BJUI
BJU International

# Clinical and genomic analysis of metastatic prostate cancer progression with a background of postoperative biochemical recurrence

Mohammed Alshalalfa, Anamaria Crisan, Ismael A. Vergara, Mercedeh Ghadessi, Christine Buerki, Nicholas Erho, Kasra Yousefi, Thomas Sierocinski, Zaid Haddad, Peter C. Black*, R. Jeffrey Karnes†, Robert B. Jenkins‡ and Elai Davicioni

*GenomeDx Biosciences, Inc. , *Department of Urologic Sciences, University of British Columbia, Vancouver, BC, Canada, †Department of Urology, and ‡Department of Pathology and Laboratory Medicine,Mayo Clinic,Rochester,MN, USA*

M.A. and A.C. contributed equally to this work.

## Objective

To better characterize the genomics of patients with biochemical recurrence (BCR) who have metastatic disease progression in order to improve treatment decisions for prostate cancer.

## Methods

The expression profiles of three clinical outcome groups after radical prostatectomy (RP) were compared: those with no evidence of disease (NED; $n$ = 108); those with BCR (rise in prostate-specific antigen [PSA] level without metastasis; $n$ = 163); and those with metastasis ($n$ = 192). The patients were profiled using Human Exon 1.0 ST microarrays, and outcomes were supported by a median 18 years of follow-up. A metastasis signature was defined and verified in an independent RP cohort to ensure the robustness of the signature. Furthermore, bioinformatics characterization of the signature was conducted to decipher its biology.

## Results

Minimal gene expression differences were observed between adjuvant treatment-naïve patients in the NED group and patients without metastasis in the BCR group. More than 95% of the differentially expressed genes (metastasis signature) were found in comparisons between primary tumours of metastasis patients and the two other outcome groups. The metastasis signature was validated in an independent cohort and was significantly associated with cell cycle genes, ubiquitin-mediated proteolysis, DNA repair, androgen, G-protein coupled and NOTCH signal transduction pathways.

## Conclusion

This study shows that metastasis development after BCR is associated with a distinct transcriptional programme that can be detected in the primary tumour. Patients with NED and BCR have highly similar transcriptional profiles, suggesting that measurement of PSA on its own is a poor surrogate for lethal disease. Use of genomic testing in patients undergoing RP with an initial rise in PSA level may be useful to improve secondary therapy decision-making.

## Keywords

prostate cancer, gene expression, metastasis, biochemical recurrence, gene functional analysis

## Introduction

The clinical course of patients who experience biochemical recurrence (BCR) after radical prostatectomy (RP) is highly variable. The detection of BCR on its own does not distinguish patients with indolent micrometastatic or locally recurrent disease from those with systemic recurrence of aggressive prostate cancer [1]; therefore the development of a rising PSA level after RP poses an important clinical challenge because indolent disease is clinically managed differently from aggressive disease [2]. For many patients with prostate cancer, a rise in PSA level leads to treatment, which has potential side effects including voiding and erectile dysfunction. It is critical, therefore, to determine which patients with rising PSA levels should receive aggressive treatment. Understanding the underlying genomic differences between men with a high metastatic potential tumour and those with more indolent disease may enable the development


BJU International © 2014 BJU International | doi:10.1111/bju.13013
Published by John Wiley & Sons Ltd. www.bjui.org

of individualized approaches to treat the cancer and not the PSA [3].

As prostate cancer has a long natural history, the assumption that BCR can be used as a surrogate to infer a patient's ultimate clinical course may in fact mask the true biological and clinical risk categories required to make appropriate treatment decisions [4]. Furthermore, even patients defined as 'at-risk' based on adverse pathological features after RP are observed to have highly variable clinical outcomes [4,5]. While current clinical guidelines recommend adjuvant radiation therapy for pT3 or margin-positive disease, many clinicians and patients prefer to delay treatment until PSA level rises [8]. Improved risk prediction models that take into account tumour biology may provide better risk stratification and utilization of adjuvant and salvage therapies [6].

We have previously reported on the development and validation of several biological signatures for clinical metastasis using a large cohort that defined three distinct clinical outcome groups [5,7]. In the initial study by Nakagawa et al. [5] we first observed that there were no statistically significant differences in gene expression between the control group with no evidence of disease (NED) and the control group with BCR (a rise in PSA level without metastasis within 5 years after RP); however, the array platform used in that initial study had focused only on ~1 000 expressed prostate cancer and oncology-related genes, primarily gleaned from previous gene expression studies. In the present study, we follow up our most recent report [7] with a high-resolution genome-wide analysis of the molecular characteristics that underlie NED, BCR and metastatic prostate cancer.

## Materials and Methods

### Patient Population and Definition of Clinical Outcomes

The patients included in the present study underwent RP as first-line treatment for prostatic adenocarcinoma at the Mayo Clinic Comprehensive Cancer Centre between 1987 and 2001. A nested case–control study was designed from this cohort of men and was previously described in Nakagawa et al. [5]; patients were excluded if they had received any neoadjuvant therapy or failed to achieve PSA nadir (PSA <0.1 ng/mL) after RP. The original study design comprised 213 men with clinical metastasis (metastasis group; within 5 years of BCR and confirmed by positive bone or CT scans), matched to one of each of the following control subjects: PSA control or BCR group (BCR with no subsequent metastasis within 5 years); and a control group with NED (no BCR with at least 7 years of follow-up). BCR was defined as a PSA measure >0.2 ng/mL and further confirmed by a subsequent measure that was ≥0.05 ng/mL above the first. The matching criterion is detailed in Nakagawa et al. [5] and includes age, time to BCR (or year of RP) and pathological Gleason score. An overview of the study design and dataset assembly for analysis is shown in Fig. 1.

For the present investigation, microarray expression profiles were available for a total of 545 patients (85%) from the original set. Two sets of analyses were performed. The first included analysis of all 545 patients (full set analysis). In the second analysis (pure set analysis), to control for adjuvant treatment effects in the NED group, patients treated with adjuvant radiation or hormonal therapies were excluded ($n$ = 61) to create a treatment-naive control group. In addition, infrequent late recurrence or metastasis events that occurred in the NED or BCR groups >10 years after RP were excluded ($n$ = 21) to focus the analysis on the biological differences between early metastatic disease progression and disease with PSA rise only or no evidence of PSA rise after RP. For validation, a cohort of 219 patients described previously by Karnes et al. [8] with the same exclusion criteria for adjuvant treatment in the NED group applied, leaving 179 patients for analysis; 69 were in the metastasis group.

