Recurrent copy number alterations in prostate cancer: an in silico meta-analysis of publicly available genomic data

Julia L. Williams a, Peter A. Greer a, Jeremy A. Squire a,b,*

a Department of Pathology and Molecular Medicine, Queen’s University, Kingston, Ontario, Canada; b Departments of Genetics and Pathology, School of Medicine of Ribeirao Preto, University of Sao Paulo at Ribeirao Preto, Sao Paulo, Brazil

We present a meta-analysis of somatic copy number alterations (CNAs) from 11 publications that examined 662 prostate cancer patient samples, which were derived from 546 primary and 116 advanced tumors. Normalization, segmentation, and identification of corresponding CNAs for meta-analysis was achieved using established commercial software. Unsupervised analysis identified five genomic subgroups in which approximately 90% of the samples were characterized by abnormal profiles with gains of 8q. The most common loss was 8p (NKX3.1). The CNA distribution in other genomic subgroups was characterized by losses at 2q, 3p, 5q, 6q, 13q, 16q, 17p, 18q, and PTEN (10q), and acquisition of 21q deletions associated with the TMPRSS2-ERG fusion rearrangement. Parallel analysis of advanced and primary tumors in the cohort indicated that genomic deletions of PTEN and the gene fusion were enriched in advanced disease. A supervised analysis of the PTEN deletion and the fusion gene showed that PTEN deletion was sufficient to impose higher levels of CNA. Moreover, the overall percentage of the genome altered was significantly higher when PTEN was deleted, suggesting that this important genomic subgroup was likely characterized by intrinsic chromosomal instability. Predicted alterations in expression levels of candidate genes in each of the recurrent CNA regions characteristic of each subgroup showed that signaling networks associated with cancer progression and genome stability were likely to be perturbed at the highest level in the PTEN deleted genomic subgroup.

Keywords Array CGH, genomic instability, pI3 kinase, bioinformatics, tumor suppressor

Prostate cancer is the most commonly diagnosed malignancy in men and a leading cause of cancer deaths in developed countries (1). Emerging prostate cancer genomic data hold great promise not only in stratifying this heterogeneous disease at biopsy, but also in providing the groundwork for future development of targeted therapies (2).

The frequencies of mutated genes in prostate cancer, which are determined by sequence-based methods, are surprisingly low, with TP53 (17%), TTN (15%), PTEN (11%), MUC16 (9%), and SPOP (8%) (Catalogue of Somatic Mutations in Cancer (COSMIC)) (3). In contrast, the frequency of large-scale copy number alterations (CNAs) and genomic rearrangements is significantly higher, suggesting that the development and progression of prostate cancer is primarily the result of an accumulation of larger-scale genomic aberrations, such as deletions, gains, and fusion gene formation (4–6), instead of more localized mutational events.

Genomic rearrangements leading to the formation of TMPRSS2-ETS gene fusions and deletion of the PTEN tumor suppressor (10q23.31) often occur concurrently, and are the most widely reported genomic biomarkers in prostate cancer (7). The TMPRSS2-ERG gene fusion is the principle ETS family prostate cancer–specific gene fusion, a characteristic signature in approximately one half of prostatic malignancies. PTEN deletions and the TMPRSS2-ERG fusion genes are independently associated with aggressive disease; likewise, concurrent exhibition portends a shorter time to biochemical recurrence and decreased prostate cancer–specific survival (8–12). Metastatic disease is portrayed by heightened genomic complexity as well as an increased frequency of CNAs (7). Hormonal therapies for
the treatment of advanced or recurrent disease often result in genomic amplification of the androgen receptor locus (AR, Xq11.2–q12), which is one of several mechanisms to overcome androgen ablation regimens that results in the development of castration-resistant prostate cancer (13).

In recent years, several prostate cancer cohort microarray studies have been published, but interpretations have been limited by the relatively small sample sizes, and rarely has there been uniformity in the methods used for data analysis between studies (5,6,14–26). Thus, the objective of this study was to perform a systematic review and selection of public domain prostate cancer genomic datasets, followed by a comprehensive meta-analysis of the pooled data of 662 tumors to derive consensus data on the common CNAs. The combined genomic data was then used to identify distinct subgroups and associated candidate pathways of prostate carcinogenesis that could be inferred from the diverse patterns of genomic imbalance. The subgroups that were the most distinct were tumors positive for PTEN deletions and/or TMPRSS2-ERG fusion status. PTEN deletions were significantly associated with a greater percent of the genome being altered (PGA). Predicted changes in expression levels of candidate genes that mapped to recurrent CNA regions showed that signaling networks and canonical pathways associated with cancer progression were more likely to be perturbed in the PTEN deleted subgroups. The large size of this meta-dataset permitted an in-depth survey and examination of concurrent losses and gains that consistently associate within tumors, suggesting that previously unrecognized relationships may exist between specific DNA changes and recurrently targeted signaling pathways.

Methods

Collecting prostate cancer public genomic datasets

High-resolution, human prostate cancer array comparative genomic hybridization (aCGH) or single nucleotide polymorphism (SNP) microarray datasets were collected from supplementary files of published manuscripts and the National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) using "prostate cancer" with "aCGH", "copy number", or "SNP" as keyword combinations. Querying the ArrayExpress (http://www.ebi.ac.uk) database did not reveal any additional datasets. Six Agilent (Santa Clara, CA) aCGH datasets (6,18,19,22–24) and five Affymetrix (Santa Clara, CA) datasets (14,15,17,20,21) were integrated to create a prostate cancer genomic copy number meta-dataset (Table 1).

A total of 879 raw genomic microarray files were collected, including primary and advanced prostate cancers, HGPIN lesions, cell lines, and xenografts, as well as matched normal tissues (Table S1). The overall workflow for this meta-analysis is shown in Figure 1.

A total of 11 datasets were incorporated to build the meta-dataset, which is referenced by GEO accession numbers and PubMed IDs in Table 1. The platform used in each case is mentioned as well as a breakdown of the sample type. The right side of the table displays the number and type of samples that passed quality control inspection, the number and type of unique samples that were included, as well as the number of samples excluded from the analysis and the reason why. High-grade prostatic intraductal neoplasia (HGPIN), cell lines, and xenograft samples were not included in the analysis due to the same sample size and matched benign controls being used as baseline when available for Affymetrix data.

