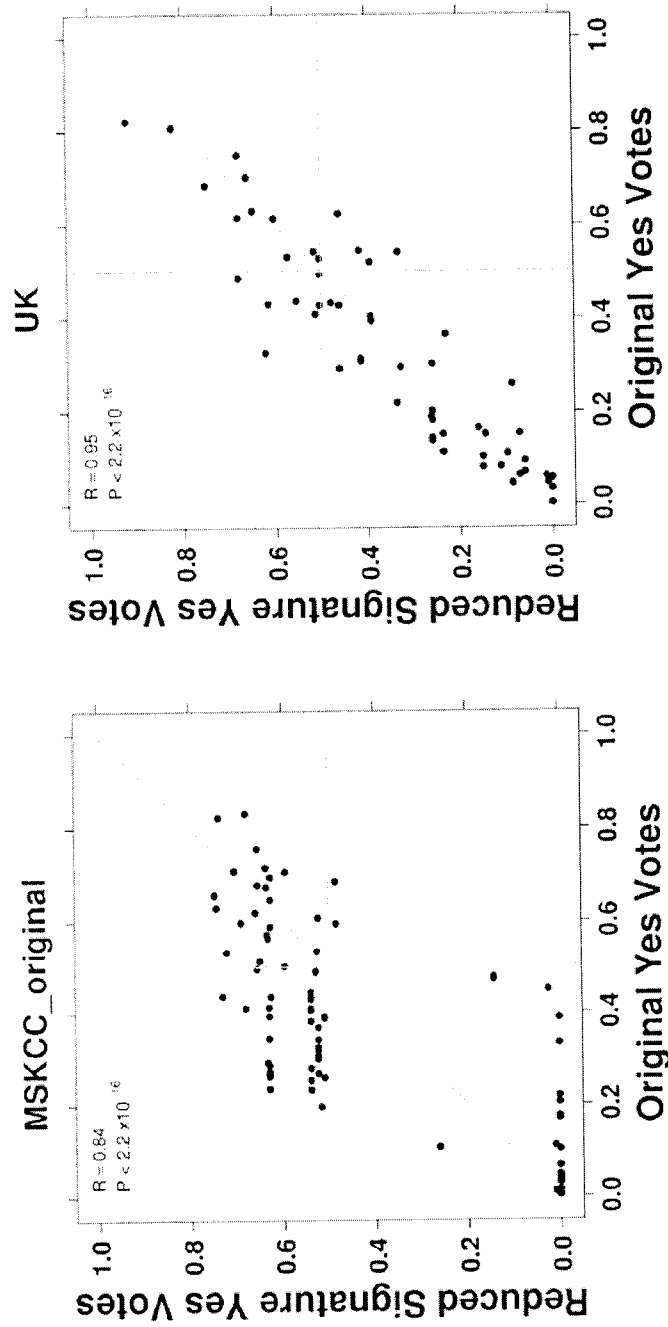


Figure 14

15/22

