Functional Analysis of Genes in Regions Commonly Amplified in High-Grade Serous and Endometrioid Ovarian Cancer

Sally J. Davis1,4, Karen E. Sheppard2,4,5,6, Richard B. Pearson2,5,6, Ian G. Campbell1,4,6, Kylie L. Gorringe1,4,6, and Kaylene J. Simpson3,4,6

Abstract

Purpose: Ovarian cancer has the highest mortality rate of all the gynecologic malignancies and is responsible for approximately 140,000 deaths annually worldwide. Copy number amplification is frequently associated with the activation of oncogenic drivers in this tumor type, but their cytogenetic complexity and heterogeneity has made it difficult to determine which gene(s) within an amplicon represent(s) the genuine oncogenic driver. We sought to identify amplicon targets by conducting a comprehensive functional analysis of genes located in the regions of amplification in high-grade serous and endometrioid ovarian tumors.

Experimental Design: High-throughput siRNA screening technology was used to systematically assess all genes within regions commonly amplified in high-grade serous and endometrioid cancer. We describe the results from a boutique siRNA screen of 272 genes in a panel of 18 ovarian cell lines. Hits identified by the functional viability screen were further interrogated in primary tumor cohorts to determine the clinical outcomes associated with amplification and gene overexpression.

Results: We identified a number of genes as critical for cellular viability when amplified, including URI1, PAK4, GAB2, and DYRK1B. Integration of primary tumor gene expression and outcome data provided further evidence for the therapeutic use of such genes, particularly URI1 and GAB2, which were significantly associated with survival in 2 independent tumor cohorts.

Conclusion: By taking this integrative approach to target discovery, we have streamlined the translation of high-resolution genomic data into preclinical in vitro studies, resulting in the identification of a number of genes that may be specifically targeted for the treatment of advanced ovarian tumors. Clin Cancer Res; 19(6); 1411–21. ©2013 AACR.

Introduction

Ovarian cancer is associated with the highest mortality rate of all gynecologic malignancies and is identified as the seventh most common cause of cancer-related death for women, with approximately 140,000 deaths annually worldwide (1). This high mortality rate is commonly attributed to the diagnosis of advanced stage disease and is often exacerbated by limited treatment options. Current treatment for ovarian cancer involves cytoreductive surgery combined with platinum-based chemotherapy. In most cases, this treatment regime is initially successful, with approximately 80% of patients responding to the first-line treatment (2). However, the majority of patients relapse within 12 months and currently there are no curative treatments for recurrent disease.

"Molecular-targeted agents” are a new class of cancer therapeutics that directly interfere with the biologic events that drive tumorigenesis. The increased selectivity of these agents towards malignant cells has the potential to not only increase the efficacy of cancer treatment, but also to facilitate a more combinatorial approach to cancer therapeutics without the risks associated with overlapping toxicities (3–5). Unlike traditional chemotherapeutics, which exert their effect upon rapidly proliferating cells, molecular-targeted agents are designed against genetic events that distinguish malignant cell populations from normal cells. Inherent to the development of molecular-targeted agents is the identification of driver genomic events unique to tumor cells. In high-grade ovarian cancer, recent studies have indicated that somatic mutual activation of oncogenes is a rare event (6). However, somatic copy number amplification is...
highly prevalent indicating that this is a common mechanism of oncogene activation in this tumor type. Copy number amplification generally results in the overexpression of genes harbored within the amplified chromosomal region. A number of high-resolution genomic studies have identified recurrent copy number amplifications that occur in ovarian tumors (6–8). However, the insights provided by such studies regarding the identity of the driver gene(s) is limited due to the fact that a large proportion of genes within amplified regions will not impart a functional effect. In this study, we describe the largest systematic functional assessment of genes located in frequently amplified regions of the ovarian tumor genome in an extensive panel of ovarian cancer cell lines that recapitulate the amplifications observed in primary samples.

We sought to capitalize on our high-resolution copy number analysis of high-grade serous and endometrioid tumors (7) by integrating those data with RNA interference (RNAi) to identify amplification-specific genetic dependencies in ovarian cancer cells. Unlike previous studies that focus on individual regions of amplification (9, 10), we have designed a boutique siRNA library targeting every protein-coding gene located within regions that are frequently amplified in high-grade serous and endometrioid ovarian cancer. Using high-throughput screening technology, we have been able to simultaneously assess the loss-of-function phenotypes associated with each gene and identify novel candidate genes including URI1, GAB2, and PAK4 that may be targeted for the treatment of high-grade serous and endometrioid ovarian tumors.

Materials and Methods

**Amplicon selection**

Amplicons were selected on the basis of the following procedure. First, a peak finding algorithm (11) on the combined copy number (CN) frequency of single-nucleotide polymorphism (SNP) array data from 398 ovarian cancer samples (7) was conducted to identify genomic regions with log2 ratio thresholds of more than 0.8 at 5% frequency and more than 1 at 2.5% frequency. Each peak was extended to encompass the segmentation boundaries of 3 samples (at a CN threshold of $>$0.8) or 1 sample (at a CN threshold of $>$1). Genomic Identification of Significant Targets in Cancer (12) was also used to identify amplification peaks in the subset of 240 cases with SNP6 data. These loci were included if they satisfied the same frequency requirement. The union of all peaks was then assessed and any loci delineated by a copy number polymorphism were removed or adjusted.