Reported clinical characteristics of patient tumors in the pooled study group

The final pool comprised 568 primary prostate cancer tumors from a total of 545 men who were diagnosed with clinically localized prostate cancer after radical prostatectomy. The primary unique tumors analyzed in this study were derived from eight published manuscripts (6,14,15,17–20,24). Further details specific to the different samples, including the patient with more than one unique tumor, can be found in the Supplemental Methods. In most cases, the tumors were staged using the 2002 TNM classification of malignant tumors (27) and graded according to the revised Gleason Grading system (28). The distribution of the Gleason Grade (available for 350 (61.6%) primary tumors) was as follows: 149 of 350 (42.6%) had Gleason Grade 5 or 6, 156 of 350 (44.6%) had Gleason Grade 7, and 37 of 350 (10.6%) had Gleason Grade 8 or higher. The 161 advanced prostate cancers (158 (98.1%) were distant metastatic lesions) included in this study were derived from 115 men obtained from five published manuscripts (6,19,21–23). Of these advanced cases, 89 (77.4%) were castration-resistant metastatic prostate cancer. Further details, as available, of the clinical characteristics of the 729 tumors included in this study are provided in Table S1.

Sample quality control, inclusion, and exclusion criteria

The raw copy number data files from 568 primary, 161 advanced, 13 HGPIN, and 120 benign control samples, as well as 17 cell line or xenograft samples, were downloaded, which amounted to 879 raw data copy number profiles (Figure 1, Tables 1 and S1). Sample exclusion criteria included: (1) corrupt raw data files (2 samples); (2) raw data of poor quality (5 samples); and, for benign samples, (3) profiles that exhibited large or prostate cancer–like CNAs believed to be potentially contaminated with adjacent tumor tissue (10 samples). A total of 563 primary, 161 advanced, 12 HGPIN, and 110 matched benign samples remained after quality control verification. Cell lines and xenografts were excluded from the analysis. To ensure the integrity of the meta-dataset, only unique samples were included in the analysis. Unique samples were defined as a single sample per patient tumor type. In cases where multiple primary, HGPIN, or metastatic foci were collected for a particular patient, incorporation of the sample harboring the greatest number of aberrations was selected (14,15,21,22). Therefore, final exclusion criteria also included one tumor focus/patient tumor type (62 samples). In only two cases (one primary and one advanced) were two samples kept for an individual patient tumor type, because each sample exhibited an aberrant yet different copy number profile (Figure S1)
Analysis of microarray copy number profiles, specifically deletions of 21q22.2–q22.3 (39.9–42.8 Mb), infers the presence of the classic \(\text{TMPRSS2-ERG}\) gene fusion rearrangement known as “Edel” (6.12) (Figure S3). “Esplit” \(\text{TMPRSS2-ERG}\) gene fusions retain the intervening sequences within the nuclei, in a copy neutral manner; consequently, microarray copy number data cannot always identify this fusion rearrangement. Nevertheless, small deletion(s) at either \(\text{ERG}\) (39.7-39.9 Mb), \(\text{TMPRSS2}\) (42.8 Mb), or both breakpoints were considered indicative of an alternative \(S^\prime\) (androgen regulated or ubiquitously expressed promoter), or \(3^\prime\) (ETS transcription factor) partner, or Esplit fusions, respectively (6).

### Unsupervised hierarchical clustering

Complete linkage hierarchical clustering (CLHC) was performed on the 546 unique primary prostate tumors, which divided the cohort into three groups of genomic alterations. The K-means clustering algorithm \((k = 3)\) was used to determine if any subgroups existed within the 8p-deleted tumors. The five unsupervised primary clusters (A, B, C1, C2, C3) represent the first set of genomic subgroups of this

### Raw data analysis and CNA assignment

Nexus Copy Number, v7.0 (BioDiscovery) is a platform-independent microarray analytical software for the incorporation and co-analysis of genomic microarray datasets from various sources (Figure S2). The algorithms contained within the copy number application of the software perform the normalization, segmentation, and identification of corresponding copy number events from the raw data of all genomic files within a single project, using build 37 (human genome 19) of the genome as the common scaffold for all tumor profiles. Details concerning methods of data integration and probe spatial relationships of the various Agilent and Affymetrix datasets are provided in the Supplemental Methods. Specific settings and algorithm details can also be found in the Supplemental Methods, as well as Tables S1 and S2. To accurately define hemizygous or homozygous deletions and, similarly, gain or amplification (high-level gain), each individual genomic profile was examined, ensuring the probes were centered at a log ratio of 0 or no copy change, while noting the extent of gain or loss in any given region of the genome and utilizing known regions of homozygous (8p, 13q, \(\text{PTEN}\)) and hemizygous (\(\text{PTEN}\), the \(\text{TMPRSS2-ERG}\) (Edel) gene fusion) deletions to accurately choose appropriate limitations of hemizygous and homozygous thresholds (Supplemental Methods).

### Detection of fusion rearrangements by genomic microarrays

Analysis of microarray copy number profiles, specifically deletions of 21q22.2–q22.3 (39.9–42.8 Mb), infers the presence of the classic \(\text{TMPRSS2-ERG}\) gene fusion rearrangement known as “Edel” (6.12) (Figure S3). “Esplit” \(\text{TMPRSS2-ERG}\) gene fusions retain the intervening sequences within the nuclei, in a copy neutral manner; consequently, microarray copy number data cannot always identify this fusion rearrangement. Nevertheless, small deletion(s) at either \(\text{ERG}\) (39.7-39.9 Mb), \(\text{TMPRSS2}\) (42.8 Mb), or both breakpoints were considered indicative of an alternative \(S^\prime\) (androgen regulated or ubiquitously expressed promoter), or \(3^\prime\) (ETS transcription factor) partner, or Esplit fusions, respectively (6).
meta-analysis. Unsupervised clustering segregates based on frequency of copy number alteration without taking biological relevance into consideration, which is why supervised classifications based on rearrangements with clinical significance were subsequently pursued.

**Supervised subgroup classification based on PTEN deletions and fusion status**

The supervised analysis comprised three main comparisons of the cohorts: (1) an overall analysis of the entire cohort of 662 samples (including both primary and advanced tumors); (2) an analysis of the 546 primary tumor samples; and (3) an analysis of the 116 advanced tumor samples. For each comparison, samples were subdivided based on either PTEN deletion or fusion status. Supervised classification based on PTEN deletions or fusion status was performed, which created PTEN intact/loss or FUSION/PTEN intact genotypes.