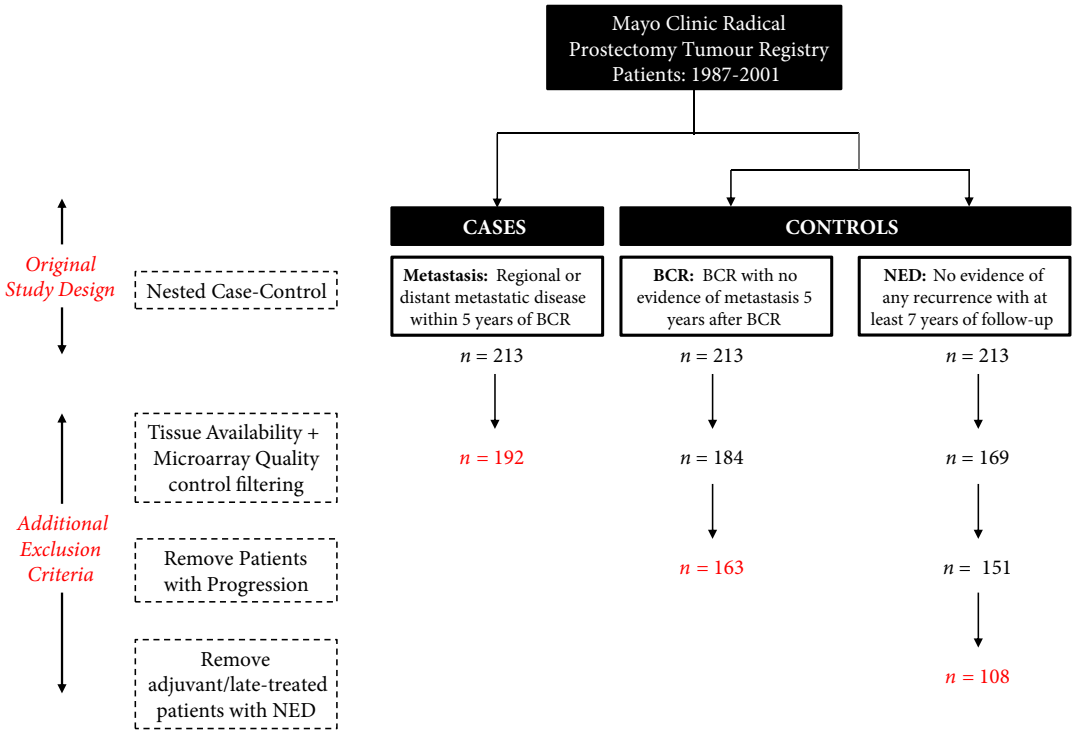
### Sample Preparation and Data Generation

Tumour specimens were obtained from archived paraffin blocks and RNA extraction was performed as described previously [5]. Generation of expression data for ~1.4 million probe selection regions evaluated in the present study, hereafter referred to as features, as well as further details on microarray processing and quality control have been previously described [7]. An overview of the study design and sample preparation steps is shown in Fig. 1.

### Expression Analysis of the Transcriptome

Differentially expressed features were obtained by pairwise comparisons among the metastasis, BCR and NED outcome groups. For each pairwise comparison, features were filtered at two thresholds of median fold difference: 1.5 and 2.0. A Wilcoxon test was subsequently applied and the resulting $P$ values were corrected for multiple testing using the Benjamini–Hochberg method and further filtered using $\alpha \leq 0.01$. Figure S1 shows the number of features filtered at each step in the pure set analysis. Differentially expressed features between pairwise analyses were compared and two lists were generated. The first was the union between metastasis/NED and metastasis/BCR (1 481 features) and the second was the overlapping list of features between metastasis/NED and metastasis/BCR (334 features) comparisons. The lists were assessed for their independent prognostic potential using a multivariable generalized linear mixed effect model with a logit link. This model used a

**Fig. 1** An overview of the case–control study design and patient selection criteria. BCR, biochemical recurrence; NED, no evidence of disease.




random-effects variable to account for matched pairs or triples resulting from the original matched nested case–control design [5] and included clinicopathological covariates as fixed effects. Features were considered to be independently prognostic of the clinical covariates if the $P$ value from the coefficient was ≤0.05 (two-sided test). All statistical analyses were performed with R v3.0 software (R Foundation for Statistical Computing, Vienna, Austria).

## Functional Annotation of the Transcriptome

The annotation of features to genes was based on the alignment provided by xmapcore against the hg19 version of the human genome. Based on the proximity of a feature relative to the genes annotated with Ensembl v62, features were categorized as either coding, untranslated region, intronic, antisense and intergenic. Features overlapping with known non-coding transcripts were annotated as non-coding transcripts. More than 75% of the features on the microarray cover regions were annotated as non-protein coding [7]. Functional annotation of biological processes and pathways was performed using the Reactome FI Cytoscape Plugin (version 4) using default settings [9]. Functional clustering of genes into relevant pathway modules was performed and functional annotation data were extracted per module with a minimum module size ≥3 and with an enrichment $P$ value of ≤0.05.

## Clinical Covariates Used in Analysis

Clinicopathological covariates were analysed in conjunction with genomic features and included: surgical margin status; seminal vesicle invasion; lymph node involvement; extracapsular extension; pathological Gleason Score; PSA doubling time after RP; and the preoperative PSA measure immediately before RP (ng/mL). Information on adjuvant and salvage therapies was also included.

For the purposes of the present investigation, clinicopathological covariates and treatment data were calculated, defined or transformed in the manner described below. PSA doubling time was calculated by dividing the natural log of 2 by the slope of the linear regression of the log PSA measures over time [1]. Patients in the NED group were given a PSA doubling time value of 'not applicable' and the rest were dichotomized into groups with the threshold of ≥9 months [10]. For multivariable analyses, pathological Gleason score was dichotomized by Gleason score ≥ 8; although it is conventional to define three Gleason score groups (≤6, 7, ≥8), the dearth of patients with Gleason score ≤6 prompted analysis of two groups. Primary and secondary Gleason pattern were not available for this cohort. Adjuvant therapy was defined as any treatment (hormone or radiation) administered within 6 months of RP with undetectable PSA, whereas salvage therapy was any treatment undergone before confirmed clinical metastasis. Therapies administered after

confirmed clinical metastasis were not included in the analysis.

## Results

### Study Population and Clinical Course

A total of 545 (85% of the original cohort described in Nakagawa et al. [5]) patient expression profiles from the three outcome groups were available for the full set analysis (NED group, $n$ = 169; BCR group, $n$ = 184; metastasis group, $n$ = 192). The clinical characteristics of these patients have been previously described by Erho et al. [7]. From this, refined outcome group criteria were applied to perform the 'pure set' analysis (Fig. 1). In the pure set analysis, to control for adjuvant treatment effects in the NED group, patients treated with adjuvant therapy were excluded to create a treatment-naive control group. In addition, infrequent late recurrence or metastasis events that occurred in the NED or BCR groups (>10 years after RP) were also excluded, resulting in 463 patients (NED group, $n$ = 108; BCR group, $n$ = 163; metastasis group, $n$ = 192) who met these refined outcome group criteria. The clinicopathological characteristics and clinical courses of these patients are shown in Table 1. The patients in the BCR and metastasis groups were originally matched based on their time to BCR [5] and, accordingly, no significant differences in time to BCR were found ($P$ = 0.56). In the metastasis group, the median time of metastasis diagnosis by radiographic imaging was 5.2 years and median time to prostate-cancer specific mortality was 10.5 years after surgery. Half of the metastasis group had died from any cause by 9.2 years after RP, while the patients in the NED group did not reach median survival on follow-up and the BCR group took >20 years to reach median survival.