**Cell culture**

Ovarian cell lines were obtained from the National Cancer Institute (OVCA3, OVCA8, OVCAR8, SKOV3, Caov3), American Type Culture Collection (ATCC: Caov4), European Collection of Cell Cultures (ECACC: OAW28, 59M), The Institute of Physical and Chemical Research (RIKEN: JHOC9, JHOS3), Health Science Research Resources Bank (HSRRB: Kuramochi, MCAS), Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ: EFO21, FUOV1), Stephen Howell at University of California, San Diego (2008), and Dr Nuzhat Ahmed at the Women’s Cancer Research Centre, Royal Women’s Hospital, Melbourne (OVCA432). Cell lines were maintained according to the conditions outlined in Supplementary Table S1. We used short tandem repeat (STR) profiling to verify the authenticity of the panel of cell lines. We amplified 8 STR loci using primer sequences listed in Supplementary Table S2. All profiles were compared with the ATCC database or STR profiles published on Catalogue of Somatic Mutations in Cancer (COSMIC; http://www.sanger.ac.uk/genetics/CGP/cosmic), where possible. STR analysis occurred both before and after siRNA screening to ensure that cell line misidentification was avoided.

**High-throughput siRNA screening**

Cell lines were grown and transfected with SMARTpool siRNAs using Dharmafect (DF) 1, 2, or 3 lipid transfection reagents (Thermo Fisher Scientific) as indicated in Supplementary Table S1. The boutique siRNA library of 272 siGENOME complexes (gene targets and catalog numbers in Supplementary Table S3) was obtained in a 384-well plate format from Dharmacon RNAi Technologies (Thermo Fisher Scientific), hydrated, and diluted in the Victorian Centre for Functional Genomics (VCFG) to 1 nmol using 1× siRNA buffer (Dharmacon). Each well in this library contained a SMARTpool of 4 distinct siRNA species targeting different sequences of the target transcript. Cells were reverse transfected (13) at 30% confluence to a final concentration of 40 nmol/L with the SMARTpool library using the CellClone ALH3000 (Caliper Life Sciences) and BioTek 406 (BioTek) liquid handling robotics. Cell lines were transfected in duplicate and 2 independent biologic replicates were screened. Each cell line was optimized to reach confluence at 72 hours after transfection. At that point, cell viability for each well was quantified on the basis of the direct measurement of intracellular ATP using the Cell Titer Glo luminescent assay (Promega) at a 1:4 dilution. All luminescent measurements were taken on the Synergy H4 high-throughput multimode microplate reader (BioTek). To identify loss of viability, data was normalized and processed as described by Brough and colleagues (14).
Briefly, raw luminescent signals from each well were log transformed, plate centered by median, and normalized by Z score. The Z scores generated in each replicate screen were averaged to generate a single Z score for each data point that was then ranked to identify screening positives or "hits".

Validation of siRNA gene silencing using real-time quantitative PCR

Quantitative real-time PCR (qRT-PCR) was used to validate siRNA-mediated gene silencing efficiency and specificity. Transfection conditions were scaled to a 96-well format, cells were transfected with 40 nmol/L SMARTpool siRNA targeting UR11, DYRK1B, GAB2, BMP8B, ZFP36, CACNA1C, SAMD4B, and PAK4, and RNA was collected 24 hours after transfection. RNA was extracted from cells using the RNeasy Mini Kit (Qiagen), in accordance with the manufacturer's instructions. cDNA was generated from 100 ng of RNA using the Superscript III VILO kit (Invitrogen), in accordance with the manufacturer's instructions. Expression of UR11, DYRK1B, GAB2, BMP8B, ZFP36, CACNA1C, SAMD4B, and PAK4 was assessed using primers listed in Supplementary Table S2 on the Lightcycler 480 (Roche Diagnostics). PCR was conducted using the Brilliant II SYBR Green QPCR Mix (Agilent Technologies) and relative expression was determined according to the comparative Ct (ΔΔCt) method: ΔΔCt = Gene siRNA DNA − Ct Non-targeting siRNA DNA. ΔCt (threshold cycles) is the Ct of the reference gene minus the Ct of the target gene. Remaining gene expression was calculated by 2−ΔΔCt.

Hypoxanthine phosphoribosyltransferase 1 (HPRT1) was used as an endogenously expressed reference gene for the purpose of quantifying relative gene expression. Each qRT-PCR was conducted in triplicate using cDNA derived from 3 independent experiments.

Copy number and gene expression in primary tumors

We obtained both SNP copy number data (n = 388) and Agilent gene expression data (n = 399) of high-grade serous ovarian tumors from The Cancer Genome Atlas (TCGA) and high-grade serous and endometrioid ovarian tumor Affymetrix U133Plus2 expression (n = 254) and SNP copy number (n = 159) from the Australian Ovarian Cancer Study [AOCS, Tothill and colleagues ref. (15) and Etemadmoghadam and colleagues ref. (16)] and Gorringe and colleagues (7). Gene 1.0ST expression data with matching copy number (n = 48) but with no survival information was obtained from Ramakrishna and colleagues (17). For Affymetrix U133 array data, the probe with the highest median value for each gene across all samples was used. Copy number data (TCGA and Gorringe and colleagues; ref. 7) was normalized and segmented as described (7). The genomic segment with the highest copy number ratio overlapping each gene was selected to represent that gene. Gene expression values were compared with matching copy number values using a Spearman correlation.

Survival analysis was conducted on copy number and expression data from the TCGA and AOCS datasets using the Cox proportional hazards model. Genes with a low median expression value across samples within a platform were excluded: TCGA Agilent <4 (BMP8B, CACNA1C, SAMD4B, DYRK1B) and AOCS U133 <7 (BMP8B, CACNA1C, DYRK1B). Gene expression was treated as a continuous variable, with residual disease at none, up to 1 cm, or more than 1 cm as a covariate. Copy number was treated as categorical comparing loss/neutral (CN log2 ratio <0.1) with gained (CN log2 ratio >0.5) with residual disease as a covariate.