**Figure 1** Workflow for prostate cancer meta-dataset and combined analysis. The gray and three light red panels depict the steps for sample selection and initial CNA profiling. The pale blue panels demonstrate the analytical steps used to investigate the genomic groupings and association of CNAs with signaling pathways. Superimposed squares indicate the collection, inclusion, and distribution of samples throughout the various steps and analyses. NCBI’s Gene Expression Omnibus (GEO) was searched using the keywords “prostate cancer” with “aCGH”, “SNP” or “copy number”. A total of 11 high-resolution human prostate cancer genomic datasets were collected, which contained 879 individual samples related to prostate cancer patients (Table 1). The 879 genomic samples were queried using BioDiscovery’s (Santa Clara, CA) Nexus Copy Number microarray software, and the FASST2 segmentation algorithm was applied to determine CN profiles for each sample. Following quality assessment, 546 unique primary and 116 unique advanced samples were incorporated. Copy number events were ascertained for each sample, which permitted an examination of differences within and between the primary and advanced populations. Stratification of the primary cohort (blue panels) was performed using several different mechanisms to determine the most informative genomic groupings of prostate cancer. Unsupervised clustering of the primary prostate cancers produced five genomic subgroups: A, B, C1, C2, and C3. Supervised classification based on PTEN deletions or the gene fusion was also performed, which created PTEN intact/loss and FUSION−/−ve genotypic subdivisions. PTEN and fusion gene levels were compared between primary and advanced samples. Assessment of genomic instability and exploration of perturbed signaling pathways was performed to explore biological and potential clinically significant differences between the various genomic subdivisions.
CNA associations were graphically depicted using circos plots from the concordance function in Nexus Copy Number, v7.0 (BioDiscovery).

Genomic instability assessment

The frequency of samples harboring losses and gains at specific genomic intervals commonly gained or lost in primary prostate cancer were enumerated and compared between genomic groups (see Table S3). The genomic instability metric, the percentage of genome altered (PGA), is the percentage of base pairs lost or gained relative to the entire genome (NCBI hg19 database (GRCh37.p13) comprising 3,095,677,412 base pairs; http://www.ncbi.nlm.nih.gov/projects genome/assembly/grc/human/data/index.shtml). Examining genomic instability by these two approaches (frequency of CNAs across the genome and PGA) provided a much more comprehensive appreciation of the integrity of the genome.

Pathways implicated in different genomic subgroups based on imputed expression levels

Since only one of the eight CNA datasets had accompanying gene expression data, it was necessary to impute relative levels of gene expression weighted by whether a particular region of interest was subject to loss (i.e., reduced expression levels assigned) or to gain (i.e., increased expression assigned). To select genes within a given genomic interval, arbitrary cut-offs above background were chosen to identify the genes specific to each subgroup and to identify signaling pathways that may be perturbed by the CNAs that are characteristic of a particular subgroup. Gene lists were compiled for each genomic subgroup: genes recurrently deleted (in ≥25% and 5% of the samples for hemizygous and homozygous deletions, respectively) or gained (in ≥10% of the samples). In this way, simulated expression levels were created based on the genes that were present in a given subgroup, taking into account the direction of copy number change and proportion of samples within that subgroup with the specific alteration (Table S4). Genes in regions of polymorphic copy number variation or in close proximity to telomeres or centromeres were excluded. Benign profiles were consulted and demonstrated similar peaks in these same regions (Figure S4). The “Core Analysis” function in Ingenuity Pathway Analysis (Ingenuity Systems, http://www.ingenuity.com) was used to explore each of the eight genomic subgroups in the context of biological functions and canonical pathways. For the general settings, the “Ingenuity Knowledge Base (Genes Only)” reference set was used as the population of genes to consider for P-value calculations, and both direct and indirect relationships were included. For the analysis filter summary, we considered only molecules and/or relationships where species equaled human or mouse or rat. The significance of the biological functions and canonical pathways implicated as perturbed in a subgroup’s gene list was determined by the Fisher exact test P-value. For each analysis, a maximum of 25 molecular interaction networks containing a maximum of 35 nodes each were generated.

Statistical analysis

The association between particular CNAs was assessed using the chi-square test, and differences in the frequencies of CNAs between subgroupings were computed using the Student t-test. The comparisons between the various subgroups were performed using the Fisher exact test. All statistical analyses were performed using IBM SPSS Statistics, v21 (IBM Corporation, Armonk, NY).

Results

The frequencies of recurrent CNAs were calculated across the entire genome and grouped into deletions and gains (Table 2, Figures 2 and S5) for all 662 prostate cancer samples (546 primary tumors and 116 advanced cohort) derived from the 11 genomic copy number datasets. A total of 14 regions of recurrent deletion and 5 regions of recurrent gain were identified in the combined dataset (Table 2). The cytobands spanning the region, the entire region (Location), the minimal region of overlap (MRO), and genes of particular interest to prostate cancer are listed for each CNA.

Systematic PubMed literature and cancer Gene Census (29) searches supplemented by recent COSMIC (3,30) findings were used to establish candidate driver genes for each region of imbalance shown under the profiles in Figure 2. Mining of Progenetix prostate cancer data (592 chromosomal CGH and 95 aCGH samples) confirmed the genes and regions of imbalance identified in this meta-analysis (31,32).

Recurrent deletions

As previously reported (4), deletion of chromosome 8p (8p23.3–p11.21; 1–43 Mb) was the most recurrent CNA observed in the prostate cancer genome, with 409 of 662 (61.8%) prostate tumors being affected. Losses of 8p were found in 304 of 546 (55.7%) and 105 of 116 (90.5%) cases of localized and advanced prostate carcinoma, respectively. The minimal region of deletion occurs in the interval 8p21.3–p21.2, which contains the prostate-specific tumor suppressor, NKX3.1, at cytoband 8p21.2. Deletions in 8p display a unimodal distribution of loss extending in both directions along the 8p arm. Telomerically, the deletion frequency diminishes to approximately 40% mid-8p22.2 before decreasing to approximately 35% of samples exhibiting deletion of the 8p telomeric region. Centromerically, the deletion frequency decreases to approximately 25% (approximately 20% in primary and approximately 35% in advanced) at 8p11.21. The deletion profile showed a sharp decline at the 8p pericentromeric region in the primary tumors and returned to a copy neutral or gained state on 8q; however, in approximately 19% of advanced tumors, the deletion interval was much larger, extending into the q arm at 8q11.21 and diminishing below background by 8q13.1 (66 Mb).

The second most frequent deletion detected in prostate cancer is the extensive genomic region surrounding the RB1 tumor suppressor gene, comprising the following cytobands, 13q13.1–q31.1 (33–86 Mb), with the deletion frequently
encompassing the 13q telomeric region in advanced tumors. A total of 350 of 662 (52.9%) prostate cancers, (245 of 546 (44.9%) primaries, and 105 of 116 (90.5%) advanced cases) harbored a deletion within this genomic interval, with the minimal region of deletion occurring at cytoband 13q14.2, which contains the RB1 gene locus. The configuration of the 13q deletion has a bimodal trend, with a weak secondary peak at 13q21.33 (73 Mb) in the primary cohort.