In multivariable analysis comparing the three groups, the BCR group was 2.6 times more likely to have a positive surgical margin than the NED group and 1.42 times more likely to have a high preoperative PSA level. Patients in the metastasis group, however, had a lower preoperative PSA (odds ratio 0.76), higher Gleason scores (odds ratio 2.99) and were more likely to have lymph node invasion (odds ratio 3.7; Table S1).

**Table 1** Overview of the patient characteristics of the no evidence of disease, biochemical recurrence and metastasis clinical outcome groups.

| | Patient group | | |
|---|---|---|---|
| | NED | BCR only | Metastasis |
| No. of patients | 108 | 163 | 192 |
| Median follow-up time, years | 16.48 | 17.04 | 17.6 |
| Pathological Gleason sum, *n* (%) | | | |
| ≤6 | 26 (24) | 27 (16.5) | 6 (3) |
| 7 | 69 (64) | 92 (56.5) | 70 (36.5) |
| ≥8 | 13 (12) | 44 (27) | 116 (60) |
| Pathological stage, *n* (%) | | | |
| T2aN0 | 29 (27) | 23 (14) | 20 (10.5) |
| T2bN0 | 45 (41.5) | 47 (29) | 32 (16.5) |
| T3aN0 | 25 (23) | 47 (29) | 42 (22) |
| T3b4N0 | 8 (7.5) | 39 (24) | 53 (27.5) |
| TxN+ | 1 (1) | 7 (4) | 45 (23.5) |
| Adverse pathological features, *n* (%) | | | |
| Extracapsular extension | 29 (27) | 77 (47) | 118 (61.5) |
| Seminal vesicle invasion | 9 (8.3) | 43 (26) | 87 (45) |
| Lymph node involvement | 1 (1) | 7 (4) | 45 (23.5) |
| Surgical margin status | 24 (22) | 82 (50) | 103 (53.5) |
| Treatment, *n* (%) | | | |
| Adjuvant hormone therapy | 0 | 20 (12) | 62 (32) |
| Adjuvant radiation therapy | 0 | 4 (2.5) | 28 (14.5) |
| Salvage hormone therapy | 0 | 72 (44) | 90 (47) |
| Salvage radiation therapy | 0 | 41 (25) | 33 (17) |
| Median (IQR) time to BCR, years | NA | 2.53 (1.10–4.59) | 2.92 (1.23–5.03) |
| Median (IQR) time to first rise in PSA level, years | NA | 1.99 (0.75–3.59) | 2.26 (0.76–3.94) |
| Preoperative PSA, *n* (%) | | | |
| <6 | 78 (72) | 75 (46) | 93 (48.5) |
| ≥10 to <20 | 21 (19.5) | 45 (27.5) | 46 (24) |
| ≥20 | 9 (8.5) | 43 (26) | 53 (27.5) |
| PSA doubling time, *n* (%) | | | |
| ≤9 months | NA | 96 (59) | 150 (78) |
| >9 months | NA | 67 (41) | 42 (22) |
| Prostate cancer death | | | |
| Yes | 0 | 0 | 118 (61.5) |

*NA, not applicable; IQR, interquartile range; BCR, biochemical recurrence; NED, no evidence of disease.*

## Distinct Transcriptional Programme in the Metastasis Group

After quality control filtering for detection above background and cross hybridization, a total of 674 521 features (~48% of the array) were used for the analysis [10]. Pairwise comparisons of the outcome groups resulted in 1 069, 761 and 18 differentially expressed features for metastasis/NED, metastasis/BCR, and BCR/NED group comparisons, respectively, at a median fold-change threshold of 1.5 after multiple testing correction (Figure S1). A total of 1 481 differentially expressed features for the metastasis/NED and metastasis/BCR comparisons were observed, with 334 overlapping between the comparisons. The metastasis group generally had higher expression of the 1 481 differentially expressed features compared with the NED and BCR groups (Fig. 2A). Furthermore, 15/18 differentially expressed features between the BCR and NED groups were also differentially expressed between metastasis and NED groups (Fig. 2B), but none of these met the threshold for differential expression between the metastasis and BCR groups (Fig. 2C).

These results show that patients in the metastasis group had a distinct transcriptional profile compared with those in the NED and BCR groups; however, as the metastasis patient group had a high number of patients with high Gleason score and nodal disease, we next tested whether the metastasis signature could be explained primarily by the skewness of the distribution of these pathological variables among the three outcome groups. Using the same filtering parameters used in the identification of the metastasis signature, we found 81 differentially expressed features between high (Gleason score $\geq 8$) and low (Gleason score $\leq 7$) Gleason score, of these, 46 overlapped with the metastasis signature. Similarly, we found 117 differentially expressed features between tumours with lymph node-positive and lymph node-negative disease, of which, 87 overlapped with the metastasis signature. This analysis shows therefore that only a small fraction of the metastasis signature can be accounted for by differences in distribution of Gleason score or lymph node invasion among the outcome groups.

There were also 720 and 427 features that did not overlap between the metastasis/NED and metastasis/BCR analyses (Fig. 2D), respectively; for these features the BCR and NED groups tended to have similar expression to each other, relative to the metastasis group. A concordance analysis was conducted; features were considered concordant if the fold-change ratios for metastasis/NED and metastasis/BCR were both >1 or both <1, and discordant if not. Of the 720 features, none were discordant and of the 427 only one was discordant. When applying the more stringent threshold of ≥2-fold change, 49, 20 and 0 features were found to be differentially expressed between metastasis/NED, metastasis/BCR and BCR/NED comparisons (Fig. 2E), respectively.

There were 16 overlapping features between the metastasis/BCR and metastasis/NED comparisons (Table S1). Overall, the extent of concordance showed that the expression levels in the NED and BCR groups were very similar to each other, whereas the metastasis group was the most distinct. The analysis was repeated on the whole cohort of 545 patients to exclude any selection bias introduced by the additional exclusion criteria (i.e. the 61 patients in the NED group who received adjuvant treatment or the 21 patients in the BCR group who developed late metastasis). Compared with the pure set analysis, the results showed that there was >80% overlap of metastasis/BCR and metastasis/NED genes (Figure S2).