Results

Selection of ovarian cell lines for analysis

We identified 25 regions of the genome with frequent high-level copy number amplifications in high-grade serous and endometrioid ovarian cancer based on our previous study (ref. 7; Table 1). To specifically address the functional role of genes targeted by these amplifications, we first sought to select ovarian tumor cell lines that recapitulate the genetic changes that were observed in clinical samples. We obtained data from SNP-based high-resolution copy number studies for 36 ovarian cell lines (K.E. Sheppard, unpublished data; COSMIC) to identify lines with gene amplifications similar to those identified in clinical samples (data not shown). We identified 17 cancer cell lines (Table 1) for our analysis, with most amplicons represented in at least 3 independent cell lines. However, chromosome 19 amplicons at 19p13.13 and 19p13.12 were only represented in a single cell line and no appropriate cell line was identified for the chromosome 20p13 amplicon.

Epithelial ovarian cancer encompasses a class of ovarian tumors and includes a number of different histologic subtypes including serous, endometrioid, mucinous, and clear cell, which are further differentiated by low- and high-grade status. Although we defined our amplified regions based upon data collected from primary high-grade serous and endometrioid tumors, cell lines for this analysis were selected on the basis of their copy number characteristics independent of histologic subtype, which has resulted in some heterogeneity in the subtypes represented in this panel. Cell lines derived from clear cell (2/17, 12%), low-grade (Type I) serous (1/17, 6%), high-grade (Type II) serous (11/17, 64%), endometrioid (2/17, 12%), and mucinous (1/17, 6%) tumors are included in this analysis (Supplementary Table S1).

Screening the same set of candidate genes through 17 different cell lines provides internal controls for the specific detection of amplicon-associated phenotypes. An additional nontumorigenic ovarian epithelial cell line (IOSE532) was also included in this study to identify genes that are essential for normal ovarian survival. This cell line does not harbor any of the amplifications found in the tumor-derived cell lines. The identity of all cell lines was confirmed by microsatellite genotyping (data not shown, see Materials and Methods).
Identification of amplicon target genes using siRNA knockdown screening

A siRNA library targeting 272 annotated genes located within regions of genomic amplification in high-grade serous and endometrioid ovarian cancer was used to transiently silence the expression of these genes in 18 ovarian cell lines. The siRNA library was arrayed in a 384-well format to facilitate the simultaneous transfection of all genes in a single experiment (see Materials and Methods). After 72 hours, cell viability was measured using the Cell Titer Glo luminescent assay (Promega), which allows for quantitation of cell number based on levels of cellular ATP. The heterogeneity of cell line morphology and growth characteristics made it difficult to integrate all the data generated by the screen. To compare the data collected across multiple cell lines, we applied the normalization strategy of Brough and colleagues (14) to our data set to generate $Z$ scores as a normalized distribution.

