Comparative oncogenomics identifies PSMB4 and SHMT2 as potential cancer driver genes.

Genee Y. Lee¹, Peter M. Haverty³, Li Li³, Noelyn M. Kljavin², Richard Bourgon³, James Lee¹, Howard Stern⁴, Zora Modrusan², Somasekar Seshagiri², Zemin Zhang³, David Davis¹, David Stokoe¹, Jeffrey Settleman¹, Frederic J. de Sauvage⁵, Richard M. Neve¹*

¹Department of Discovery Oncology
²Department of Molecular Biology
³Department of Bioinformatics
⁴Department of Pathology
⁵Department of Molecular Oncology

Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA.

*Corresponding author: email: neve.richard@gene.com

Running title.

Oncogenomic target discovery.

Keywords.

Genomics, SHMT2, PSMB4, copy number, oncogene.

Conflict of Interest: The authors disclose no potential conflicts of interest.

Final word count: 4,699.

Total number of figures: 5.

Total number of tables: 2.
Abstract.

Cancer genomes maintain a complex array of somatic alterations required for maintenance and progression of the disease, posing a challenge to identify driver genes amongst this genetic disorder. Toward this end, we mapped regions of recurrent amplification in a large collection (n=392) of primary human cancers and selected 620 genes whose expression is elevated in tumors. An RNAi loss-of-function screen targeting these genes across a panel of 32 cancer cell lines identified potential driver genes. Subsequent functional assays identified SHMT2, a key enzyme in the serine/glycine synthesis pathway, as necessary for tumor cell survival but insufficient for transformation. The 26S proteasomal subunit, PSMB4, was identified as the first proteasomal subunit with oncogenic properties promoting cancer cell survival and tumor growth in vivo. Elevated expression of SHMT2 and PSMB4 was found to be associated with poor prognosis in human cancer supporting the development of molecular therapies targeting these genes or components of their pathways.
Introduction.

Genome-wide documentation of tumor gene copy number, gene expression and somatic mutations has become standard practice, driven by remarkable technological advances over the last decade (1, 2). By elucidating the “omic” landscape of tumor cells we are only now realizing the vast complexity of the diseases that are classed as cancer (3). Assigning function to any gene in this morass of observed changes can be a challenge, and years of collective study have begun to hone in on core functionalities regulating specific hallmarks of cancer (4). However, if the long term goal of diagnostic-driven, individualized therapy is to be realized, we need to begin to understand how these genetic changes sustain tumor growth, progression and survival.

Oncogenes and tumor suppressors are the classical drivers of oncogenesis, but it is still an open question as to which genes are the drivers in any single tumor when as much as 10-15% of the genome can be altered by recurrent copy number abnormalities (5). The task of finding “driver genes” is further hampered by the fact that the majority of copy number abnormalities are low-level gains, often encompassing large regions of the genome, which contrasts with high-level, focal copy number gains as exemplified by HER2 (6). Evidence also suggests individual amplicons contain multiple genes that collectively alter processes important for tumor growth and survival (7) and that there may be equivalent co-operativity between genes in separate but co-amplified genomic regions (8, 9). The distinct challenge is to separate passenger genes from driver genes (10). In the quest to find new driver genes and hence potential therapeutic targets, a number of groups have used either a reductionist approach by identifying a region of copy number gain/loss of clinical interest and testing the functionality of each gene within that region.
(11, 12), or a holistic approach using RNA interference library screens in a single cell line to find
driver genes (13, 14).

The primary goal of this study was to develop a genome-wide oncogenomic screening strategy to
identify candidate oncogenic driver genes. Our approach was to identify recurrent regions of
copy number gain and increased expression across multiple solid tumor types, with the
hypothesis that these genes represent fundamental processes for sustaining solid-tumor growth
and survival. We selected 620 genes from commonly amplified regions whose expression is
elevated in tumor versus normal and determined the effect on cell viability by RNAi targeting in
32 cancer cell lines. Here we report the findings of this study which represents one of the most
comprehensive functional oncogenomic screens to date.
Materials and Methods.