## Evaluating the Robustness of the Metastasis Signature

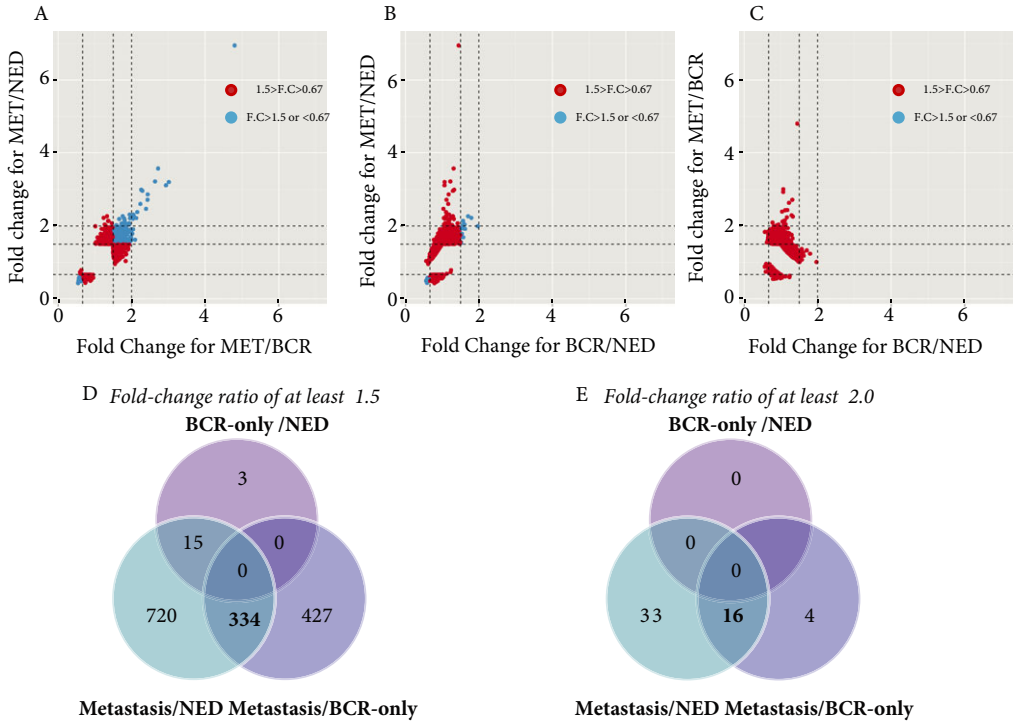
Next, the putative metastasis signature comprising all 1 481 features differentially expressed in the pairwise analyses was evaluated in an independent cohort of patients at high risk who had undergone RP. The cohort in the study by Karnes et al. [8], excluding the patients with NED who had received adjuvant hormonal or radiation treatment ($n = 171$) was evaluated using agglomerative hierarchical clustering with Pearson's correlation as a distance metric and Ward linkage. The resulting dendogram of patients, which is a tree diagram to illustrate similarity between patients based on the expression of the 1 481 features, has three main branches. Patients within each branch have similar gene expression profiles. Cluster 1 had 57 patients, 70% of these men developed metastasis on study follow-up (median 6.9 years; Fig. 3A). In Cluster 2 ($n = 47$) and Cluster 3 ($n = 67$) 12 and 34% of patients had metastasis, respectively.

Kaplan–Meier survival analysis further shows that patients in Cluster 1 developed metastasis rapidly after RP, with a median metastasis-free survival time of <4 years (log-rank, $P < 0.0001$) and half of the patients dying from prostate cancer by 10 years after RP (log-rank $P < 0.0001$) (Fig. 3B). Hazards analysis shows that Cluster 1 had a hazard ratio for metastasis of 7.1 (95% CI 3.2–15.9) compared with Cluster 2 as reference. Cluster 3 had slightly higher but significant hazard ratio of 1.5 compared with Cluster 2 (95% CI 1–2.3). These results suggest that those patients with the highest expression levels of the metastasis signature in their primary tumour are most at risk for rapid metastatic progression after RP.

## Role of Non-Coding RNA in Metastatic Disease

Next, we investigated the representation of protein coding and non-coding regions of the genome in the differentially expressed genes detected in this analysis. At a fold-change threshold of 1.5, 68% of the 334 features were associated with exons or untranslated regions of protein coding genes. A total

**Fig. 2** Scatter plots summarizing the overlap from pairwise analysis of transcriptome differential expression. (**A**) scatter plot between metastasis/no evidence of disease (NED) and metastasis/biochemical recurrence (BCR) showing that (22%) of the 1 481 features have Fold Change (FC) > 1.5 in both experiments, and 16 features have FC > 2. (**B**) Only 15 features of the features differentially expressed in metastasis/NED have FC > 1.5 in BCR/NED analysis. (**C**) None of the features that have FC > 1.5 in metastasis/BCR have FC > 1.5 in the BCR/NED analysis. (**D**) Venn diagram showing the overlap between the comparisons at FC 1.5. (**E**) Venn diagram showing the overlap between the comparisons at FC 2.




of 12% of features were found among intronic regions and 8.0% in intergenic regions. Nine out of the 16 prognostic features, common between the 49 and 20 features in the metastasis/NED and metastasis/BCR comparative analysis, were derived from the 2q31.3 locus, with the majority of intergenic features derived from this locus (Table S2).

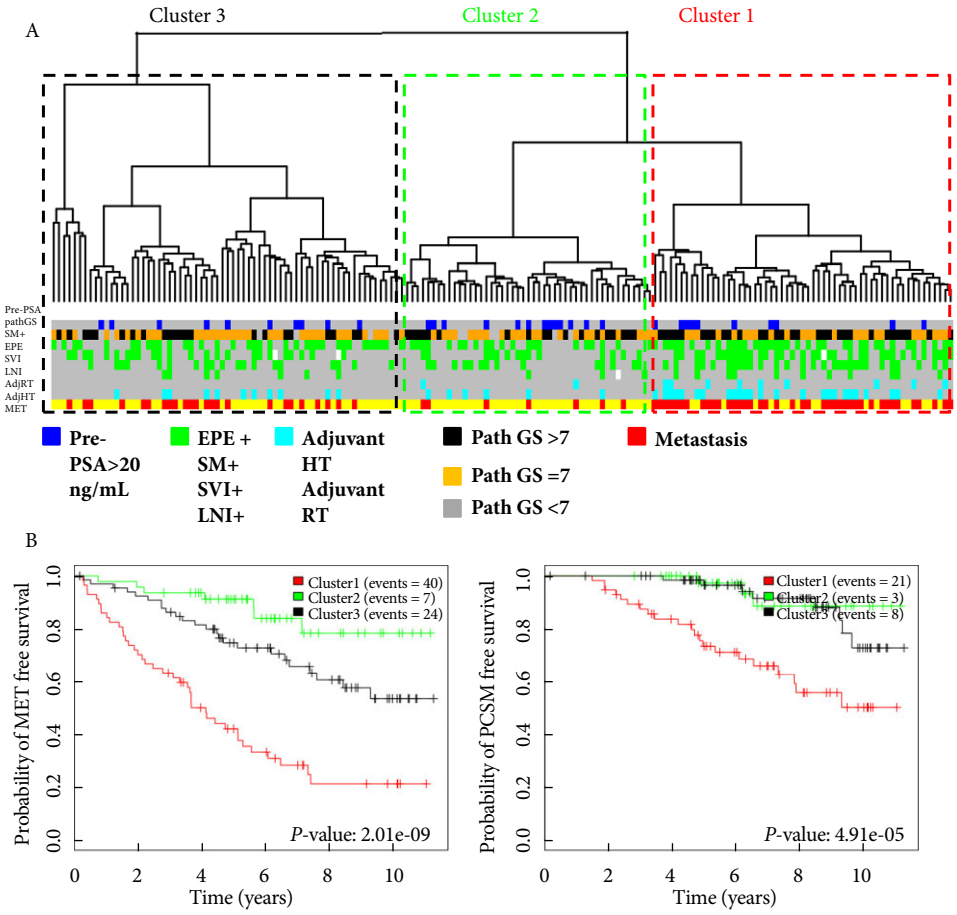
## Functional Characterization of Metastasis Signature