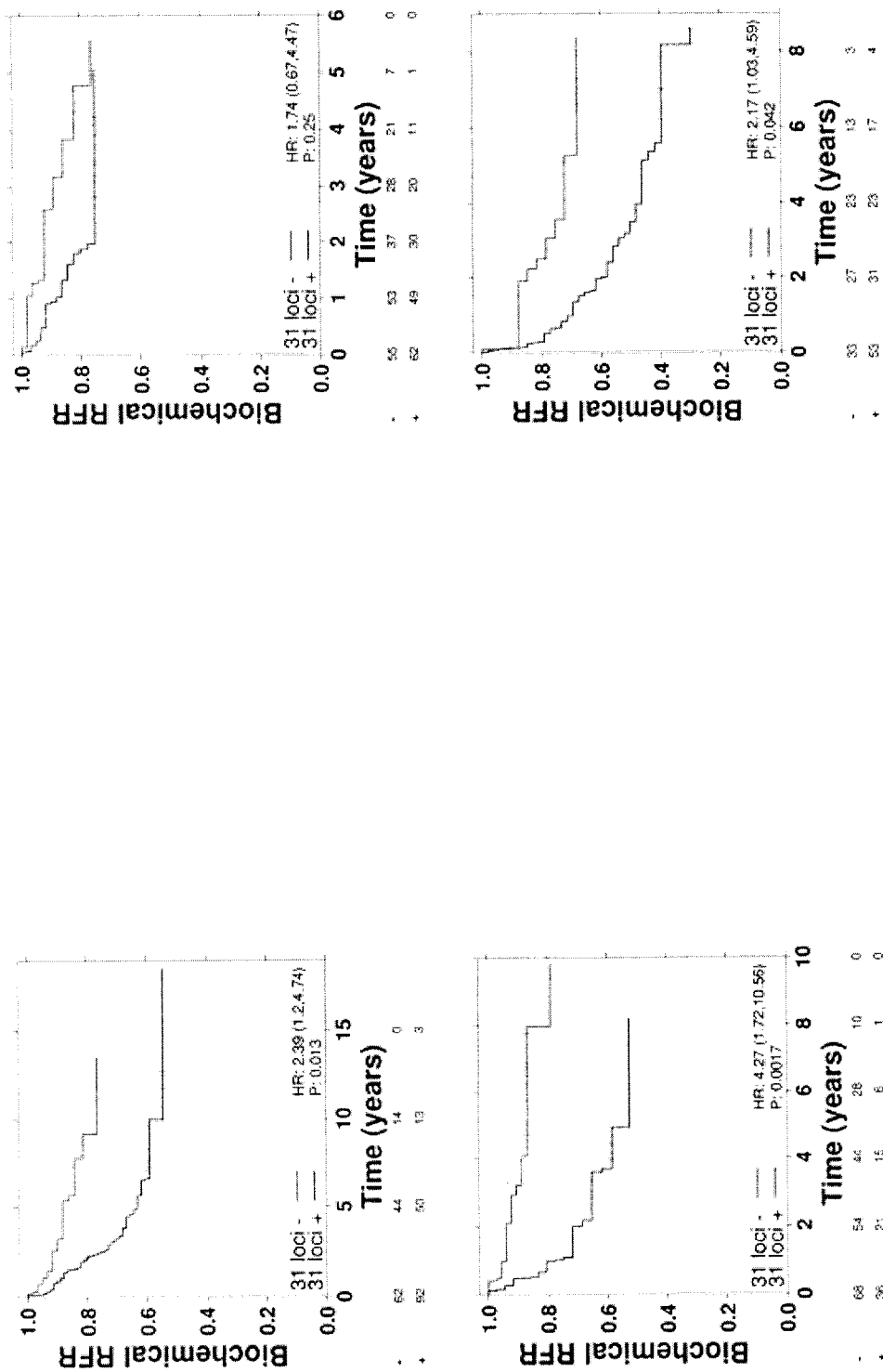


Figure 15

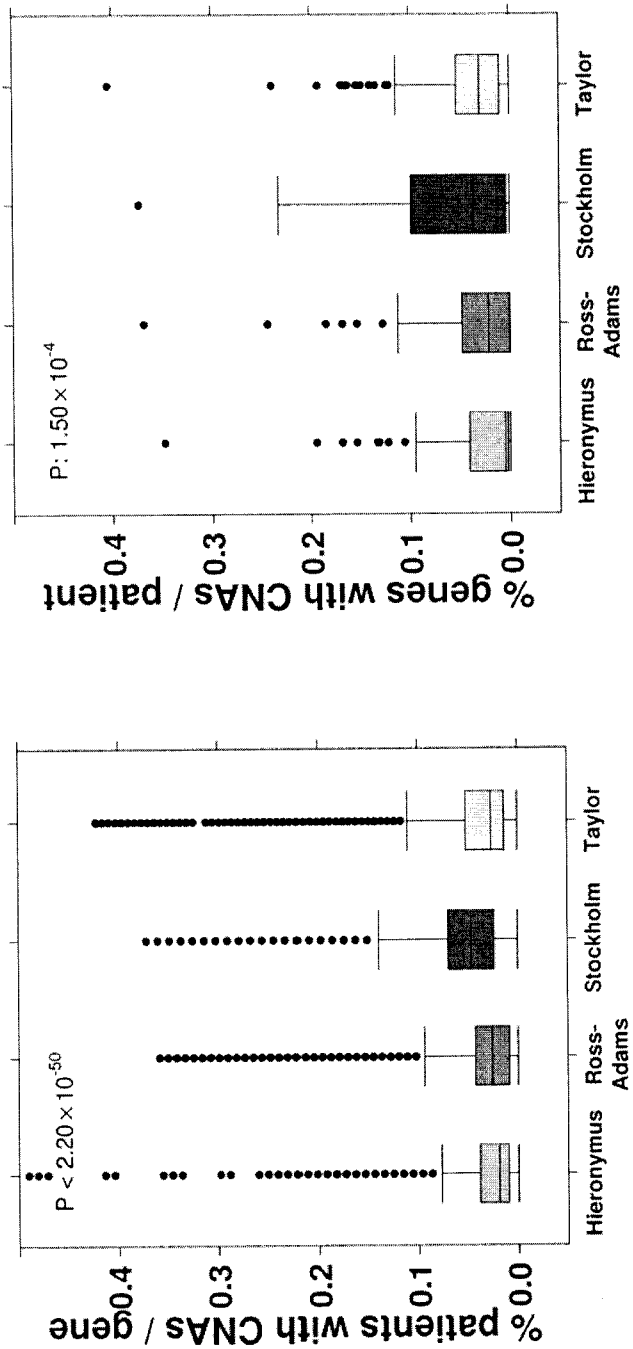