Cell lines

See Supplemental Table 7 for full list of cell lines, cell line STR profiles, growth conditions and transfection conditions used. All cell lines are tested for mycoplasma, cross contamination and genetically fingerprinted when new stocks are generated to ensure quality and confirm ancestry.

Cell line fingerprinting: SNP fingerprinting. SNP genotypes are performed each time new stocks are expanded for cryopreservation. Cell line identity is verified by high-throughput SNP genotyping using Fluidigm multiplexed assays. SNPs were selected based on minor allele frequency and presence on commercial genotyping platforms. SNP profiles are compared to SNP calls from available internal and external data (when available) to determine or confirm ancestry. In cases where data is unavailable or cell line ancestry is questionable, DNA or cell lines are re-purchased to perform profiling to confirm cell line ancestry. SNPs. rs11746396, rs16928965, rs2172614, rs10050093, rs10828176, rs16888998, rs16999576, rs1912640, rs2355988, rs3125842, rs10018359, rs10410468, rs10834627, rs11083145, rs11100847, rs11638893, rs12537, rs1956898, rs2069492, rs10740186, rs12486048, rs13032222, rs1635191, rs17174920, rs2590442, rs2714679, rs2928432, rs2999156, rs10461909, rs11180435, rs1784232, rs3783412, rs10885378, rs1726254, rs2391691, rs3739422, rs10108245, rs1425916, rs1325922, rs1709795, rs1934395, rs2280916, rs2563263, rs10755578, rs1529192, rs2927899, rs2848745, rs10977980.

Short Tandem Repeat (STR) Profiling. STR profiles are determined for each line using the Promega PowerPlex 16 System. This is performed once and compared to external STR profiles of cell lines (when available) to determine cell line ancestry. Loci analyzed. Detection of sixteen loci (fifteen STR loci and Amelogenin for gender identification), including D3S1358, TH01,
D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, AMEL, vWA, D8S1179 and TPOX.

**Cloning**

Full length ORF clones were obtained from Invitrogen and Genentech. C-terminus FLAG tags were added by PCR then tagged proteins were subcloned into pLPCX retroviral expression vector (Clontech). NIH/3T3 and MCF 10A stable cell lines were generated by infection with retroviral particles expressing FLAG-tagged constructs and selection with puromycin. See **Supplemental Table 7** for a full list of clones and primers used.

**Gene selection for RNAi screen.**

An independent set of tumor expression data (Source: GeneLogic, Inc., Gaithersburg, MD(15)) was used to identify genes upregulated in tumor versus normal samples (breast: 24 normals, 91 tumors; lung: 80 normals, 105 tumors; ovary: 101 normals, 82 tumors; prostate: 27 normals, 70 tumors). Percentile analysis for differential gene expression (PADGE)(16), a statistical tool that compares both the magnitude and the variability between two sample groups, was used to identify differential gene expression between normal and cancer samples in the Genelogic(15) database. Genes found within an amplicon which showed a 1.5-fold increase in cancer samples versus normal were selected for the screen.

**Tumor samples, DNA preparation and copy number arrays**

Tumor samples had a tumor percentage >80% as assessed by an expert pathologist. DNA was extracted from frozen tissues and cell lines by a standard protocol using DNA/RNA extraction
kit (Qiagen). DNA was then hybridized to the Agilent Human Genome CGH Microarray Kit 244A and genomic copy number was calculated by comparing the intensity ratio between individual tumors and the averaged normal DNA content.

**RNAi screen**

Cells were reverse lipid transfected with pools of four siRNAs; siGENOME siRNA were used in the primary screen and ON-TARGETplus siRNA was used in the secondary screen and all follow-up studies (Dharmacon). siRNA pools were used at 50 nM and singles at 25 nM. Pooled siRNA targeting PLK1 and TOX transfection reagent were used as positive controls at 25 nM. Lipids used were Dharmafects 1, 2, 3, and 4 (Dharmacon) and Lipofectamine RNAiMAX (Invitrogen). See **Supplemental Table 7** for transfection conditions for each cell line. siRNA-lipid complexes were formed for 1h, then cells were seeded on top of complexes. After 5 days, cell viability was assayed using CellTiter-Glo (Promega) and luminescence readout was performed using Envision plate reader (PerkinElmer). Gene selection criteria: Genes were selected from the screen by one of three criteria: 1) correlation of RNAi response with total RNA expression (Expression, value >0.3 Pearson’s); 2) correlation of RNAi response with copy number (CGH, value >0.3 Pearson’s); 3) genes which impacted cell viability across the panel of cell lines regardless of copy number or gene expression (Multiple). Genes were classified as a multiple hit if six or more cell lines had an RNAi response of greater than 0.65 (65% cell death).