<table>
<thead>
<tr>
<th>Region</th>
<th>Location (Mb)$^a$</th>
<th>Tumors with CN &gt;1 (%)$^b$</th>
<th>Tumors with CN &gt;0.8 (%)</th>
<th>No. genes</th>
<th>Cell lines with CN gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p34.2</td>
<td>39.372–41.577</td>
<td>3.0</td>
<td>5.0</td>
<td>31</td>
<td>CAOV3, CAOV4, OAW28, KURAMOCHI, OVCA432</td>
</tr>
<tr>
<td>1p34.2</td>
<td>42.913–43.07</td>
<td>3.0</td>
<td>4.3</td>
<td>7</td>
<td>CAOV3, CAOV4, KURAMOCHI, OVCA432</td>
</tr>
<tr>
<td>1q21.1</td>
<td>145.415–145.831</td>
<td>2.5</td>
<td>4.5</td>
<td>3</td>
<td>CAOV4, KURAMOCHI, OVCAR4, JHOS3</td>
</tr>
<tr>
<td>3q26.2</td>
<td>170.244–170.406</td>
<td>4.8</td>
<td>12.3</td>
<td>2</td>
<td>EFO21, OVCAR3, JHOS3, CAOV3, CAOV4, FUOV1, OAW28, OVCAR4, 2008, OVCA432, JHOC9, SKOV3</td>
</tr>
<tr>
<td>3q26.31</td>
<td>173.564–174.253</td>
<td>1.8</td>
<td>8.0</td>
<td>6</td>
<td>EFO21, JHOS3, CAOV3, CAOV4, FUOV1, OAW28, 2008, JHOC9</td>
</tr>
<tr>
<td>3q26.32</td>
<td>178.213–178.589</td>
<td>2.3</td>
<td>7.0</td>
<td>1</td>
<td>EFO21, OVCAR3, JHOS3, CAOV3, CAOV4, FUOV1, OAW28, 2008, SKOV3</td>
</tr>
<tr>
<td>3q29</td>
<td>194.369–199.306</td>
<td>4.8</td>
<td>7.3</td>
<td>46</td>
<td>EFO21, OVCAR3, JHOS3, CAOV3, CAOV4, FUOV1, OAW28, OVCAR4, 2008, OVCA432, SKOV3, KURAMOCHI</td>
</tr>
<tr>
<td>6p21.1</td>
<td>40.899–42.031</td>
<td>2.5</td>
<td>3.0</td>
<td>22</td>
<td>JHOS3, CAOV3, OVCAR8</td>
</tr>
<tr>
<td>8q23.3</td>
<td>112.645–114.649</td>
<td>2.5</td>
<td>9.8</td>
<td>1</td>
<td>JHOS3, 59M, OAW28, KURAMOCHI, JHOC9</td>
</tr>
<tr>
<td>8q24.11</td>
<td>117.721–118.789</td>
<td>3.3</td>
<td>11.1</td>
<td>6</td>
<td>JHOS3, 59M, CAOV4, OAW28, KURAMOCHI, OVCAR8, JHOC9</td>
</tr>
<tr>
<td>8q24.12</td>
<td>121.039–121.517</td>
<td>3.8</td>
<td>12.6</td>
<td>3</td>
<td>EFO21, JHOS3, 59M, KURAMOCHI, JHOC9, OVCAR8</td>
</tr>
<tr>
<td>8q24.21</td>
<td>128.494–129.146</td>
<td>7.0</td>
<td>18.3</td>
<td>1</td>
<td>JHOS3, 59M, CAOV4, FUOV1, KURAMOCHI, JHOC9, OVCAR4, OVCAR3, OVCAR8</td>
</tr>
<tr>
<td>8q24.3</td>
<td>141.725–142.15</td>
<td>4.8</td>
<td>15.1</td>
<td>1</td>
<td>JHOS3, 59M, CAOV3, KURAMOCHI, JHOC9, OVCAR4, OVCAR8</td>
</tr>
<tr>
<td>8q24.3</td>
<td>146.202–146.275</td>
<td>2.3</td>
<td>6.3</td>
<td>1</td>
<td>JHOS3, 59M, CAOV3, KURAMOCHI, JHOC9, OVCAR4, OVCAR8</td>
</tr>
<tr>
<td>11q14.1</td>
<td>75.913–78.31</td>
<td>3.3</td>
<td>5.0</td>
<td>25</td>
<td>EFO21, ES2, OVCAR3</td>
</tr>
<tr>
<td>12p13.33-p13.32</td>
<td>1.205–4.214</td>
<td>3.5</td>
<td>6.3</td>
<td>19</td>
<td>ES2, CAOV3, CAOV4, FUOV1, OAW28</td>
</tr>
<tr>
<td>12p12.1</td>
<td>24.236–26.396</td>
<td>4.3</td>
<td>6.5</td>
<td>13</td>
<td>EFO21, JHOS3, MCAS, CAOV4, FUOV1, OAW28, KURAMOCHI, SKOV3</td>
</tr>
<tr>
<td>19p13.13</td>
<td>12.992–13.761</td>
<td>3.5</td>
<td>7.0</td>
<td>10</td>
<td>EFO21</td>
</tr>
<tr>
<td>19p13.12</td>
<td>15.021–15.314</td>
<td>2.8</td>
<td>6.5</td>
<td>7</td>
<td>EFO21</td>
</tr>
<tr>
<td>19q12</td>
<td>34.914–35.121</td>
<td>8.8</td>
<td>11.8</td>
<td>1</td>
<td>EFO21, CAOV4, FUOV1, OVCAR4, KURAMOCHI, OVCAR3, OVCAR8</td>
</tr>
<tr>
<td>19q12</td>
<td>35.135–35.198</td>
<td>9.0</td>
<td>11.6</td>
<td>1</td>
<td>EFO21, CAOV4, FUOV1, OVCAR4, KURAMOCHI, OVCAR3, OVCAR8</td>
</tr>
<tr>
<td>19q13.2</td>
<td>44.011–45.086</td>
<td>2.8</td>
<td>5.0</td>
<td>36</td>
<td>CAOV3, CAOV4, FUOV1, KURAMOCHI, OVCAR3</td>
</tr>
<tr>
<td>20p13</td>
<td>0.086–0.178</td>
<td>1.3</td>
<td>5.5</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>20q11.21</td>
<td>29.397–30.249</td>
<td>1.8</td>
<td>5.5</td>
<td>21</td>
<td>EFO21, JHOS3, CAOV4, FUOV1, OAW28, OVCAR4, 2008, KURAMOCHI, OVCAR3, SKOV3</td>
</tr>
<tr>
<td>20q13.2</td>
<td>50.408–52.416</td>
<td>2.3</td>
<td>5.5</td>
<td>5</td>
<td>EFO21, JHOS3, MCAS, CAOV4, OVCAR4, 2008, KURAMOCHI, OVCAR3, JHOC9</td>
</tr>
</tbody>
</table>

$^a$Build 36.1 2006 locations.
$^b$From Gorringe and colleagues (36).
across all the samples where negative Z scores represent loss of viability.

Cell line screening performance was assessed using a number of statistical methods as follows. Variability between replicates was calculated according to the percentage coefficient of variation (%CV), which measured an average of 5.6% across replicate experiments for each gene in all cell lines (Supplementary Fig. S1). We also used the Z' factor (18) as a means to establish the dynamic range between the positive and negative controls to provide a measure of high-throughput screen performance. We used PLK1 as a positive control for cell death in all cell lines except EFO21, MCAS, and OAW28, which were not affected by loss of PLK1. DLL3 was used as a positive control for these cell lines. Nontargeting control siRNA was used as a negative control for all cell lines assessed. The Z' factor data for all cell lines was more than 0 with an average of 0.44, indicating a high-quality screen (ref. 19; Supplementary Fig. S1).

For the purpose of this screen, “hits” were defined as those siRNAs that resulted in an average Z score of < −1.5. This threshold was measured to be the equivalent of an average 40% decrease in cellular viability when compared with confluent cells (P < 0.0001) in all cell lines assessed (Supplementary Table S4). Applying this threshold, we identified 74 genes that when silenced cause a statistically significant decrease in cellular viability (Table 2 and Supplementary Table S5). Approximately half of the genes identified (35/74, 47%) induced significant viability effects across multiple cell lines. The gene that significantly induced cell death in the largest number of cell lines was PTK2 (also referred to as FAK), affecting viability in 9 independent cell lines. Notably, PTK2 is the subject of ongoing clinical trials in tumors that overexpress PTK2 including (but not limited to) breast, ovary, colon, liver, lung, and prostate tumors (GlassoSmithKline 2012; clinicaltrials.gov identifier NCT01138033) and in hematologic malignancies such as lymphocytic leukemia (National Cancer Institute 2012; clinicaltrials.gov identifier NCT00563290). In addition, ribosomal genes such as RPL35A, RPS16, and HNRNPL were prominent with many hits in multiple cell lines. At least 1 of these 4 ribosomal genes was detected as a hit in 67% of the cell lines analyzed (12/18; Supplementary Table S5).