Figure 16

17/22

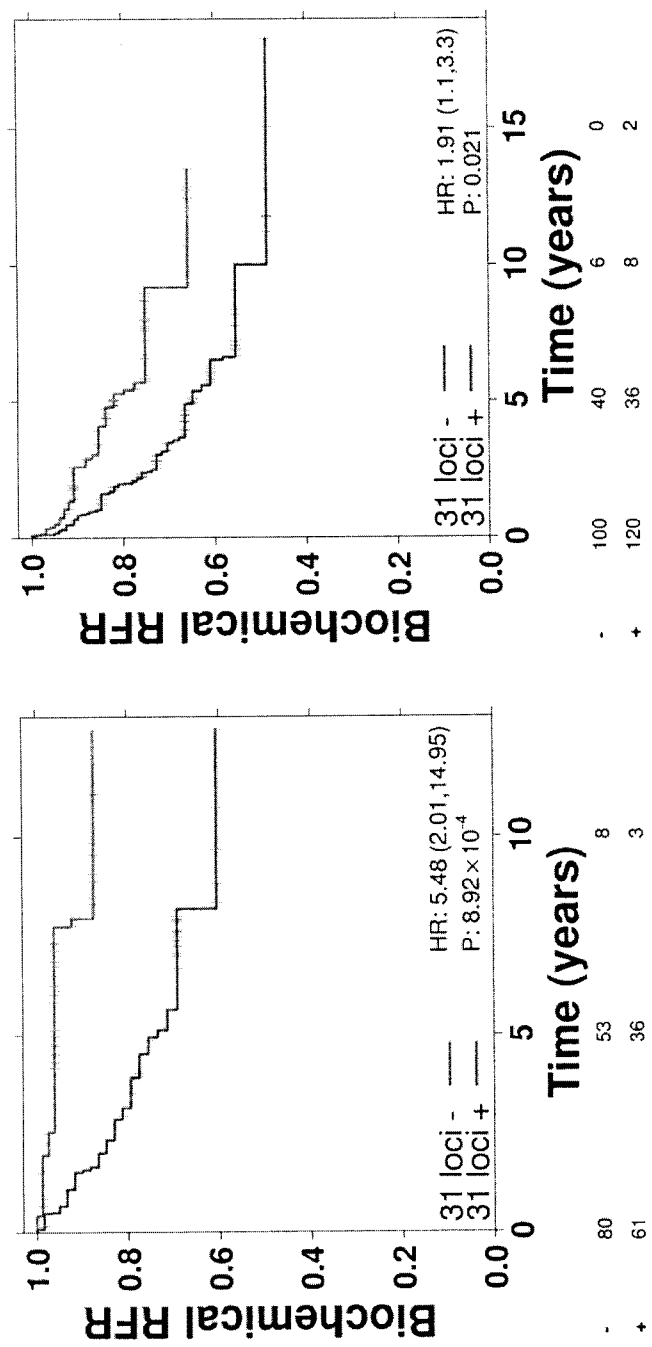


Figure 17