The 1 481 differentially expressed features of the putative metastasis signature were mapped and annotated to 965 unique genes. The gene list was first analysed by characterizing the functional interactions among the genes using Reactome FI tool [11] implemented in Cytoscape 3.1. Results show that 503 out of 965 genes are highly networked and clustered into 13 modules (Fig. 4). Enrichment analysis of GO biological processes and molecular functions terms was performed for each module (Table S3). Functional analysis of modules showed that modules representing mitotic cell cycle regulation, DNA repair, androgen receptor signalling, ubiquitin-mediated protein degradation, RNA splicing, MAPK kinase signalling, Notch, WNT and PI3K signal transduction pathways were prominent. Additional analysis showed that ATM, CDK1, MAP3K3 are key protein kinases targeting most of the metastasis signature, and UBC, RAD21, HDAC2, RELA and RAC1 appear to be hub genes of the metastasis signature. Further enrichment analysis of the metastasis signature in molecular concepts derived from experimental manipulations using the ConceptGen tool [12] showed that oestrogen stimulation and androgen-mediated signalling concepts were highly enriched in the metastasis signature (Table 2).

# Discussion

The use of BCR as a surrogate for lethal disease has emerged from the notion that nearly all patients with clinical metastasis will first have a PSA rise after RP as the only sign of disease recurrence [13]. Analysis of the natural history of prostate cancer after RP suggests that, although up to 50% of patients will experience BCR, at 15 years after RP, metastasis and prostate cancer-specific mortality rates approach 10 and 5%, respectively, even when no adjuvant or salvage therapy is administered before metastasis [1]. This calls into question the use of PSA as a surrogate endpoint in clinical trials, a prognostic biomarker and a treatment indicator. Recently Xia et al. [14], from the CISNET Prostate Cancer Modeling group, used simulation studies and modelling of patient outcomes data from the CAPSURE and Johns Hopkins registries to measure the potential for 'overdetection of

**Fig. 3** Annotated dendogram of the 1 481 features in Karnes et al. [13] validation cohort ($n$ = 171). (**A**) Using the 1 481 metastasis signature, hierarchical clustering using Pearson correlation as distance measure and ward as linkage function clusters patients from an independent high-risk RP cohort ($n$ = 171). Results show that most metastasis cases are clustered together suggesting that they are molecularly distinct from NED and BCR cases. (**B**) Kaplan–Meier survival analysis shows that Cluster 1 with the highest expression of the metastasis signature genes is enriched with patients who more rapidly develop metastasis and prostate-cancer specific mortality events. EPE, extracapsular extension; SM+, positive surgical margin status; SVI+, seminal vesicle invasion; LNI+, lymph node invasion; GS, Gleason score; RT, radiotherapy HT, hormone therapy.

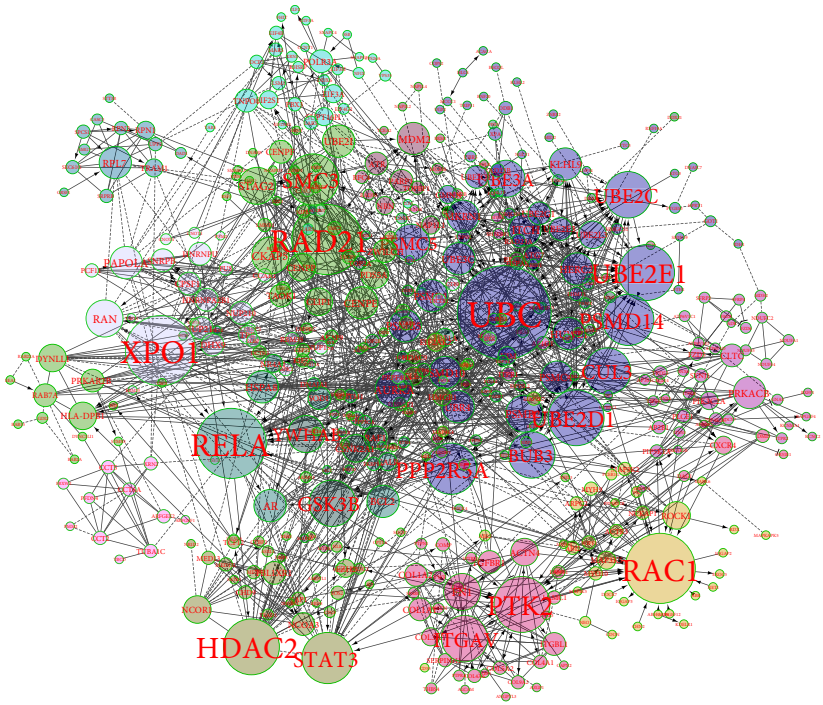



recurrence' based on BCR. They characterized overdetection of recurrence as the detection of recurrent disease that would not progress to clinical metastasis even in the absence of salvage therapy, and estimate its prevalence among men treated with BCR to range between 10 and 37%, depending on patient age at RP and Gleason grade.

Our study design separated patients with BCR into two distinct outcome groups: those without a subsequent metastasis and those that experienced clinical metastasis within 5 years of BCR. We show that patients with BCR and no subsequent clinical metastasis are molecularly similar to treatment-naïve patients who do not experience postoperative BCR with very long follow-up and very distinct from patients with subsequent clinical metastasis. We showed that features detectable at the time of RP provide prognostic information about potentially lethal prostate cancer that is independent of clinical covariates. We observed relatively few differentially expressed genes on the basis of Gleason score and lymph node invasion and most of these were captured in the metastasis signature. This suggests that the addition of genomic data as a direct biological measure of tumour metastatic potential can augment risk stratification based on use of clinical risk factors alone. Furthermore, we found that the metastasis signature identified in the discovery cohort could identify patients who developed early metastasis in an independent dataset. This suggests that the mere detection of BCR after RP may be insufficient to predict the development of clinical metastasis as the tumours of patients who eventually metastasized have unique molecular features that are distinct from those that did not metastasize, irrespective of whether or not they experienced BCR.