**Gene expression analysis**

Patient tissue samples with appropriate Institutional Review Board approval and patient-informed consent were obtained from commercial sources (**Supplemental Table S1**). The
human tissue samples used in the study were de-identified (double-coded) before their use and hence the study using these samples is not considered human subject research under the US Department of Human and Health Services regulations and related guidance (45 CFR Part 46). All tumour tissues were subject to pathology review. Tumour DNA and RNA were extracted using the Qiagen AllPrep DNA/RNA kit (Qiagen).

Total RNA was harvested from cells using RNeasy Kit with on-column DNase digestion (Qiagen). RNA for validation of siRNA function was acquired 3 days after transfection unless otherwise noted. RNA was quantified using a Nanodrop spectrophotometer, amplified with TaqMan One-Step RT-PCR Master Mix, assayed with TaqMan gene expression assays, then analyzed using an 7900HT Fast Real-Time PCR System (Applied Biosystems). See Supplemental Table 7 for full list of gene expression assays used. Normalization control assays were HPRT1 (human) and ACTB (mouse).

**Immunoblotting**

Protein was harvested from cells with RIPA buffer, passed through a syringe, and cleared by centrifugation. Protein for validation of siRNA function was acquired 3 days after transfection unless otherwise noted. Protein was quantified using BCA protein assay (Pierce). Protein was separated on 4-12% Bis-Tris gels (Invitrogen), transferred to nitrocellulose membranes, blocked with 5% BSA or milk in TBST for 30 minutes, then blotted with primary antibody overnight at 4C. See Supplemental Table 7 for list of primary antibodies used. Membranes were then washed and incubated with appropriate HRP conjugated secondary antibodies for 1 hour, washed and detected with SuperSignal West Femto Chemiluminescent Substrate (Pierce). Luminescence signal was acquired with FluorChem Q (Alpha Innotech).
3D assays

Matrigel Drip Cultures: MCF10A cells were cultured in 3D culture as previously described(17). 125,000 cells were seeded in a 6 well plate on top of a layer of polymerized Matrigel (BD), overlaid with a solution of 5% Matrigel and cultured for 21 days.

Soft agar assay: NIH/3T3 cells were cultured in soft agar colony formation assay with a base layer of 1.2% agarose and an assay layer of 0.4% agarose (BD). 4000 cells were seeded per well in a 12 well plate. Plates were scanned and analyzed after 21 days using Gelcount scanner and software (Oxford Optronics).

Xenograft studies

1 million NIH/3T3 cells were implanted subcutaneously in the right flank of NCr nude mice (Taconic), with 5 mice per group. Tumor dimensions were measured by caliper. At end of assay, tumors were harvested and flash frozen in liquid nitrogen. RNA was extracted from tumors using a TissueLyser II and RNeasy extraction (Qiagen).

Cell cycle analysis

For steady state assay, NIH/3T3 cells were cultured for 48h and labeled with 10 µM BrdU for 30 minutes. Cells were then detached and stained using BrdU Flow Kit (BD). Flow cytometry data was acquired with FACSCalibur (BD) and analyzed with FlowJo software. For cell synchronization assay (double thymidine block), NIH/3T3 cells were cultured for 48h then blocked with 2 mM thymidine (Sigma) for 15h. Cells were released into full media for 10h and reblocked with 2 mM thymidine for 15h. Cells were then released and pulsed at indicated
timepoints for 30 minutes with 10 µM BrdU in the presence or absence of 100 ng/ml nocodazole (Sigma). Cells were then analyzed for cell cycle status as described above.