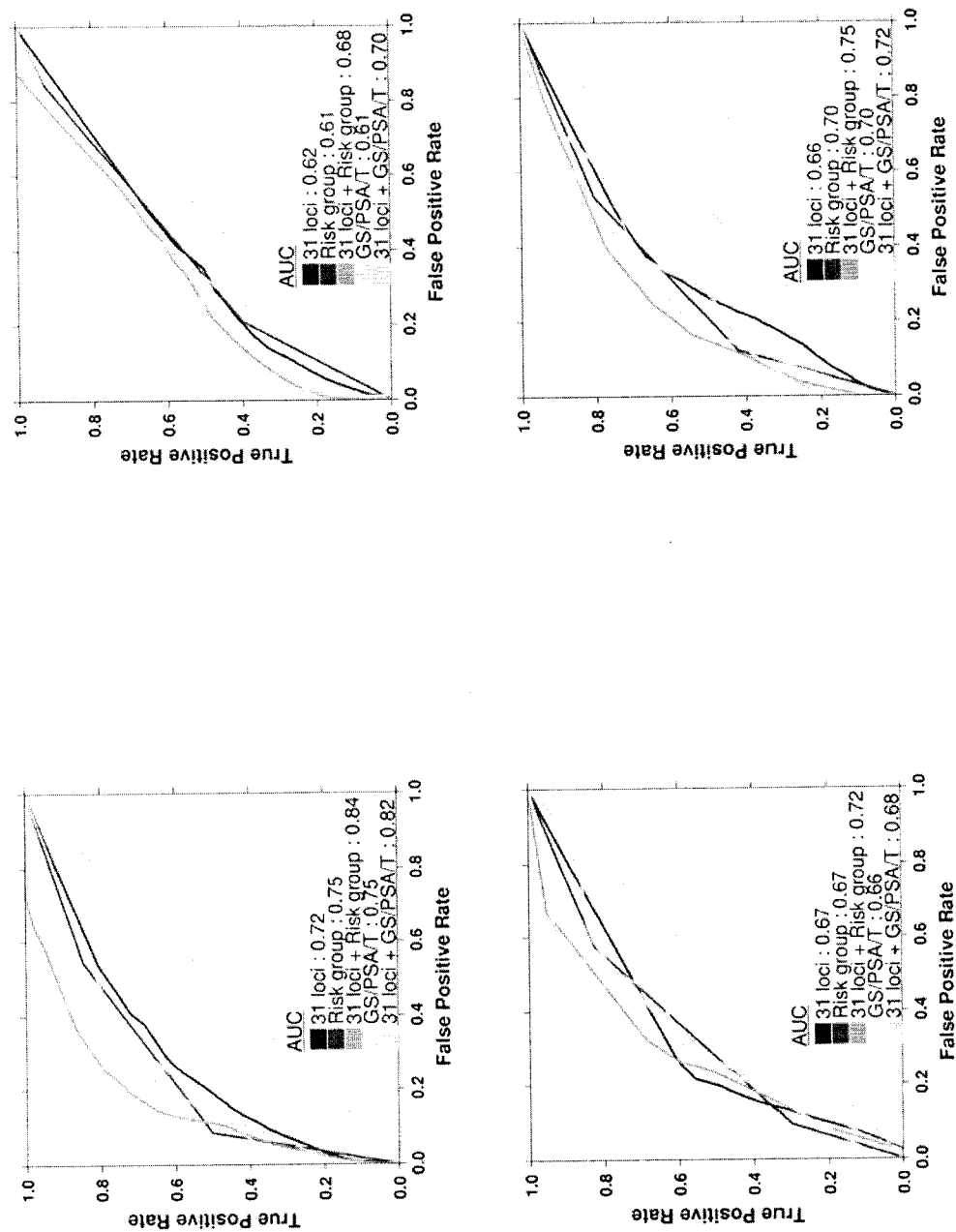


Figure 18

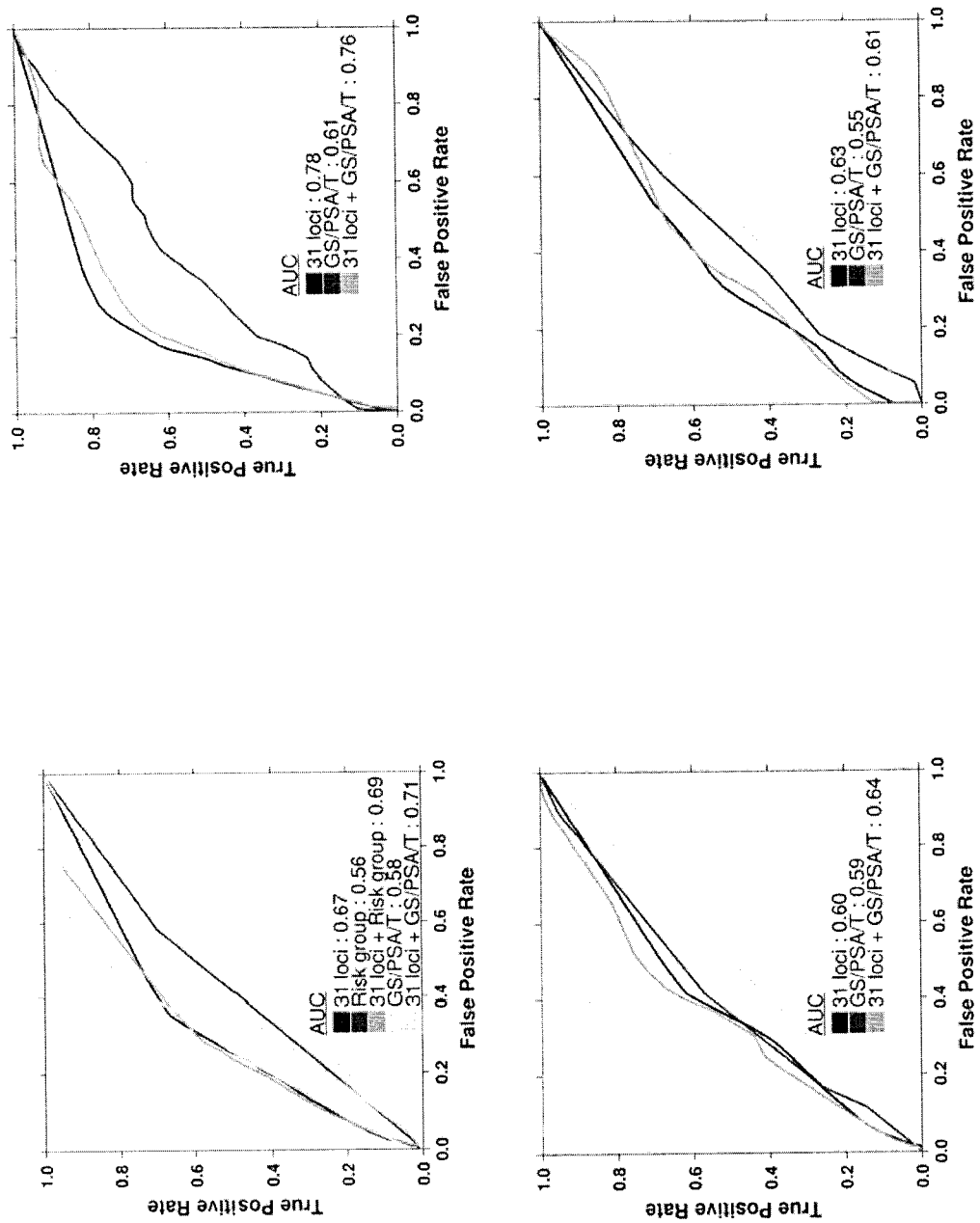