The present results seem to provide a genomic basis that supports the findings of the CISNET group that there are really two modes to PSA recurrence with only one leading to early metastasis. This may now be detected early on in the treatment of localized disease by measuring gene expression

**Fig. 4** Functional protein interaction modules of the 965 genes representing the 1 481 metastasis signature. A total of 503 genes are highly interacting and form 13 modules with distinct biology behind them. The network was constructed using Reactome FI plugin in Cytoscape.




differences in primary tumour tissue. Several tissue-based molecular tests are now commercially available in prostate cancer [15]. For example, the Decipher® prostate cancer classifier was developed as a metastasis-specific signature and has been extensively validated as an independent prognostic model outperforming standard clinical risk factors and nomograms [7,8,16,17]. It has been shown to better predict metastasis after BCR [18] and recently Den et al. [16] reported that for patients with low Decipher scores there was no disadvantage in BCR-free and metastasis-free survival when comparing salvage with adjuvant radiotherapy [16]. Furthermore, the researchers found that for patients with high Decipher scores (i.e. high expression of metastasis genes), these patients had improved metastasis-free survival with adjuvant compared with salvage radiotherapy.

The genomic differences found between patients in the BCR and metastasis groups, when deciphering the biology underlying the metastasis signature, led to intriguing observations. About a third of all prognostic features in the metastasis signature were associated with non-coding regions of the genome. Notably, the 2q31.3 locus provided 16 of the 226 prognostic features; other loci of the genome are not as enriched for prognostic markers. This locus was found to be associated with SChLAP1 (Second Chromosome Locus Associated with Prostate-1), which has been recently shown to directly bind and antagonize the activity of the SWI/SNF chromatin modifying complex, a known tumour suppressor [19]. On its own SChLAP1 has been demonstrated to be an independent predictor of metastasis. In the present study, we confirmed those findings in a completely independent analysis and have added to the growing body of evidence supporting a key role for non-coding RNA in prostate cancer disease progression [20].

Several prostate cancer-related signalling pathways, such as androgen, oestrogen and cell cycle signalling pathways, have been shown to be implicated prostate cancer progression. Although in the metastasis signature we found several genes implicated in androgen signalling pathways by gene ontology analysis (e.g. UBE2C, UBE3A, DNAJA1), we did not find significant differential expression in what would be considered canonical androgen signalling genes (i.e. AR, TMPRSS2-ERG). As most of the patients in the metastasis and BCR groups received hormonal therapy at some stage preceding (adjuvant) or after BCR (salvage) or both, the fact that the metastasis group were refractory to hormonal therapy early on (this outcome group was defined by the detection of clinical metastasis within 5 years of BCR) suggests that at diagnosis these patients harboured a tumour with an androgen-insensitive phenotype. Another line of evidence associating the metastasis signature with AR-regulation is the enrichment with gene sets derived from experimental manipulations of LNCaP prostate cancer cell lines with both androgen stimulation and deprivation. A potential mechanism that further drives metastasis in these

**Table 2** Significant ConceptGens enriched in the metastasis signature.

| Concept name | GEO ID | Concept size | Overlap | *P* | Description | PMID |
|---|---|---|---|---|---|---|
| MCF7-Oestrogen induction | GSE11324 | 999 | 176 | 6.31E-44 | Genome-wide analysis of oestrogen receptor binding sites | 17013392 |
| Melanoma cells treated with elesclomol or paclitaxel | GSE11550 | 997 | 142 | 3.65E-25 | Hs 294T cells treated with elesclomol alone or in combination with paclitaxel compared with DMSO-treated | 18723479 |
| MCF7-siTOP1 | GSE7161 | 995 | 127 | 2.29E-18 | Non-classical functions of human topoisomerase i: genome-wide and pharmacological analyses | 17875716 |
| LNCaP androgen deprivation | GSE8702 | 995 | 126 | 6.09E-18 | Longitudinal analysis of progression to androgen independence | 18302219 |
| Ovarian tumour vincristine resistance | GSE7556 | 995 | 126 | 6.09E-18 | Genetic changes in the evolution of multidrug resistance for cultured human ovarian cancer cells | 17726699 |
| LNCaP RSL1 treatment | GSE7223 | 998 | 126 | 7.77E-18 | Genes regulated by the processed form of AIbZIP/CREB3L4 in LNCaP prostate cancer cells | 17712038 |
| MCF7 cells treated with cimicifuga, oestradiol, tamoxifen | GSE6800 | 992 | 122 | 2.18E-16 | Effects of Cimicfuga racemosa, 17β-estradiol, tamoxifen in MCF-7 cells | 17880733 |
| Neuronal cells response to baculoviral | GSE9835 | 996 | 108 | 4.66E-11 | Gene expression changes in response to baculoviral vector transduction of neuronal cells *in vitro* | 19476540 |
| MSC-derived hES differentiation | GSE7879 | 993 | 107 | 8.45E-11 | Comparison of gene expression data between undifferentiated hES cells and MSC-derived hES cells | 21081659 |
| CDK4 activity in neuroblastoma | GSE8866 | 995 | 107 | 9.51E-11 | Cyclin D1 and CDK4 activity contribute to the undifferentiated phenotype in neuroblastoma | 18413728 |
| Effects to HMPV on A549 | GSE8961 | 978 | 105 | 1.61E-10 | Identification of human metapneumovirus-induced gene networks in airway epithelial cells by microarray analysis | 18234263 |
| LNCap DHT treatment | GSE7868 | 992 | 106 | 1.71E-10 | Genome-wide analysis of transcription factors governing androgen receptor-dependent prostate cancer growth | 17679089 |
| Lipoprotein-stimulated human THP-1 macrophages | GSE9101 | 995 | 106 | 2.04E-10 | Expression data in native lipoprotein-stimulated human THP-1 macrophages | NA |
| P63 transcriptional profile | GSE5993 | 994 | 105 | 4.07E-10 | mRNA expression profiling of p63-depleted vs wildtype ME180 cells | 17188034 |
| Response of cycloheximide on MCF7 cells | GSE8597 | 996 | 105 | 4.56E-10 | Gene expression analysis of hormone treated MCF7 breast cancer cells in the presence or absence of cycloheximide | 17986456 |
| Oestradiol effect on MCF10 | GSE5116 | 995 | 104 | 9.02E-10 | Genomic pathways of 17-β-estradiol induced malignant cell transformation in human breast epithelial cells | 18056439 |
| SH-SY5Y neuronal maturational differentiation | GSE4600 | 992 | 100 | 1.32E-08 | Identifying targets of MeCP2 during neuronal maturational differentiation | 16682435 |
| Endometrial response to E2 or MPA | GSE12446 | 997 | 98 | 6.51E-08 | Transcriptome analysis of human endometrial tissues from healthy post-menoupausal women reflecting the endometrial response to 3-weeks treatment with estradiol (E2) and E2+MPA (Medroxy Progesteroneacetate) | 17226044 |
| Myc-regulation in MCF-7 | GSE11791 | 990 | 95 | 3.24E-07 | Oestrogen- and Myc-regulated genes in MCF-7 breast cancer cells | 18714337 |
| Hypoxia response in lymphoma | GSE4086 | 995 | 94 | 7.53E-07 | Hypoxia responsive genes in human Burkitt's lymphoma cell line | 16517405 |
| TGFβ1 modulation in IOSE cells | GSE6653 | 886 | 86 | 7.99E-07 | Gene expression analysis of IOSE cells treated with TGFβ1 | 19615063 |
| MCF7 oestrogen deprivation | GSE10879 | 991 | 82 | 0.0004326 | Expression data of hormone-responsive MCF-7 cells vs oestrogen-deprived MCF-7:5C and MCF-7:2A breast cancer cells | NA |