### Table 2

<table>
<thead>
<tr>
<th>Region</th>
<th>Cytobands</th>
<th>Location (Mb)</th>
<th>MRO (Mb)</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusion</td>
<td>21q22.2–q22.3</td>
<td>40.0–42.8</td>
<td>39.9 &amp; 42.8</td>
<td>ERG, TMPRSS2</td>
</tr>
<tr>
<td>PTEN</td>
<td>10q23.31</td>
<td>Gene only</td>
<td>89.6–89.7</td>
<td>PTEN</td>
</tr>
<tr>
<td>Del_2q</td>
<td>2q14.1-q24.2</td>
<td>117–160</td>
<td>136.2–137.2</td>
<td>CXCR4</td>
</tr>
<tr>
<td>Del_3p13</td>
<td>3p13</td>
<td>71–74</td>
<td>71.9–72.9</td>
<td>FOXP1, RYBP, SHQ2</td>
</tr>
<tr>
<td>Del_5q</td>
<td>5q11.2–q23.3</td>
<td>54.8–128</td>
<td>66.8–73.2 &amp; 98–103</td>
<td>CHD1, APC</td>
</tr>
<tr>
<td>Del_6q</td>
<td>6q12–q22.33</td>
<td>63.9–129.7</td>
<td>88.1–93.1</td>
<td>MAP3K7</td>
</tr>
<tr>
<td>Del_8p</td>
<td>8p23.3–p11.21b</td>
<td>2.9–42.1</td>
<td>21.9–25.1</td>
<td>NIKK2-J</td>
</tr>
<tr>
<td>Del_10q</td>
<td>10q23.2–q26.12</td>
<td>88–131.4</td>
<td>89.6–89.7</td>
<td>PTEN</td>
</tr>
<tr>
<td>Del_12p</td>
<td>12p13.31–p12.3</td>
<td>7.8–17.9</td>
<td>11.8–13.8</td>
<td>CDKN1B</td>
</tr>
<tr>
<td>Del_13q</td>
<td>13q12.3–q13.31</td>
<td>32.1–110.3</td>
<td>48.1–50.1</td>
<td>BRCA2, RB1</td>
</tr>
<tr>
<td>Del_16q</td>
<td>16q11.2–q24.3</td>
<td>46.6–88.8</td>
<td>85.2–90.2</td>
<td>CDH1</td>
</tr>
<tr>
<td>Del_17p</td>
<td>17p13.3–17p11.2</td>
<td>2.2–18.3</td>
<td>7.5</td>
<td>TP53</td>
</tr>
<tr>
<td>Del_17q</td>
<td>17q11.2–q21.31</td>
<td>40.5–42.3</td>
<td>41.1–41.7</td>
<td>BRCA1, ETV4</td>
</tr>
<tr>
<td>Del_18q</td>
<td>18q12.1–q23</td>
<td>26.1–76.7</td>
<td>71.2–76.7</td>
<td>SMAD4, BCL2</td>
</tr>
<tr>
<td>Gain_3q</td>
<td>3q13.33–q27.3</td>
<td>120.9–187.8</td>
<td>140–170</td>
<td>PIK3CA, ETV5</td>
</tr>
<tr>
<td>Gain_5q</td>
<td>5q11.2–q23.3</td>
<td>48.8–142.9</td>
<td>128.8</td>
<td>MYC</td>
</tr>
<tr>
<td>Gain_7</td>
<td>7p22.3–q36.2</td>
<td>1.5–153</td>
<td>whole chromosome gain</td>
<td>ETV1, EGRF, MCM7, BRAF</td>
</tr>
<tr>
<td>Gain_8q</td>
<td>8q11.21–q24.3</td>
<td>48.8–142.9</td>
<td>128.8</td>
<td>MYC</td>
</tr>
<tr>
<td>Gain_16q</td>
<td>16q11.2–q23</td>
<td>26.1–76.7</td>
<td>71.2–76.7</td>
<td>SMAD4, BCL2</td>
</tr>
<tr>
<td>Gain_17q</td>
<td>17q11.2–q21.31</td>
<td>40.5–42.3</td>
<td>41.1–41.7</td>
<td>BRCA1, ETV4</td>
</tr>
<tr>
<td>Gain_18q</td>
<td>18q12.1–q23</td>
<td>26.1–76.7</td>
<td>71.2–76.7</td>
<td>SMAD4, BCL2</td>
</tr>
<tr>
<td>Gain_20q</td>
<td>20q11.22–q13.33</td>
<td>32.2–62.9</td>
<td>55–62.9</td>
<td>-</td>
</tr>
</tbody>
</table>

Abbreviation: MRO, minimal region of overlap.
*Primary and Advanced combined.
Aberration extends in advanced cases.

Figure 2 Recurrent genomic copy number alterations in prostate cancer. Snapshot from Nexus Copy Number microarray analytical software (BioDiscovery). The genome is displayed horizontally, and the frequency of CNA at any given location is displayed on the y-axis. Red and dark red represent regions of hemizygous and homozygous deletions, respectively; blue and dark blue represent regions of gain and amplification, respectively. Systematic COSMIC, PubMed literature, and cancer Gene Census searches identified candidate driver genes for each region of imbalance. *Genes that have been identified as alternative 3' or 5' prostate cancer gene fusion partners.
Deletions at 16q22.1–q24.3 (entire 16q arm in advanced disease) and 6q12–q22.33 followed as the third and fourth most common deletions observed in prostate cancer. The remaining recurrent deletions, in order of prevalence in primary disease were: 5q11.2–q23.3, 17p13.3–p11.2, 10q23.2–q26.12, 18q, 21q22.2–q22.3 (gene fusion), 12p13.31–p12.1, 2q14.1–q24.1, 3p13, and 17q21.2–21.31.

**Recurrent gains**

Chromosome 8q was the most frequent gain observed in this meta-analysis dataset, with 211 of 662 (31.9%) tumors harboring extra copies of 8q. Chromosome 8q gain was identified in 114 of 546 (20.9%) and 97 of 116 (83.6%) primary and advanced cases, respectively. The pattern of gain along the long arm of chromosome 8q has an extensive region of consistent gain across the entire arm (8q11.21–q24.3) in both primary and advanced disease. Within this region of 8q gain, advanced disease cases often have a small focal region of high gain at the cytoband 8q24.21, which corresponds to the MYC oncogene (Table 2). Additional gains observed in primary disease are quite infrequent but include gain of 16p (100 of 546, 18.3%), the entire chromosome 7 (77 of 546, 14.1%), 3q (53 of 546, 9.7%), and 20p (23 of 546, 4.2%).