Figure 19

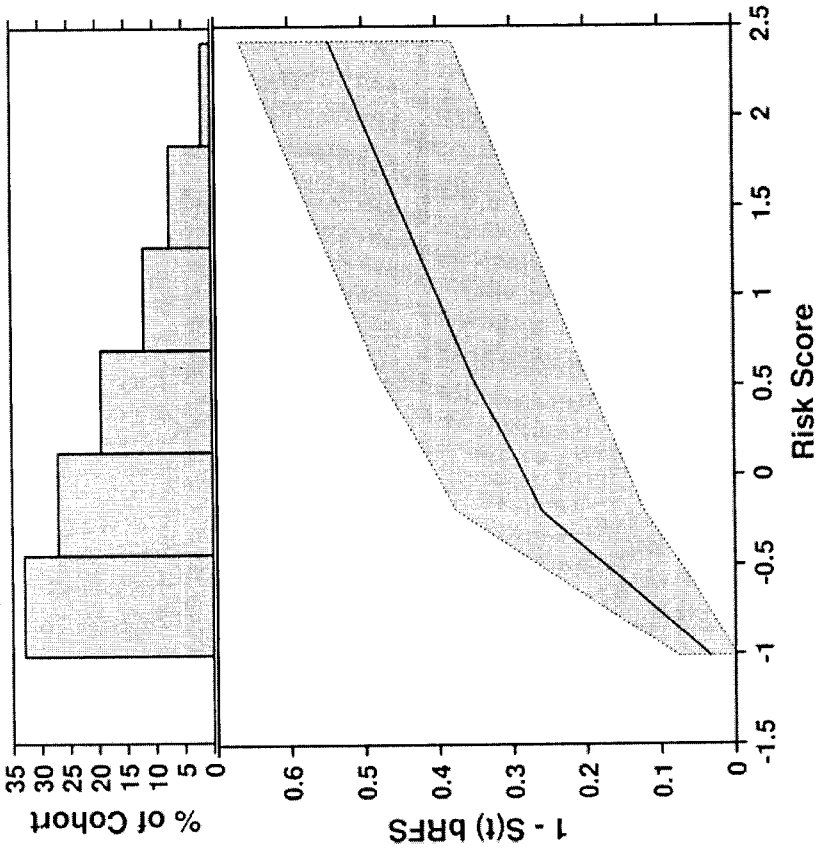


Figure 20

21/22