*Enriched gene concepts from experimental settings were identified using the metastasis signature (965 genes). Concepts related to androgen and oestrogen receptor signalling are highly significant, suggesting a role of both of these sex steroid pathways in prostate cancer progression and metastasis.*

patients could actually be hormonal therapy itself, acting as a stress that further selectively favours clones or 'unmasks' the androgen-insensitive phenotype, as has been observed with neo-adjuvant tumour models [21]. Gene set enrichment analysis shows oestrogen-mediated signalling may also be implicated in prostate cancer progression as the metastasis signature is significantly enriched with genes stimulated by estradiol and tamoxifen in breast cancer cell lines. These genes and the ER pathway may therefore represent alternatives to canonical AR signalling pathways active in primary tumours long before metastasis can be detected [22]. The drug perturbation analysis shed light on the potential of oestrogen-mediated regulation as therapeutic targets in high-risk prostate cancer treatment.

Furthermore, we found that that the metastasis signature has many genes associated with cell cycle regulation, DNA damage sensing, cytoskeleton organization and mitogen-activated protein kinases. Kinase enrichment analysis showed that genes in the metastasis signature are highly phosphorylated by key cell cycle kinases such as ATM, CDK and MAP3K3, which play role in cell cycle checkpoint and control of G2/G1-M transition [23]. DNA damage sensors may be activated in early stages of prostate cancer progression through PRKDC and ATR protein kinases [24], which can induce RAD21 a nuclear phosphoprotein that becomes hyperphosphorylated in cell cycle M phase [25] and has been previously described as a prognostic factor in prostate cancer [5]. Additional investigation of the role of cytoskeleton organization in early prostate cancer progression showed that several metastasis signature genes such as AURKA, CKAP5, TPX2 and RAN were involved in cytoskeleton organization. These genes are involved in microtubule organization and localization to spindle microtubules [26]. Several mitogen-activated protein kinases such as RAF1, MAP3K3, MAP4K4 and MAP3K5 function downstream of Ras family of membrane associated GTPases (e.g. RAC1) [27]. In the metastasis signature functional network, RAF1 that has been shown to activate protein phosphatase 6 (PPP6C), Rho-associated coiled protein kinase (ROCK1) and NF-kB P65 subunit (RELA). These pathways may represent novel targets for biological therapy approaches in high-risk prostate cancer. As chemotherapy is particularly effective in tumours with high proliferation, detection of the metastasis signature therefore may be important for the identification of prostate cancer tumours that may be sensitive to combined chemohormonal therapy, which was shown to be efficacious in the recently completed CHAARTED trial [28].

Exploration of the interactions among the differentially expressed genes that characterized metastasis may guide future studies of drug-targetable pathways and novel agents to address this potentially lethal phenotype. Aparicio et al. [29], have proposed a molecular classification of prostate cancer progression that includes transitioning from an androgen-dependent (endocrine phase) to an androgen-independent (paracrine phase) through a 'tumour-cell autonomous' phase that mimics small cell carcinomas or neuroendocrine prostate cancer. A hallmark of this final cell autonomous phase is the expression of markers such as AURKA, PARP1 and UBE2C. Intriguingly, we found these markers to be overexpressed in the primary tumours of patients who developed metastasis in comparison with the control groups, suggesting that these may be early events in some patients that drive future tumour behaviour and response to therapy. Although chemotherapy, which is the first-line treatment for small-cell or neuroendocrine prostate cancer, is not appropriate for most early-stage localized prostate cancer it is possible that a small subset of cases with overexpression of such 'tumour-cell autonomous' genes already in the primary tumour may benefit from earlier intervention with chemotherapy. Future studies will be needed to test the hypothesis that men with primary tumours that have increased expression of such a metastasis signature may benefit from earlier intervention with chemo-hormonal therapy.

Although our analyses showed that patients with metastasis are molecularly and clinically distinct from those in the BCR and NED groups, there are several limitations of these analyses. First, RNA was extracted from archived (1987–2001) formalin-fixed paraffin-embedded specimens. As the fixation process is known to degrade RNA, some of the expression in the prostate tumour will not be detected, therefore, the metastasis signature identified in the present study may represent only a subset of all the important genes for metastasis. In addition, the gene enrichment analysis based primarily on data mining and computational biology approaches maybe useful for hypothesis generation. Further biological experiments are needed to validate the importance of these genes to the metastatic disease process. Finally, studies in additional cohorts of high-risk prostate cancer are required to confirm the robustness of this genomic information for metastasis prediction across different patient populations.

In summary, we observed diverse biological functions and interactions that may represent important networks in the development of metastatic disease that are established early in the primary tumours of some patients. Mining and analysing the function of the interacting proteins shows that the putative metastasis transcriptional programme stems from several key biological processes not exclusively limited to cell cycle progression. Future efforts to characterize prostate cancer metastasis signatures and the biological significance of these genes and pathways may enable better, targeted therapy approaches tailored to individual patients who are most at risk of potentially lethal prostate cancer.

In conclusion, the present investigation provides further evidence that the biological potential for metastasis can be measured by RNA expression already in the primary tumour and that, in patients with tumours that did not develop metastasis after long follow-up, few if any differences in gene expression could distinguish patients with BCR from those with NED. These findings therefore support the notion that the majority of men treated with surgery have disease that can be effectively managed with local therapy approaches, regardless of whether or not they have rising PSA or BCR. In these cases, if BCR occurs it may signal disease progression that will follow a relatively indolent course that can be effectively managed with postoperative radiation; however, in the subset of patients where BCR precedes rapid onset of metastasis, earlier intervention with systemic therapy may be

required to delay disease progression. Further investigation of the biological pathways that manifest early in the primary tumours of these patients may lead to new therapeutic approaches that can prevent or further delay metastasis and potentially lethal prostate cancer. Genomic signatures of metastasis may play a role in treatment and decisions for adverse pathology and PSA rise in order to better prioritize patients to secondary treatment.

## Acknowledgements

The authors thank Voleak Choeurng, Lucia Lam and Darby Thompson for constructive feedback and discussions during preparation for this manuscript. This study was supported in part by the National Research Council of Canada, Industrial Research Assistance Program, and the Mayo Clinic Prostate Cancer SPORE P50 CA91956.