Gain of chromosome 7 was also frequent, with 164 of 662 (24.8%) prostate tumors harboring gains encompassing the entire chromosome when primary and advanced tumors were combined. Gains in the 16p arm (100 of 546, 18.3%, and 74 of 116, 63.8%, in primary and advanced tumors, respectively) were identified at a higher frequency than chromosome 7 gains in the primary cohort.

**Gleason Scores and CNAs**

Genomic copy number profiles of the primary cohort stratified by Gleason Score were created for the 309 primary tumor samples from which data were available for comparison. The 133 cases of Gleason 6 were indistinguishable from the 141 Gleason 7 cases when overall profiles were examined. There was a more marked distribution of CNAs, which suggested a greater frequency of alterations in the 33 tumors with Gleason ≥8 (Figure S6).

**Genomic rearrangements specific to advanced disease**

Gain of chromosome 7 was the second most common gain and was identified in 74 of 116 (63.8%) advanced disease cases; this is in stark contrast to the low frequency and third ranked position of chromosome 7 gains in primary disease. Subsequently, the AR gene locus on chromosome X (Xq12; 76 of 116, 65.5%) ranks third in advanced disease gains. The extra copies of the AR gene found in the advanced disease cohort relative to the 2.9% observed in the primary tumor cohort was consistent with mechanisms to overcome the androgen blockade imposed by antiandrogens and chemical castration treatments to combat advanced recurrent prostate carcinoma. Chromosomes 16p (63.8%), 3q (61.2%), and 20q (51.7%) were also identified as frequent regions of gain in advanced disease.

**Unsupervised clustering identified three major genomic subgroups**

Unsupervised hierarchical clustering was used to generate a dendrogram of the 546 unique primary prostate cancer samples (Figure 3). The cluster analysis produced three main genomic branches. Further segregation beyond three clusters using the CLHc algorithm produced an increasing number of small clusters, while always maintaining a large proportion of samples in a single cluster. This was not informative, for example, when the CLHc was used to define the six clusters into which the cohort was divided: 52 (9.5%), 55 (10.1%), 356 (65.2%), 38 (6.9%), 6, (1.1%), and 39 (7.1%) (Figure S7). The most informative CLHc produced three subdivisions: a group that was genomically benign and lacked any consistent CNAs ("A"); a group that lacked the most common CNA, 8p deletions ("B"); and the 8p-deleted tumors that contained a wide range of additional CNAs ("C").

Subgroup A lacked any consistent genomic imbalances (n = 55, 10.1% of the samples). Subgroups B and C each had 6q losses accompanied by gains of chromosomes 7 and 8q. The subgroup B (n = 52, 9.5% of samples) branch of the dendrogram was classified by the absence of 8p deletions. The other CNAs within the branch included losses at 2q, 5q, 13q, and 16q. The most prevalent subgroup, C (n = 439, 80.6% of samples), was characterized by the presence of 8p deletions together with several other large-scale CNA alterations.

The 8p-deleted samples ("C", n = 439) were partitioned into three related genomic subgroup branches (C1-3), based on their individual CNA profiles. K-means clustering permitted segregation of the large cohort of 8p-deleted tumors (K-means was uninformative when used on the entire primary cohort; see Figure S8). Subgroup C1 was characterized by 2q, 6q, 8p, and 13q deletions; however, chromosomal losses at 5q, 16q, 17p, and 18q, as well as PTEN and fusion rearrangements, were not apparent in this subgroup (n = 55, 10.1%). In contrast, subgroups C2 and C3 harbored the majority of the PTEN deletions and fusion genes, with several additional chromosomal losses commonly identified in prostate cancer. One apparent distinction of C2 tumors is the absence of 13q deletions (n = 52, 9.5%), whereas C3 tumors (n = 332, 60.8%) had a greater frequency of CNAs across the entire genome, including the common 13q deletion. Although the overall levels of differential genomic CNA observed within each of the subgroup C arms was minor, the pattern of alterations suggested that acquisition of PTEN genomic deletions and/or the TMPRSS2-ERG fusion may underlie the distinction between C1 and C2/C3 tumors. We therefore undertook a more detailed evaluation of genomic subgroupings based on the presence or absence of the PTEN deletion and TMPRSS2-ERG fusions in the cohort.

Unsupervised clustering (CLHc and K-means) of the advanced cohort was unsuccessful because samples were disproportionately allocated into the derivative clusters (data not shown).
PTEN deletions and TMPRSS2-ERG gene fusions

PTEN deletions and ETS gene fusions are frequently concurrent events in prostate cancer; thus, CNAs affecting these regions were examined in more detail (8,33). PTEN deletions were identified in 205 of 662 (31.0%) prostate cancer cases (124 of 546 (22.7%) primary and 81 of 116 (69.8%) advanced cases). The PTEN deletions were then characterized as either hemizygous or homozygous deletions (Table 3), where 96 and 28 of 546 (17.6% and 5.1%) primary cancers harbored a detectable hemizygous or homozygous deletion, respectively (Figures S9 and S10).

PTEN deletion frequency and fusion status for the meta-dataset are presented in Table 3. PTEN loss is stratified by the extent of deletion, and positive fusion status is arranged according to the mechanism of rearrangement: Edel or another mechanism. The frequencies of the PTEN and fusion supervised classifications are displayed in the table. A significantly greater proportion of advanced samples harbored both rearrangements when compared with the primary cohort.

Conversely, advanced disease samples not only harbored an elevated frequency of PTEN genomic losses (69.8% vs. 22.7%), but also a significantly greater proportion (Fisher exact test, \( P > 0.0001 \)) of homozygous deletions: 18 (15.5%) and 63 (54.3%) of 116 advanced tumors harbored a detectable hemizygous or homozygous deletion, respectively. The deletion pattern observed on chromosome 10 in primary disease corresponds very specifically to an interstitial focal deletion centered on the PTEN genomic locus. In cases where additional regions are involved, the deletion more frequently extends in a telomeric direction, which is consistent with the PTEN deletion fluorescence in situ hybridization (FISH) findings from our laboratory (34). The rearrangements on chromosome 10 also exhibit a more complex pattern of chromosome 10 imbalance in the advanced cohort compared to the overall trend of focal interstitial deletions targeting the PTEN genomic locus in primary disease. Furthermore, when comparing PTEN deleted samples to those without PTEN deletions in the primary cohort, tumors with an intact PTEN locus had almost no detectable CNAs targeting chromosome 10 (Figure 4viii and 4ix).