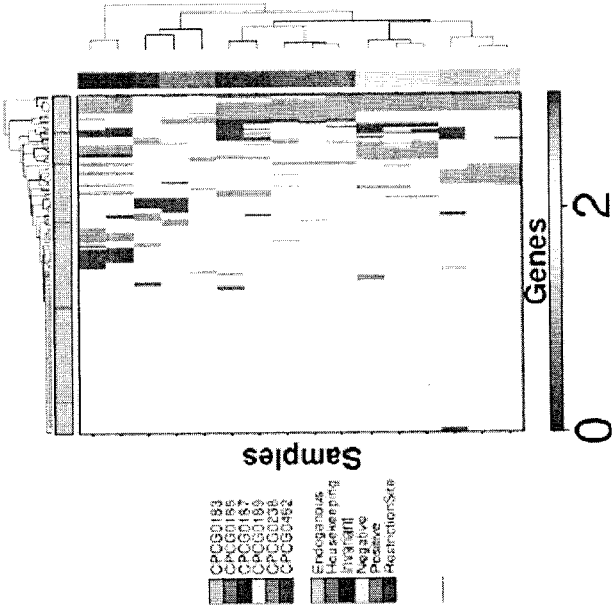
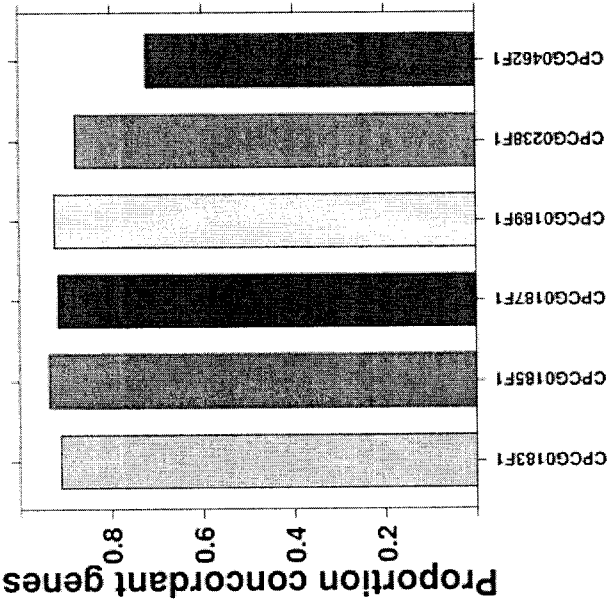