Since the primary purpose of this study was to identify genes that are amplified in ovarian cancers and also have a functional affect on tumorigenesis, all data were subjected to a 2-tailed t test to compare the Z scores in cell lines with amplification at a specific gene locus to cell lines that were copy number neutral or showed copy number loss at that locus. Applying this test, 8 of 74 (11%) of the genes were significantly positively associated with gene amplification at P < 0.05, namely CACNA1C, BMP8B, Dyrk1B, Gab2, Pak4, Samd4b, URI1, and Zfp36 (Fig. 1 and Supplementary Table S5). Interestingly, Hnrnppl and Tpx2 were found to have an inverse association with amplification. That is, cell lines harboring amplification of the HNRNPL or TPX2 locus were protected from the decreased cellular viability effects that were observed in cell lines without amplification of this locus. The effect observed with siRNA HNRNPL may be a more general response as the viability of IOSE532 cells was also significantly impaired following loss of HNRNPL expression (Supplementary Table S5). Loss of Tpx2 did not induce a significant effect in IOSE532 cells (Supplementary Fig. S2 and Supplementary Table S5) and this inverse relationship between Tpx2 amplification and Tpx2 gene knockdown has also been reported by others (9). HNRNPL and Tpx2 were excluded from further analysis.

To ensure that the phenotypes observed in the screen were attributable to on-target gene silencing, we quantified the efficiency of silencing by qRT-PCR in the 8 amplification-associated hits. Expression of each target gene was evaluated by qRT-PCR in the 8 amplification-prone cell lines (ZLOC2, BACAS1, BCAT1, BCL9, BMP8B, RASSF8, URI1, CACNA1A, CASC1, CASP14, CLDN19, CYF24A1, DLG1, Dyrk1B, ECH1, ELKS, SLFN1, XXYLT1, RINL, FR35, Gab2, GMF5, MED29, NCCRP1, KIAA0754, Kras2, MDF1, C1orf50, Ndufc2, Spata16, Pak4, Plagl2, Sars2, Scmh1, Sox5, SynC, Tfe2, Thrsp, Med20, Zdhhc19). Additionally, PTX2 was also significantly impaired following loss of HNRNPL expression (Supplementary Table S5). Loss of Tpx2 did not induce a significant effect in IOSE532 cells (Supplementary Fig. S2 and Supplementary Table S5) and this inverse relationship between Tpx2 amplification and Tpx2 gene knockdown has also been reported by others (9). HNRNPL and Tpx2 were excluded from further analysis.

### Table 2. Summary of candidate genes identified by siRNA screen

<table>
<thead>
<tr>
<th>Number of affected cell lines (Z &gt; −1.5)</th>
<th>Genes*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BCAS1, BCAT1, BCL9, BMP8B, RASSF8, URI1, CACNA1A, CASC1, CASP14, CLDN19, CYF24A1, DLG1, Dyrk1B, ECH1, ELKS, SLFN1, XXYLT1, RINL, FR35, Gab2, GMF5, MED29, NCCRP1, KIAA0754, Kras2, MDF1, C1orf50, Ndufc2, Spata16, Pak4, Plagl2, Sars2, Scmh1, Sox5, SynC, Tfe2, Thrsp, Med20, Zdhhc19</td>
</tr>
<tr>
<td>2</td>
<td>ATP13A5, BCL2L1, CAP1, CPN2, PIGX, FOXP4, CEP19, MYC, MYLK2, NCBP2, OPA1, PAK1, PP1H, SELV, TRFC, USP49, Zfp36</td>
</tr>
<tr>
<td>3</td>
<td>BYSL, CACNA1C, DEFB127, FTtd1, C11orf67, SupT5h</td>
</tr>
<tr>
<td>4</td>
<td>FBL, MEOM, SAMD4B, ZNF684</td>
</tr>
<tr>
<td>5</td>
<td>RAD21, Tpx2</td>
</tr>
<tr>
<td>6</td>
<td>Dll3, Hnrnppl, Rps16</td>
</tr>
<tr>
<td>7</td>
<td>Rpl3A</td>
</tr>
<tr>
<td>8</td>
<td>Atp13A4</td>
</tr>
<tr>
<td>9</td>
<td>Ptk2</td>
</tr>
</tbody>
</table>

*Genes highlighted in bold are significantly associated with gene amplification.

Assessment of gene hits in primary tumors

Having identified 8 genes with amplification-associated viability effects (Fig. 1 and Supplementary Table S5) in cell lines, we sought to validate this association in primary high-grade serous and endometrioid primary ovarian tumors.
Normalized gene expression data, copy number data, and survival information from patients with ovarian cancer was obtained from recent publications [TCGA; ref. 6], AOCs (15, 16), Ramakrishna and colleagues (17), Gorringe and colleagues (7)]. A significant positive correlation between copy number and gene expression was observed with all 8 genes in at least 1 dataset (Table 3) and for 5 genes across all 3 datasets (GAB2, URI1, SAMD4B, DYRK1B, and PAK4). Discordant results between datasets were frequently observed when the measured expression of the gene was low, which could indicate lack of expression or poor probe performance (CACNA1C, BMP8B).