**Gene Ontology analysis**

Gene sets were classified by Gene Ontology using the DAVID pathway analysis online tool and standard parameters (18).

**Cell based assays**

Apoptosis was assayed using Caspase-Glo 3/7 Assay (Promega). Proteasome activity was assayed using Proteasome-Glo Chymotrypsin-Like Cell Based Assays (Promega).

**Kaplan-Meier Analysis.**

Data were generated with KMPlotter (http://kmplot.com/analysis/) using a median cut-off for gene expression. Probesets used for analysis: 202244_at (PSMB4); 214095_at (SHMT2).
Results

DNA copy number analysis of human tumors.

Anonymous tumor specimens were collected and appraised for tumor content and pathology by expert pathologists. The collection consisted of 161 breast tumors (HER2-positive (n=53), hormone receptor (HR)-positive (n=54) and triple-negative (TN, n=54)), 51 ovarian (serous (n=37), papillary serous (n=14)) tumors, 52 lung tumors (adenocarcinoma (n=5), squamous cell carcinoma (n=47)), 51 melanomas and 57 prostate tumors (Supplemental Table S1). DNA copy number was measured using Agilent arrays (see materials and methods). GLAD (R package from Bioconductor) (19) was applied to infer the segmented copy numbers for each sample normalized to two copies. Figure 1a compares the frequency of recurrent abnormalities across all five indications. The genomic identification of significant targets in cancer (GISTIC) algorithm (20) was applied to calculate significant regions of copy number change and identify the genes within each region. Table 1 contains the summary of this analysis; gene lists are supplied in Table S2.

Common regions of gene amplification and deletion across tumor types.

Our data provided a direct comparison amongst the genomes of five solid tumor types analyzed on a single platform and it was evident that the frequency of genomic aberrations found in breast, ovarian, lung and melanoma tumors had strikingly similar architecture in several chromosomal regions (Figure 1). In contrast, the prostate dataset shared few common features with the other tumor types, showing greater frequency of copy number loss overall and few regions of significant copy number gain. In order to select genes for an RNAi-based screen to identify
oncogenic drivers, we focused on genomic regions which exhibited recurrent amplification across the breast, ovarian, lung and melanoma datasets. Figure 1b shows an overlay of the GISTIC Q-scores for these four tumor types and Figure 1c shows peaks derived from a cross-analysis of these significant peaks. Genes within these 86 peaks (Table S2) were filtered by gene expression levels (see Materials and Methods) to select genes within amplicons which are over-expressed in cancer compared to normal tissues resulting in a final list of 620 genes (Table S3). Gene Ontology analysis (18) of these genes showed enrichment for several cancer-related functions such as regulation of DNA, protein and glucose metabolism (Table S4).

Loss of function RNAi screen to identify oncogenic driver genes.

To identify amplified and over-expressed genes required for initiation and/or maintenance of tumor cell proliferation, we performed an RNAi loss of function screen targeting 620 genes across a panel of 32 tumor cell lines representing the four solid tumor types used to generate the gene list as outlined in Figure 2a. Each of the 86 amplified regions identified by GISTIC was represented by at least three cell lines. Two rounds of RNAi screening were conducted, the results of which are summarized in Figure 2b and Table S3. Three independent criteria were used to select genes; (i) correlation of RNAi response with copy number (CGH; Figure 2c), (ii) correlation of RNAi response with total RNA level (expression; Figure 2d) and (iii) genes which affect multiple lines (multiple; Figure 2e) independent of copy number or expression correlations (Table S3). Using these selection criteria, 105 genes passed moved to second round screening of which twenty-five validated (Table 2). This list was significantly enriched for genes with functions associated with the proteasome, spliceosome, DNA replication, cell cycle and metabolism (Table S4). ERBB2 and MYC were among the genes meeting the validation criteria, confirming the screen was able to identify amplified/over-expressed driver oncogenes, although
these were both selected based on a correlation between phenotype and total RNA levels, not phenotype and copy number. Overall correlation of phenotype with expression levels yielded more hits than that with copy number, perhaps due to the noted lack of linearity between copy number/mRNA levels/protein levels for the majority of genes (21, 22).