Figure 21

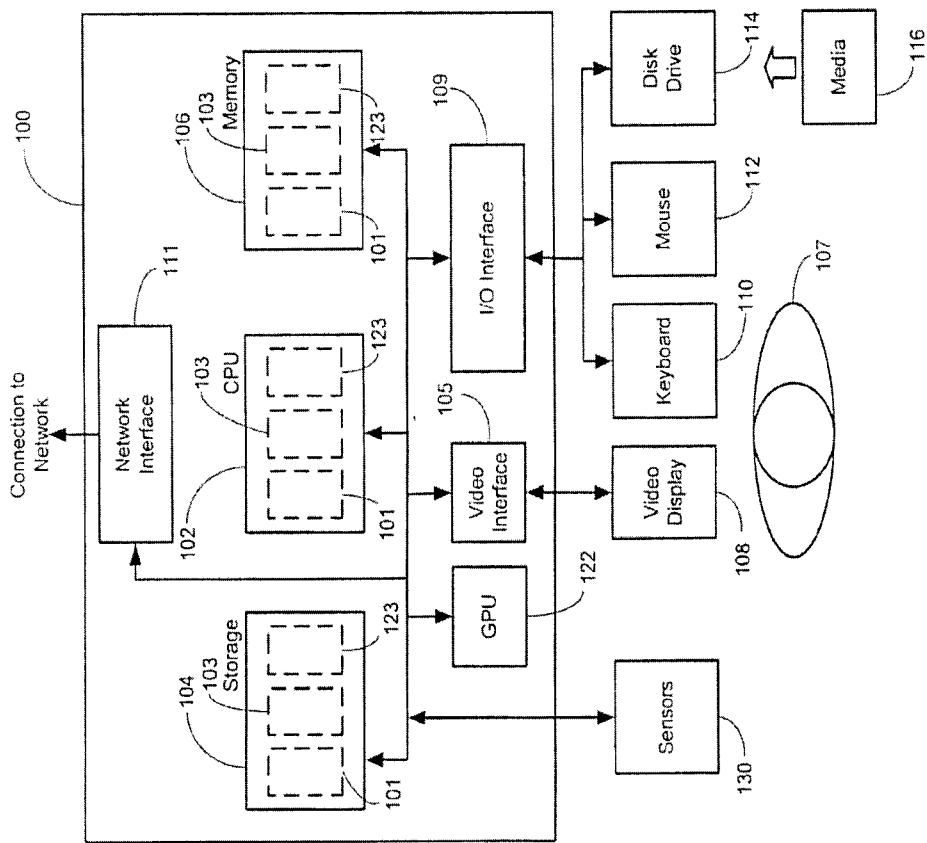


Figure 22

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CA2017/000095

A. CLASSIFICATION OF SUBJECT MATTER

IPC: **C12Q 1/68** (2006.01), **G01N 33/48** (2006.01), **G06F 17/18** (2006.01), **G06F 19/18** (2011.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: **C12Q 1/68** (2006.01), **G01N 33/48** (2006.01), **G06F 17/18** (2006.01), **G06F 19/18** (2011.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used)

Database(s): STN (Medline, CaPlus, Biosis)

Keywords: prostat?, recur?, relaps?, risk?, genom?, hybrid?, RNA, copy number, Bristow, Lalonde, Boutros

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO2015106341A1 (BOUTROS et al.) 23 July 2015 (23-07-2015) see claims 1, 6, 22 and 23; Table 1; Figure 4; Example 5; paragraph 88	1-11
Y	LALONDE, E. et al., Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. Lancet Oncol., December 2014, 15(13): 1521-1532, ISSN 1474-5488 see whole document, especially Summary, Table 2, Figure 5 and page 1530, right column, first full paragraph	1-11

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* "A" "E" "L" "O" "P"	Special categories of cited documents: document defining the general state of the art which is not considered to be of particular relevance earlier application or patent but published on or after the international filing date document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) document referring to an oral disclosure, use, exhibition or other means document published prior to the international filing date but later than the priority date claimed	"T" "X" "Y" "&"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art document member of the same patent family
--------------------------------------	--	--------------------------	--

Date of the actual completion of the international search
18 May 2017 (18-05-2017)Date of mailing of the international search report
15 June 2017 (15-06-2017)Name and mailing address of the ISA/CA
Canadian Intellectual Property Office
Place du Portage I, C114 - 1st Floor, Box PCT
50 Victoria Street
Gatineau, Quebec K1A 0C9