The most prevalent TMPRSS2-ERG fusion rearrangement, "Edel", results from deletion of 21q22.2--q22.3 fusing the 5’ untranslated region of TMPRSS2 (21q22.3) to the coding sequence of ERG (21q22.2). This typical CNA was identified in 145 of 662 (21.9%) prostate cancers, with 104 of 546 (19.0%) and 41 of 116 (35.3%) samples from primary and advanced disease, respectively (Figures S3 and S11).
Another 33 of 546 (6.0%) and 15 of 116 (12.9%) were found to harbor a deletion at either the ERG breakpoint (suggestive of an alternate ETS partner), the TMPRSS2 breakpoint (indicating an alternate 3' ETS partner), or at both breakpoints (representative of the “Esplit” TMPRSS2-ERG gene fusion rearrangement where the intervening sequences are retained but relocated elsewhere in the genome) (6,12). A significant association occurred between the PTEN deletions and gene fusions by the chi-square statistic ($P < 0.001$) in both primary and advanced cohorts. A significant shift ($P > 0.0001$) also occurred in the distribution when comparing primary and advanced cohorts with the smallest fraction of primary tumors (9%), but the majority (42%) of the advanced tumors harbored both rearrangements (Table 3). The advanced cohort was significantly enriched for both PTEN deletions and/or the gene fusion (75.9% compared with 38.8% of primary samples). These data suggested that a supervised analysis of the genomic influence of a PTEN deletion and gene fusion rearrangements on the primary cohort may be highly informative.

Supervised classification with respect to PTEN deletions and fusion genes

To determine how the presence or absence of a PTEN deletion and fusion rearrangement affected the overall CNA distribution within the primary cohort, we undertook a series of comparisons of their combined effect on profiles in Figure 4i–4iv, followed by an evaluation of their individual effects in Figure 4v–4x.

The genomic profiles were initially subjected to a four-way supervised classification in which the 546 primary tumors were used to generate CNA profiles from the subset bearing neither a PTEN deletion nor fusion rearrangements (see Figure 4i). Similarly, the subset derived from samples with just the gene fusion is shown in Figure 4b, those with just a PTEN deletion in Figure 4iii, and those samples bearing both a PTEN deletion and a fusion rearrangement in Figure 4iv. By comparing the two profiles in Figure 4i and 4iv to the two profiles in Figure 4ii and 4v, it is possible to directly infer the overall influence of a PTEN deletion in comparison to the fusion rearrangement. It can be seen that strong similarities exist between profiles in Figure 4i and 4ii (both with intact PTEN), with only minor CNAs evident along the genome. Likewise, the profiles in Figure 4iii and 4iv, both with PTEN deletions, are very similar. The presence of a PTEN deletion seems to be strongly associated with more extreme copy number differences and a wider distribution of CNAs along the genome. Collectively, the profiles in Figure 4i–4iv suggest that the presence of a PTEN deletion may have a more general influence on the extent of CNAs in other regions of the genome. In contrast, the fusion rearrangement has less overall influence on the distribution or level of CNAs along the genome.

Figure 4v–4vii provides profiles of the CNA intensity and distribution along the genome in which a pairwise comparison of the consensus from samples without a fusion (Figure 4v) is made against samples with a fusion rearrangement (Figure 4vi). In Figure 4vii, the fusion-negative imbalance profile was used as a baseline reference against the fusion-positive profile to obtain the enriched CNA distribution that is strongly associated with the presence of a fusion rearrangement. It can be seen that losses at 3p13, 8p, 10q, 16q, and 17p accompany the 21q deletion that was used to define fusion events.

Table 3 Frequency of PTEN and fusion rearrangements in prostate cancer

<table>
<thead>
<tr>
<th></th>
<th>Prostate cancer (N=662)</th>
<th>Primary (N=546)</th>
<th>Advanced (N=116)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTEN hemizygous loss</td>
<td>114 (17.2)</td>
<td>96 (17.6)</td>
<td>18 (15.5)</td>
</tr>
<tr>
<td>PTEN homozygous loss</td>
<td>91 (13.7)</td>
<td>28 (5.1)</td>
<td>63 (54.3)</td>
</tr>
<tr>
<td>PTEN loss</td>
<td>205 (31.0)</td>
<td>124 (22.7)</td>
<td>81 (69.8)</td>
</tr>
<tr>
<td>TMPRSS2-ERG Edel</td>
<td>145 (21.9)</td>
<td>104 (19.0)</td>
<td>41 (35.3)</td>
</tr>
<tr>
<td>Fusion other</td>
<td>48 (7.3)</td>
<td>33 (6.0)</td>
<td>15 (12.9)</td>
</tr>
<tr>
<td>Fusion (total)</td>
<td>193 (29.2)</td>
<td>137 (25.1)</td>
<td>56 (48.3)</td>
</tr>
<tr>
<td>PTENintact/fusion-nega</td>
<td>362 (54.7)</td>
<td>334 (61.2)</td>
<td>28 (24.1)</td>
</tr>
<tr>
<td>PTENintact/fusion-positi</td>
<td>95 (14.4)</td>
<td>88 (16.1)</td>
<td>7 (6.0)</td>
</tr>
<tr>
<td>PTENloss/fusion-negative</td>
<td>107 (16.2)</td>
<td>75 (13.7)</td>
<td>32 (27.6)</td>
</tr>
<tr>
<td>PTENloss/fusion-positi</td>
<td>98 (14.8)</td>
<td>49 (9.0)</td>
<td>49 (42.2)</td>
</tr>
</tbody>
</table>

*Primary and Advanced combined.*
Figure 4 Influence of PTEN deletion and fusion rearrangement on CNAs. The genome is displayed horizontally, and the frequency of CNA at any given location is displayed on the y-axis. Panels i–iv show a four-way classification of CNAs, which is stratified according to status of fusion gene rearrangements and PTEN deletion status as indicated in each panel. It can be seen that strong similarities exist between panels i and ii (both PTEN intact), with only a small number of CNAs evident along the genome. Panels iii and iv (both PTEN deleted) also have very similar overall profiles, but a much more extensive and extreme distribution of CNAs is present. In contrast, the presence (panels ii and iv) or absence (panels i and iii) of the fusion rearrangement did not seem to influence the overall CNA profile. Panel v shows all samples with a fusion, and panel vi shows all samples without a fusion rearrangement. In panel vii, the fusion-negative imbalance profile was used as a baseline reference against the fusion-positive profile to obtain the consensus CNA distribution that is associated with the presence of the fusion rearrangement. In panel viii, PTEN gene intact samples can be compared to PTEN deleted samples (panel ix). In panel x, PTEN intact imbalances are used as a baseline reference against PTEN deleted samples to show the consensus CNAs associated with PTEN deleted samples. It can be seen that the influence of PTEN deletion leads to more extreme copy number differences and a wider distribution of CNAs along the genome (panel x). (Figure Continues)
Genomic profiles emerge from the primary cohort: genotypically unstable PTEN deleted tumors and simpler PTEN intact tumors.