**Experimental and functional validation of RNAi screen candidates**

Fifteen of the twenty-five hits from the RNAi screen were selected for further validation based on function and/or potential druggability (Table 2, Materials and Methods). Protein levels of these genes were assessed in the cell line panel by western analysis (Figure S1). Of the eight candidate genes for which specific antibodies were found, only three, ACTN4, ERBB2 and EMILIN1, showed correlation between mRNA (detected by Affymetrix expression arrays or TaqMan) and protein (Table S5). RNAi-mediated knockdown was confirmed for thirteen of the fifteen target genes. Levels of two genes, ENO3 and WARS, were unaffected by RNAi and dropped from further evaluation (Figure S2). To assess the transforming activity of the remaining thirteen genes, each was cloned and expressed in NIH/3T3 cells to perform soft agar colony formation assays (Figure 3a). Of these, ELOVL1 could not be expressed, PSMA6 expression was only detected by TaqMan, and all other genes expressed at or above physiological levels. Two of the thirteen genes (PSMB4 and SHMT2) promoted colony formation in addition to the HRAS.G12V positive control (Figure 3b). Overall proliferation rates of PSMB4 and SHMT2 expressing cells were higher than parental cells (Figure S3a). A more detailed analysis of cell cycle effects of these genes in both steady-state and after release from synchronization revealed that the expression of PSMB4, but not SHMT2, resulted in aberrations in cell cycle profile. In the steady-state condition, the cell cycle profile of SHMT2 expressing cells was indistinguishable from that of control cells, while PSMB4 expressing cells...
spent a significantly lower proportion of time in G(0)/G(1) phase in favor of more time in both S and G(2)/M phases (Figure S3b). Synchronization of the cells using a double thymidine block and release into media with or without nocodozole revealed that PSMB4-expressing cells have a similar cell cycle profile as HRAS.G12V-expressing cells, with accelerated transition through the G(2)/M phase and bypass of the G(2) DNA damage and mitotic spindle checkpoints (23) (Figure S3c, d). PSMB4 and SHMT2 were expressed in MCF 10A cells to assess effects on epithelial cell morphogenesis in three-dimensional cultures. These cells exhibited no morphological differences compared to control (Figure 3c), in contrast to the highly invasive phenotype of the HRAS.G12V control suggesting these genes have minimal effect on polarity, migration, and invasion. Knockdown efficiency for the individual RNAis within pools were confirmed by TaqMan and western for PSMB4 and SHMT2 (Figure S4a). In vivo assessment of oncogenic properties of PSMB4 and SHMT2 demonstrated that PSMB4 but not SHMT2 expression was sufficient to promote NIH/3T3 xenograft tumor growth (Figure 3d). Expression of each transgene in the tumors was confirmed using human-specific TaqMan probes on mRNA extracted from the tumors (Figure S4b).

**Relevance of SHMT2 and PSMB4 in human cancer.**

SHMT2 and PSMB4 were primarily selected for our screen as amplified in lung and ovarian cancer respectively. To investigate the broader clinical relevance, we compared expression of *SHMT2* and *PSMB4* in cancer and normal tissue in two published datasets (15, 24). For the five primary indications surveyed in this study, *SHMT2* expression is significantly increased in cancer samples compared to normal tissue in all five indications in both datasets (Figure 4a and b, left panels). Normal samples were not available for ovary and skin in the TCGA RNAseq dataset (24), however Agilent expression array data from the same group contained three normal
ovary samples and showed an increase in expression in cancer (p-value = 0.0226, Table S6). In comparison, the cytosolic form of serine hydroxymethyltransferase, SHMT1, was found overexpressed in only four tumor types, and showed decreased expression in several others. PSMB4 showed significant up-regulation in breast, lung, ovarian and skin tumors (Figure 4a and b, right panels). Overexpression of the three proteasome beta- ring catalytic subunits was also observed (Table S6). PSMB5 overexpression followed a similar pattern to PSMB4, whereas levels of PSMB6 and 7 were unrelated and less frequently elevated in cancer. Increases in expression of SHMT2 and PSMB4 were observed in a wide range of other tumor types supporting the involvement of these genes in the etiology of multiple cancers (Table S6). Kaplan-Meier analysis (25) indicated high expression of SHMT2 is associated with worse relapse- free survival (RFS) and distant metastasis-free survival (DMFS) and overall survival (OS) in breast cancer, and increased time to first progression (FP) and decreased OS in lung cancer (Figure 4c and d, left panels, Table S6). Increased PSMB4 expression was associated with worse RFS in breast cancer and decreased OS in ovarian cancer (Figure 4c, right panel, Table S6).