Facsimile No.: 819-953-2476

Authorized officer

Ralph Salvino (819) 639-7650

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CA2017/000095

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ROSS-ADAMS et al., Integration of copy number and transcriptomics provides stratification in prostate cancer: A discovery and validation cohort study. EBioMedicine, 29 July 2015, 2(9):1133-1144 Epub [retrieved on 2017-05-19] Retrieved from the Internet: <URL>:https://dx.doi.org/10.2016/j.ebiomed.2015.07.017 See Abstract, Figures 2-4; Section 3.7; Discussion	1-11
P,X	FRASER, M. et al., Genomic hallmarks of localized, non-indolent prostate cancer. Nature, 19 January 2017, 541(7637):359-364, ISSN 1476-4687 see whole document	1-11
P,X	LALONDE, E., Translating a prognostic DNA genomic classifier into the clinic: Retrospective validation in 563 localized prostate tumors. European Urology, 1 November 2016, S0302-2838(16)30716-3 Epub [retrieved on 2017-05-19] Retrieved from the Internet: <URL>:https://dx.doi.org/10.2016/j.eururo.2016.10.013 See whole document	1-11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CA2017/000095**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of the first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claim Nos.: 1-3, 10 and 11(partially)
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

The application provides disclosure within the meaning of Article 5 for only one type of cancer, prostate cancer. As such, a meaningful search over the whole scope claimed is not warranted. Consequently, the search has been established for the parts of the application which are sufficiently enabled, namely the claimed methods and system wherein the cancer is prostate cancer.

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claim Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CA2017/000095

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
WO2015106341A1	23 July 2015 (23-07-2015)	WO2015106341A1 CA2937051A1 EP3095055A1 US2016333421A1	23 July 2015 (23-07-2015) 23 July 2015 (23-07-2015) 23 November 2016 (23-11-2016) 17 November 2016 (17-11-2016)