We also looked for an association of altered copy number and expression with overall and progression-free survival using the TCGA (6) and AOCs (15) datasets. It is well recognized that an inverse relationship exists between post-operative residual disease and patient survival (20–22); consequently, we included residual disease in the model as a known prognostic factor (Table 4). Elevated URI1 was consistently associated with worse outcome. GAB2 amplification and increased expression was associated with improved overall survival, but not progression-free survival (Fig. 2). Expression of BMP8B or CACNA1C was not significantly associated with overall or progression-free survival; however, their expression was low across the cohorts. The 4 genes on the 19q13.2 amplicon (DYRK1B, SAMD4B, PAK4, and ZFP36) were not consistently associated with survival across the 2 cohorts, but this may be a reflection of the limited power of the AOCs CN cohort.

**Discussion**

Copy number amplification has long been recognized to be a mechanism of oncogene activation. However, the cytogenetic complexity and heterogeneity of high-grade serous and endometrioid ovarian tumors has made it difficult to determine which gene(s) within an amplicon represent the genuine oncogenic driver. Previous studies that have used copy number amplification as an aid to functional identification of novel oncogenes have been limited to single regions of recurrent amplification (9, 10, 23). We have used high-throughput siRNA screening technologies to conduct the first systematic analysis of all recurrent and highly amplified genes in high-grade serous and endometrioid ovarian cancer in a large panel of ovarian tumor cell lines. We further enhanced our dataset by integrating data from large clinical cohorts of primary tumors that have been assessed for copy number, gene expression, and patient outcome. By taking this integrative approach, we have streamlined the translation of high-resolution genomic data into preclinical in vitro studies, resulting in the identification of a number of genes that may be specifically targeted for the treatment of advanced ovarian tumors.

As proof-of-concept for our experimental and analysis approach, one of the most robust hits was URI1 prefoldin-like chaperone (URI1), which has recently been identified as an ovarian oncogene (10). URI1 encodes a molecular chaperone that is specifically associated with the mitochondrial inhibition of PPIy and PPIy-mediated inhibition of S6K1-BAD survival signaling. Using in vitro and in vivo analyses, Theurillat and colleagues showed that ovarian cancer cells that harbor amplification of chromosome 19q12 were "addicted" to the aberrant expression of URI1 (10), such that short hairpin RNA (shRNA) knockdown of URI1 in ovarian tumor cells overexpressing URI1 resulted in proliferative arrest and cell death. Interestingly, URI1 lies approximately 100 kb downstream of CCNE1, which has
previously been identified as the target gene for the 19q12 amplicon in high-grade serous and endometrioid ovarian tumors (9). CCNE1 was included in our boutique siRNA library, but did not induce a significant decrease in cell viability using our stringent criteria in cell lines harboring amplification at 19q12. This outcome is consistent compared with assays of anchorage-independent growth (9). Therefore, we provide evidence to support the conclusion that other genes within the 19q12 amplicon, including URI1, may have important functionality in ovarian tumorigenesis.

In addition to URI1, we identified 7 genes located in 4 amplicons that are strong candidates for therapeutic targeting of high-grade serous and endometrioid ovarian tumors including BMP8B, CACNA1C, Dyrk1B, GAB2, Pak4, Samd4B, and Zfp36. Copy number gain was correlated with aberrant expression in primary tumors in at least one independent cohort of primary tumors for all genes [TCGA, AOCS, and Ramakrishna and colleagues (17)]. Importantly, 4 of these genes were significantly correlated across all 4 datasets, namely, Dyrk1B, GAB2, Pak4, and Samd4B. Furthermore, when clinical outcome was taken into consideration, GAB2 and Pak4 emerged as strong candidates for therapeutic intervention due to altered overall and/or progression-free survival when copy number and gene expression were assessed using the TCGA and AOCS datasets. It should be noted that the TCGA dataset was a substantially larger cohort of primary tumors; therefore,

### Table 4. Survival analyses by Cox proportional hazards

<table>
<thead>
<tr>
<th>Gene</th>
<th>Locus</th>
<th>TCGA Agilent</th>
<th>AOCS U133</th>
<th>TCGA CN</th>
<th>AOCS CN</th>
<th>TCGA Agilent</th>
<th>AOCS U133</th>
<th>TCGA CN</th>
<th>AOCS CN</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI1</td>
<td>19q12</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>ns</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>GAB2</td>
<td>11q14.1</td>
<td>*</td>
<td>**</td>
<td>P = 0.09</td>
<td>**</td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Pak4</td>
<td>19q13.2</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>ns</td>
<td>P = 0.05</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Zfp36</td>
<td>19q13.2</td>
<td>*</td>
<td>ns</td>
<td>**</td>
<td>ns</td>
<td>*</td>
<td>ns</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Samd4B</td>
<td>19q13.2</td>
<td>NA</td>
<td>**</td>
<td>**</td>
<td>ns</td>
<td>NA</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Dyrk1B</td>
<td>19q13.2</td>
<td>NA</td>
<td>NA</td>
<td>**</td>
<td>ns</td>
<td>NA</td>
<td>NA</td>
<td>P = 0.09</td>
<td>**</td>
</tr>
<tr>
<td>Bmp8B</td>
<td>1p34.2</td>
<td>NA</td>
<td>NA</td>
<td>ns</td>
<td>ns</td>
<td>NA</td>
<td>NA</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Cacna1C</td>
<td>12p13</td>
<td>NA</td>
<td>NA</td>
<td>**</td>
<td>ns</td>
<td>NA</td>
<td>NA</td>
<td>ns</td>
<td>P = 0.07</td>
</tr>
</tbody>
</table>

NOTE: *, P < 0.05; **, P < 0.01; ***, P < 0.001.

Abbreviations: A, Agilent 244K Custom Gene Expression; CN, copy number; NA, median gene expression low; ns, not significant; U133, Affymetrix Gene Chip Human Genome U133 Plus 2.0 expression array.