## Conflict of Interest

M. A., A. C., I. A. V., M. Gh., C. B., N. E., K. Y., T. S., Z. H. and E. D. are employees of GenomeDx Biosciences Inc. The remaining authors have no conflict of interest to declare .

## References

1 **Antonarakis ES, Feng Z, Trock BJ et al.** The natural history of metastatic progression in men with prostate-specific antigen recurrence after radical prostatectomy: long-term follow-up. *BJU Int* 2012; 109: 32–9

2 **Punnen S, Cooperberg MR, D'Amico AV et al.** Management of biochemical recurrence after primary treatment of prostate cancer: a systematic review of the literature. *Eur Urol* 2013; 64: 905–15

3 **Prensner JR, Rubin MA, Wei JT, Chinnaiyan AM.** Beyond PSA: the next generation of prostate cancer biomarkers. *Sci Transl Med* 2012; 4: 127rv3

4 **Cooperberg MR, Broering JM, Carroll PR.** Time trends and local variation in primary treatment of localized prostate cancer. *J Clin Oncol* 2010; 28: 1117–23

5 **Nakagawa T, Kollmeyer TM, Morlan BW et al.** A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. *PLoS One* 2008; 3: e2318

6 **Bastian PJ, Boorjian SA, Bossi A et al.** High-risk prostate cancer: from definition to contemporary management. *Eur Urol* 2012; 61: 1096–106

7 **Erho N, Crisan A, Vergara IA et al.** Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS One* 2013; 8: e66855

8 **Karnes RJ, Bergstralh EJ, Davicioni E et al.** Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at risk patient population. *J Urol* 2013; 190: 2047–53

9 **Wu G, Feng X, Stein L.** A human functional protein interaction network and its application to cancer data analysis. *Genome Biol* 2010; 11: R53

10 **Schroeck FR, Aronson WJ, Presti JC et al.** Do nomograms predict aggressive recurrence after radical prostatectomy more accurately than biochemical recurrence alone? *BJU Int* 2009; 103: 603–8

11 **Taylor BS, Schultz N, Hieronymus H et al.** Integrative genomic profiling of human prostate cancer. *Cancer Cell* 2010; 18: 11–22

12 **Sartor MA, Mahavisno V, Keshamouni VG et al.** ConceptGen: a gene set enrichment and gene set relation mapping tool. *Bioinformatics* 2010; 26: 456–63

13 **Swanson GP, Basler JW.** Prognostic factors for failure after prostatectomy. *J Cancer*. Ivyspring International Publisher 2010; 2: 1–19

14 **Xia J, Trock BJ, Gulati R et al.** Overdetection of recurrence after radical prostatectomy: estimates based on patient and tumor characteristics. *Clin Cancer Res* 2014; 20: 5302–10

15 **Davis J.** Novel commercially available genomic tests for prostate cancer: a road map to understanding their clinical impact. *BJUi* 2014; 114: 320–2

16 **Den RB, Feng FY, Showalter TN et al.** Genomic prostate cancer classifier predicts biochemical failure and metastases in patients after postoperative radiation therapy. *Int J Radiat Oncol* [Internet] 2014; 89: 1038–46. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0360301614005525

17 **Klein EA, Yousefi K, Haddad Z et al.** A genomic classifier improves prediction of metastatic disease within 5 years after surgery in node-negative high-risk prostate cancer patients managed by radical prostatectomy without adjuvant therapy. *Eur Urol* 2014; doi: 10.1016/j.eururo.2014.10.036. [Epub ahead of print]

18 **Ross AE, Feng FY, Ghadessi M et al.** A genomic classifier predicting metastatic disease progression in men with biochemical recurrence after prostatectomy. *Prostate Cancer Prostatic Dis* [Internet] 2014; 17: 64–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24145624

19 **Prensner JR, Iyer MK, Balbin OA.** The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet*. 2013; 45: 1392–8

20 **Vergara IA, Erho N, Triche TJ et al.** Genomic "Dark Matter" in prostate cancer: exploring the clinical utility of ncRNA as biomarkers. *Front Genet* 2012; 3: 23

21 **Sun Y, Wang B-E, Leong KG et al.** Androgen deprivation causes epithelial-mesenchymal transition in the prostate: implications for androgen-deprivation therapy. *Cancer Res* 2012; 72: 527–36

22 **Nelles JL, Hu W-Y, Prins GS.** Estrogen action and prostate cancer. *Expert Rev Endocrinol Metab* 2011; 6: 437–51

23 **Bardelle C, Boros J.** Development of a high-content high-throughput screening assay for the discovery of ATM signaling inhibitors. *J Biomol Screen* 2012; 17: 912–20

24 **Wang Y-G, Nnakwe C, Lane WS, Modesti M, Frank KM.** Phosphorylation and regulation of DNA ligase IV stability by DNA-dependent protein kinase. *J Biol Chem* 2004; 279: 37282–90

25 **Hoque MT, Ishikawa F.** Human chromatid cohesin component hRad21 is phosphorylated in M phase and associated with metaphase centromeres. *J Biol Chem* 2001; 276: 5059–67

26 **Bayliss R, Sardon T, Vernos I, Conti E.** Structural basis of Aurora-A activation by TPX2 at the mitotic spindle. *Mol Cell* 2003; 12: 851–62

27 **Baccarini M.** Second nature: biological functions of the Raf-1 "kinase". *FEBS Lett* 2005; 579: 3271–7

28 **Sweeney C, Chen Y-H, Carducci MA et al.** Impact on overall survival (OS) with chemohormonal therapy versus hormonal therapy for hormone-sensitive newly metastatic prostate cancer (mPrCa): an ECOG-led phase III randomized trial. *J Clin Oncol* 2014; 32(suppl; abstr LBA2): 5s

29 **Aparicio A, Logothetis CJ, Maity SN.** Understanding the lethal variant of prostate cancer: power of examining extremes. *Cancer Discov* 2011; 1: 466–8

Correspondence: Elai Davicioni, GenomeDx Biosciences, Inc., Vancouver, BC, Canada.

e-mail: elai@genomedx.com

Abbreviations: BCR, biochemical recurrence; RP, radical prostatectomy; NED, no evidence of disease.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Fig. S1** Flow diagram of pairwise analyses and fold-change threshold of 1.5 and 2.0.
**Fig. S2** Venn diagram summarizing the overlap from pairwise analysis between using the data with excluded samples and with all samples using F.C:1.5.

**Table S1.** Multivariable generalized linear mixed effect model pairwise comparisons of clinical risk factors.
**Table S2.** Features prognostic of metastasis at a fold-change threshold of 2.0. A total of 16 features were found to be independently prognostic of clinicopathologic variables by a multivariable generalized linear mixed effect model. The cytogenetic band, strand, coding status, associated gene and pathways (where available) and the median fold-change ratio are shown for each feature.
**Table S3.** Gene ontology enrichment analysis of the 13 network modules from the metastasis signature.