**CNAs and candidate genes associated with PTEN deletions or fusion positivity**

A statistical comparison of the regions of CNAs associated with either the presence of the fusion gene or the presence of a PTEN deletion was made using circos plots (Figure 4i—4ii). In these analyses, the candidate driver genes (see Figure 2) were assigned to regions of recurrent imbalance. The fusion-positive samples were associated with deletions at 3p13 (FOXP1, EIF4E3, RYBP, and SHQ1, $P = 5.95 \times 10^{-5}$), 8p (NKX3-1, $P = 0.003$), 10q23.31 (PTEN, $P = 7.7 \times 10^{-4}$), 16q ($P = 2.4 \times 10^{-4}$), and 17p (TP53, $P = 4.8 \times 10^{-5}$) (Figure 4xi). PTEN deleted samples (Figure 4xii) had a high concordance with deletions at 3p13 (FOXP1, EIF4E3, RYBP, and SHQ1, $P = 1.95 \times 10^{-6}$), 8p (NKX3-1, $P = 2.3 \times 10^{-12}$), 10q regions adjacent to PTEN ($P = 1.1 \times 10^{-11}$), 13q (RB1, $P = 3.4 \times 10^{-6}$), 16q ($P = 5.4 \times 10^{-5}$), 17p (TP53, $P = 5.9 \times 10^{-19}$), and 21q22.2–q22.3 (gene fusion, $P = 1.95 \times 10^{-6}$), as well as gain of 8q (MYC, $P = 6.8 \times 10^{-7}$). Based on the magnitude of the P-values, the likelihood of association between PTEN and its recurrent CNA regions was more significant than the association levels found for the fusion gene. Taken together, these findings continue to support the notion that PTEN deletions are sufficient to impose higher levels of CNA and overall genetic diversity.

**Genomic instability metric**

To compare overall levels of genomic instability in those tumors with a PTEN genomic deletion to those without an apparent deletion, the percentage of base pairs lost or gained relative to the entire genome was used to calculate the PGA. A significant difference ($P < 0.001$) in PGA was observed when segregating the primary cohort based on deletion of the PTEN genomic region (Table S5, PTEN intact, PGA = 5.4%; and PTEN loss, PGA = 13.3%). There was also a significantly greater frequency and number of CNAs as well as a larger PGA in samples with deletion of the PTEN locus (Figure 5A and 5B). Conversely, division of the cohort based on fusion status did not yield a significant difference ($P = 0.25$) in terms of CNA frequency or PGA (Table S5, fusion-negative, PGA = 6.9%; and fusion-positive, PGA = 7.9%). The frequency of deletions was higher in the fusion-positive samples only for 8p and 16q. Similarly, no difference was observed with respect to the frequency of gains across the genome (Figure 5C and 5D).

**Pathways implicated by the influence of CNAs on gene expression in genomic subgroups**

Ingenuity Pathway Analysis “Core Analysis” determined significant networks, biological functions, and canonical pathways associated with each genomic subgroup: B, C1, C2, C3; PTEN intact, PTEN loss; fusion-positive, and fusion-negative (Figure 6, Figure S12).

Cancer was the most significant biological function across all eight subgroups; however, the PTEN loss group had the most significant number (1,461) of input genes overlapping with the “Cancer” function. “DNA Replication, Recombination, and Repair” and “Cell Cycle” were found in the top five “Molecular and Cellular Functions” and were significantly altered in the unsupervised genomic subgroups containing PTEN genomic deletions (B (33 genes), C2 (46 genes), and C3 (44 genes)) but not in the C1 group, which lacked PTEN deletions. The “Telomere Extension by Telomerase”

---

Figure 4  (Continued) Panels xi–xii provide circos plots of a statistical comparison of the regions of CNAs associated with PTEN deletion (panel xi) and fusion rearrangement (panel xii). In these analyses, the candidate driver genes from Figure 2 have been assigned to the regions of recurrent imbalance.
canonical pathway was found to be significantly associated with the C genomic subgroups, which suggested a connection with the 8p deletion, the common factor between these samples. Differences between the supervised classification subgroups based on the fusion gene were almost completely lacking; 3, 2, and 2 of the top “Diseases and Disorders”, “Molecular & Cellular Functions”, and “Physiological System Development and Function”, respectively, were the same between both fusion-negative and fusion-positive samples.

The PTEN deletion that was used to supervise classification of the primary cohort successfully established that the majority of PTEN deleted genomes were associated with cancer progression and metastasis. The top five PTEN deleted “Molecular & Cellular Functions”, which comprised 1,029 genes, were: Cellular Movement; Cell Death and Survival; Cellular Growth and Proliferation; Cellular Development; and Drug Metabolism. The top five PTEN intact “Molecular & Cellular Functions”, which comprised only 163 genes, were:
unsupervised genomic subgroups and supervised classification, despite the fact that it was found in the top five. Both somatic mutations (37,40,41). Although localized sequence physical position (rearrangements) being more common than level remains relatively low, with changes in copy number or (36
subtypes based on somatic mutations in prostate cancer individual subgroups.

be deregulated by changes in gene expression arising from further support the significance of stratifying primary prostate cancer based on PTEN deletion status.

Discussion

Prostate cancer has a variable clinical course, and recent characterizations of CNAs have revealed marked genomic heterogeneity that may help explain the differences in prognosis (35). One of the historical limitations of attempting genomic subgroup analysis has been the small cohort size of individual studies. This is the first meta-analysis to address this problem and to systematically examine publicly accessible prostate cancer genomic datasets to derive a consensus map of the most frequent CNAs in 662 prostate cancers comprising 546 primary tumors and 116 from advanced disease. This large cohort size allowed us to stratify the CNAs into distinct genomic subgroups, and to demonstrate that there were potential pathways that might be deregulated by changes in gene expression arising from the recurrent genomic imbalance that characterized the individual subgroups.