Figure 2. GAB2 amplification correlates with increased gene expression and is associated with patient outcome. Correlation between GAB2 copy number and gene expression measured by microarray in (A) TCGA \( R^2 = 0.48, P < 0.05 \) and (B) AOCS \( R^2 = 0.63, P < 0.05 \) datasets. Kaplan-Meier analysis of overall survival of (C) TCGA and (D) AOCS datasets using GAB2 copy number as a categorical variable comparing loss/neutral copy number (CN log2 ratio <0.1) with gained (CN log2 ratio >0.5). Cox proportion hazards test with residual disease as a copredictor P values reported.
some of the associations observed in the TCGA dataset could not be replicated in the AOCS dataset possibly due to a lack of statistical power.

\textit{GAB2} is a scaffolding adapter protein that binds the p85 regulatory subunit of phosphoinositide-3-kinase (PI3K), stimulating PI3K activity. We previously reported that \textit{GAB2} was gained in 31\% (log2 ratio > 0.3); amplified in 10\% (log2 ratio > 0.6), and highly amplified in 5\% (log2 ratio > 0.8) of high-grade serous and endometrioid tumors (7), which is consistent with other studies that have reported a frequency of approximately 15\% amplification in ovarian tumors (24, 25). This gene is located on 11q14, a region also amplified in other tumor types such as breast (26). The amplicon is complex and distinct from the proximal \textit{CCND1} locus, which is seldom highly amplified in ovarian cancer, in contrast to breast cancer. A number of genes have been implicated as drivers for this amplicon, including \textit{EMSY}, \textit{C11orf30} and \textit{PAK1} (24) suggesting an amplified gene cassette. However, we found no effect of \textit{EMSY} knockdown, and while reduced \textit{PAK1} showed an effect in 2 cell lines, this was not amplification associated. Thus, the functional data presented here combined with the genomic data provide convincing evidence to support \textit{GAB2} as an important driver of ovarian cancer for this locus that may be targetable in the treatment of advanced tumors. Our clinical data suggest that increased \textit{GAB2} copy number results in a significant survival advantage in both the TCGA and AOCS datasets (P < 0.05, respectively) indicating that \textit{GAB2} amplification could drive a less aggressive cancer phenotype. However, this survival analyses is conducted relative to other cancers of the same type; therefore, the outcome remains poor even when \textit{GAB2} amplification is present.

Mechanistically, recent data suggest that \textit{GAB2} overexpression enhances epithelial-to-mesenchymal (EMT)-like behavior of ovarian tumor cells by enhancing properties such as migration and invasion via activation of the PI3K-Akt pathway and E-cadherin (27). This direct association between \textit{GAB2} overexpression phenotypes and activation of PI3K has led to the use of small-molecule inhibitors for PI3K and its downstream signaling molecule mTOR for the treatment of ovarian tumors that overexpress \textit{GAB2}. In preclinical studies, the use of such inhibitors significantly reduces the in vitro EMT-like behavior of \textit{GAB2} overexpressing ovarian tumors cells (27). Furthermore, increased sensitivity to a dual PI3K and mTOR inhibitor (PF-04691502) has been shown to correlate with increased expression of \textit{GAB2} mRNA in a panel of ovarian cancer cell lines (Sheppard and colleagues, unpublished data). These data suggest that PI3K/mTOR inhibitors may be used to treat ovarian tumor cells with amplification of 11q14 and consequent overexpression of \textit{GAB2} despite the inverse correlation we observed between \textit{GAB2} copy number and overall survival.

The 19q13.2 amplicon was notable for identifying 4 amplification-associated hits: \textit{PAK4}, \textit{Dyrk1b}, \textit{Zfp36}, and \textit{Samd4b}. This amplicon measures approximately 1 Mb but is gene-rich with 36 genes included in the screen. Our data suggests that multiple genes may be drivers in this amplicon and could be potential therapeutic targets. \textit{Dyrk1b} knockdown showed the most significant effect in association with amplified cell lines, and represents an attractive therapeutic target as a kinase; however, its expression was not strong by microarray analysis. In contrast, \textit{PAK4} expression was significantly correlated with copy number, and showed the most consistent association with clinical outcome of the 4 genes in the locus. \textit{PAK4} encodes a serine/threonine p21-activating kinase that is a major effector of Rho GTPases including Rac1 and Cdc42, which play an essential role in cytoskeleton organization, cell morphology, apoptosis, survival, and angiogenesis (28). Rho GTPases are generally considered to be difficult to directly target therapeutically and activating mutations in this family of proteins are rare (29). We observed \textit{PAK4} copy number gain (log2 ratio > 0.3) in 21\%, amplification in 8\% (log2 ratio > 0.6), and high amplitude amplification in 5\% (log2 ratio > 0.8) of high-grade serous and endometrioid ovarian tumors (7). Abrupt expression of \textit{PAK4} has been observed in a broad range of tumor types including lung (30), ovary (30), colon (30, 31), breast (30), and pancreas (32, 33). This has led to the suggestion that \textit{PAK4} could be an excellent molecule for targeted therapy and a number of small-molecule inhibitors targeting \textit{PAK4} have been developed (34, 35). In the context of ovarian cancer, overexpression of \textit{PAK4} has been reported at both the RNA and protein levels (36). In ovarian tumor cells, \textit{PAK4} has been shown to regulate cell migration and invasion in a kinase-dependent fashion through the activation of c-Src, MEK1, ERK1/2, and MMP2. Furthermore, \textit{PAK4} was also shown to regulate cell proliferation in a c-Src- and EGF-dependent manner (36). Despite the genetic and functional evidence supporting the role of \textit{PAK4} in cancer, little is known about the mechanism under which \textit{PAK4} operates in a biologic context. Specifically, there is little in the literature regarding \textit{PAK4} that gives a clear indication of molecules that operate downstream. This is a particular problem for the therapeutic targeting of \textit{PAK4} due to the fact that there is no \textit{bona fide} target or biomarker to accurately detect the on-target inhibition of \textit{PAK4} function. This issue must be addressed for \textit{PAK4} as a target to progress through to clinical development.

The genes included in our screen covered 25 distinct amplicons that occur frequently in ovarian cancer. We have identified the potential drivers (strictly defined in this study by an amplification-associated effect of gene knockdown) in 5 of these regions: 19q12 (\textit{Urir}), 11q14 (\textit{GAB2}), 12p13 (\textit{Caca1c}), 1p34 (\textit{Bmp8b}) and 19q13 (\textit{Dyrk1b}, \textit{Pak4}, \textit{Zfp36}, and \textit{Samd4b}). However, it is recognized that the cell lines used for this analysis are limited in the extent to which they recapitulate the types of amplicons that are observed in patients with high-grade serous and endometrioid ovarian cancer. For example, amplicons on 19p and 20p are less frequently observed in cancer cell lines than would be expected from the primary tumor data. In addition, it is plausible that many genes may only impart a phenotypic effect when silenced in conjunction with adjacent genes in the same region of amplification or with genes that lie on distal chromosomal locations. Functional
assessments of combinations of genes within single or multiple regions of amplification is outside the scope of this study. Therefore, it is likely that we may have failed to identify a number of driver genes within these tumor types.

Recent studies have indicated (37, 38) that while primary tumors show clear evidence of a monoclonal ancestry, there are often multiple subclones with varying genetic landscapes. Thus, the genetic and expression profiles obtained from a primary tumor depend on the part of the tumor taken for nucleic acid extraction. The amplicons studied here were chosen on the basis of frequency (indicating positive selection in multiple individual instances) and high level of amplification (suggesting that they represent the dominant clone in the tissue taken) and are therefore highly likely to represent genuine drivers and valid targets for therapy. Nonetheless, future applications of targeted therapies derived from such data will need to recognize that a single target is unlikely to be effective in all tumor cells, and multiple simultaneous strands of therapy will be required. This multiplicity of targets will require complex screening strategies and careful design of clinical trials as eligible patients will likely be limited.

These studies provide important targets for rational design of next generation therapeutics. Although the loss of viability data presented here is not sufficient to conclusively characterize driver genes in high-grade serous and endometrioid ovarian cancer, these data provide additional evidence to justify a more detailed biologic and molecular investigation into the target genes identified. Further investigation will be required to define drivers for other loci, with either insufficient cell lines with gain (20p13, 19p13.12, 19p13.12, and 19p13.12) or conversely with too many cell lines with gain (≥8 cell lines: 4 regions on 3q, 1q21, 8q24.21, and 2 regions on 20q). Both situations lead to a lack of power and in the latter case perhaps a lack of specificity in the identification of amplification-associated hits. Other regions may contain drivers activated by alternative mechanisms in addition to amplification, again reducing the capacity to detect them using an amplification-associated criterion. In addition, there are likely biologic reasons why some amplicons had no hits, including amplicons where the driver was selected for a cancer phenotype not measured in this assay such as anchorage-independent growth (e.g. CCNE1), cell motility, invasion, or other interactions with surrounding tissues, angiogenesis, and immune evasion. Nonetheless, this study represents a strong basis for further analysis of amplified genes in alternative assays and model systems.

In summary, we have described the first systematic functional analysis of all amplified genes in high-grade serous and endometrioid ovarian tumors. The integration of high-throughput siRNA loss-of-function screening with comprehensive copy number and expression genomic analyses and clinical outcome has resulted in the identification of a number of convincing targets for therapeutic intervention.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors' Contributions
Conception and design: S.J. Davis, I.G. Campbell, K.L. Gorringe, K.J. Simpson
Development of methodology: S.J. Davis, I.G. Campbell, K.J. Simpson
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.J. Davis, K.E. Sheppard, K.J. Simpson
Analysis and interpretation of data (e.g., statistical analysis, bioinformatics, computational analysis): S.J. Davis, R.B. Pearson, K.L. Gorringe, K.J. Simpson
Writing, review, and/or revision of the manuscript: S.J. Davis, K.E. Sheppard, R.B. Pearson, I.G. Campbell, K.J. Gorringe, K.J. Simpson
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.J. Davis, I.G. Campbell, K.J. Simpson
Study supervision: I.G. Campbell, K.L. Gorringe, K.J. Simpson

Acknowledgments

The authors thank Daniel Thomas and Yanny Handoko (Victorian Centre for Functional Genomics, Peter MacCallum Cancer Centre) for their technical assistance and screening support, Joshy George (Peter MacCallum Cancer Centre) for bioinformatics advice, and Jason Ellul (Peter MacCallum Cancer Centre) for bioinformatics support.

Grant Support

This study was funded by an Australian National Health and Medical Research Council project grant (#566603) and the Emer Casey Foundation. S.J. Davis is supported by a Cancer Council of Victoria Postgraduate Scholarship. The Victorian Centre for Functional Genomics (VCfG) is funded by the Australian Cancer Research Foundation (ACRF), the Victorian Department of Industry, Innovation and Regional Development (DIIRD), the Australian Phenomics Network supported by funding from the Australian Government’s Education Investment Fund through the Super Science Initiative, the Australasian Genomics Technologies Association (AMATA), and the Brockhoff Foundation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 5, 2012; revised January 13, 2013; accepted January 16, 2013; published OnlineFirst January 29, 2013.

References

Clinical Cancer Research

Functional Analysis of Genes in Regions Commonly Amplified in High-Grade Serous and Endometrioid Ovarian Cancer


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-12-3433

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2013/01/29/1078-0432.CCR-12-3433.DC1

Cited articles
This article cites 37 articles, 10 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/19/6/1411.full.html#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
/content/19/6/1411.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.