At present, less evidence exists regarding molecular subtypes based on somatic mutations in prostate cancer according to next generation sequencing approaches (36–39). The recurrent mutation frequency at the sequence level remains relatively low, with changes in copy number or physical position (rearrangements) being more common than somatic mutations (37,40,41). Although localized sequence mutations do occasionally target important cancer genes, albeit at a low frequency, CNAs seem to be much more prevalent. Since recurrent CNAs can significantly reduce the expression levels of tumor suppressors (PTEN, TP53, RB1) by deletion, as well as increase the expression of oncogenes (MYC, ERG) by gains, amplifications, or rearrangements, there remains considerable interest in classifying prostate cancer based on the patterns of recurrent CNAs.

In keeping with the published literature, deletions at 8p (NKX3-1), 13q (RB1), 16q, and 6q were the most frequent overall CNA losses observed in the cohort. Consistent interstitial focal deletions included the tumor suppressor genes PTEN at 10q23.31 and TP53 at 17p13.1. Other frequent focal deletions involved 3p13 (FOXP1, RYBP, and SHQ1 gene losses) and the 2.9 Mb deletion at 21q22.2–q22.3, which is required to generate the common TMPRSS2-ERG genomic fusion. The most frequent CNA gains affected the majority of the 8q arm (MYC) and all of chromosome 7.

PTEN deletions were reported in 20–30% of primary prostate cancers, and were recognized to have prognostic significance with respect to earlier biochemical relapse (9,41) and prostate cancer—specific death (33). Studies of castration-resistant prostate cancer also showed that PTEN deletions were present at a much higher frequency (50–70%) (42,43). Although the prognostic significance of ETS rearrangements remains unresolved (reviewed in (44)), it is well-established that ERG fusion rearrangements are more likely to occur when PTEN is deleted, and that the presence of both alterations influences prostate cancer prognosis (8,33,42). Consistent with these findings, both of these genomic alterations were found to be significantly associated in the overall cohort in this study. Moreover, there was a significant marked increase in the proportion of cases harboring both rearrangements in the 116 advanced disease samples. Specific regions particularly associated with PTEN deletions included deletions of 8p (NKX3-1, prostate-specific tumor suppressor), 10q, 13q (RB1), 16q (TP53), and the TMPRSS2-ERG gene fusion, as well as gain of 8q (MYC).

Regions associated with the fusion rearrangement were deletions at 3p13 (FOXP1), 8p (NKX3-1), PTEN deletions, 16q, and 17p (TP53), whereas fusion-negative samples had a weak association with deletions of 2q (CXC4).

Our analysis of the influence of PTEN deletion on the overall pattern of CNAs in the primary meta-analysis cohort demonstrated that more marked copy number differences occurred, which affected wider regions of the genome, when PTEN was lost. Thus, the overall pattern of CNA appeared to be strongly influenced by the presence or absence of PTEN genomic deletion. In contrast, the presence or absence of the gene fusion did not seem to influence the overall trends of CNAs in the entire dataset.

The PTEN protein is thought to maintain genomic stability through pAKT-dependent mechanisms that inactivate CHEK1 by pAKT, thus permitting cell cycle progression to continue in the presence of DNA damage (45). If this model is correct, then when genomic deletion of PTEN takes place, the absence, or reduced levels, of PTEN protein will lead to pAKT activation and the accumulation of CNAs as a result of unresolved DNA damage. However, there have been conflicting results with respect to non-canonical mechanisms of PTEN’s function in maintaining genomic integrity. Several groups have shown that lack of PTEN protein correlates with a reduction in RAD51 foci (46,47) and an increase in γH2AX foci (48,49), whereas other groups have demonstrated no association of PTEN with RAD51 expression or foci (50,51).

| Subgroup B | 7 | 656 |
| Subgroup C1 | 13 | 249 |
| Subgroup C2 | 24 | 535 |
| Subgroup C3 | 14 | 498 |
| Fusion-negative | 14 | 399 |
| Fusion-positive | 14 | 469 |
| PTEN intact | 17 | 227 |
| PTEN loss | 50 | 773 |

Figure 6 The number of significantly perturbed canonical pathways and biological processes for each primary prostate cancer genomic subgroup are identified in this figure.
Thus, it seems likely that PTEN loss could be involved in multiple cell cycle progression checkpoints, some of which have downstream effects leading to defective DNA double-strand break repair.

In addition to genomic deletion, the PTEN protein can be lost as a result of indels, small DNA point mutations, or promoter methylation of the PTEN gene (30,42). Thus, the possibility exists that our supervised analyses of PTEN genomic losses based on CNA detection may not have detected all tumors with a PTEN protein loss, and may have introduced bias into our analysis of the influences of PTEN deletion on overall levels of CNAs. However, since it is estimated that only 11% of prostate cancers have intragenic mutations of PTEN (3) and aberrant promoter methylation of the PTEN gene appears to be rare (42), our supervised classification of PTEN deletions is unlikely to have been significantly affected.

Collectively, this in silico meta-analysis has demonstrated that deletion of PTEN is associated with genomic instability and other oncogenic changes, such as formation of the TMPRSS2 gene fusion and loss of tumor suppressor genomic regions containing RB1 and TP53. Moreover, the overall pattern of genomic change found in PTEN deleted primary tumors most closely resembles the CNA distribution present in metastatic disease (Figure 2). These findings are in agreement with a crucial oncogenic role for PTEN as a primary driver gene in prostate cancer (10,52,53). Interestingly, a recent in silico study of expression datasets, based on integrative network analysis, showed that the PTEN protein was an independent prognostic biomarker of biochemical recurrence-free survival in prostate cancer (54). This finding is in keeping with our pathway analysis, which also indicated that PTEN deleted tumors had the most significantly distorted biological functions, highlighting cancer, cell death/survival, and cellular movement in keeping with the PTEN protein’s pivotal prognostic role.

The costs and feasibility of sequencing and CNA analysis are continually decreasing. Thus, using publicly available genomic data to provide novel genomic subclassifications in tumors will become increasingly important. Such approaches provide combined evidence from multiple studies to indicate which biologically relevant pathways should be pursued for clinical utility, prognostic significance, and stratification of patients.

Acknowledgments

The authors acknowledge the assistance of the BioDiscovery team for their continued technical support. Our research was supported, in part, by the Terry Fox Foundation training program in transdisciplinary cancer research, in partnership with the Canadian Institutes of Health Research (CIHR) (J.W.), the Canadian Cancer Society (grant no. 018124) (J.S.), and the Brazilian CNPq (grant nos. 478492/2011-1 and 33.654.831/0001-36) (J.S.).

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cancergen.2014.09.003.

References