Suppression of PSMB4 inhibits proteasomal activity and processing of β-ring catalytic subunits

Inhibition of proteasome activity is an established anti-cancer therapeutic strategy (26, 27). While PSMB4 does not contain intrinsic hydrolytic activity, it is rate-limiting for 20S proteasome assembly (28). Cell viability and protease activity was assessed after PSMB4 knockdown revealing a time-dependent decrease in both cell number and protease activity (Figure 5a, upper panels). Cells treated with the proteasome inhibitor, MG132, showed a similar
effect in the cells tested (Figure 5a, lower panels). Expression of PSMB5, -6 and -7 mRNA increased over time in response to PSMB4 knockdown and MG132 treatment (Figure 5b) and an accumulation of polyubiquinated (K48-linked) protein products were observed (Figure 5c). This did not result in an obvious increase of the total protein levels of PSMB5, -6 and -7, but a decrease in the processed (lower molecular weight, higher mobility) forms with a concomitant accumulation of the precursor forms of these subunits (Figure 5c) was observed. This data indicates loss of PSMB4 disrupts the formation of the 20S proteasome, since assembly of 20S half-mers requires PSMB4 and propeptide removal precedes maturation of the 20S proteasome (29). This effect appears to be specific to loss of PSMB4, since knockdown of PSMA6 had little or no effect on the protein levels of the core beta- ring catalytic subunits (Figure S5a). To compare the effects of PSMB4 loss to proteasome inhibitor activity, we selected 5 cell lines with a range of sensitivities to bortezomib (PS-341, Velcade) and compared the effect on cell viability of bortezomib against PSMB4 knockdown. While one line (NCI-H1838) was exceptionally resistant to both Bortezomib treatment and loss of PSMB4, other lines tested showed variable sensitivities to both treatments with little correlation between the two results (Figure 5d). A similar result was observed with twenty of the lines used in the RNAi screen (Figure S5c).

Together these data support a central role for PSMB4 in maintaining a functional 20S proteasome and suggest that targeting PSMB4 may offer an alternative therapeutic option to existing proteasome inhibitors.

Discussion.

In this report we used genomics-based selection and a loss-of-function screen to identify genes required for the survival and growth of human cancer cells. Two rounds of RNAi screens
reduced our initial pool of 620 candidates to 25 genes, fifteen of which were put through a stringent series of follow-up screens to assess oncogenic potential. Known oncogenes ERBB2 and MYC were identified using this approach, confirming the validity of this screen. A number of shared biological functions were enriched in the 25 candidate genes; SHMT2 and ALDOA have glycolytic/metabolic functions; PHF5A, THOC4 and SNRPD2 are involved in mRNA processing; and 40% (6/15) of the genes are components of the 26S proteasome (four from the 19S regulatory particle and one each from the 20S α- and β-rings). We used stringent criteria to identify potential targets, and in addition to SHMT2 and PSMB4, our data revealed a number of genes which cancer cells from multiple tissues rely on for continued proliferation and warrant further investigation (Table 2).

Based on our results, we established that the mitochondrial serine hydroxymethyltransferase gene (SHMT2) is required for cancer cell survival, but is not sufficient to promote tumorigenesis. SHMT enzymes catalyze the conversion of serine to glycine by catalyzing the transfer of the β-carbon of serine to tetrahydrofolate (THF) generating 5,10-methylene-THF and glycine (30, 31). Phosphoglycerate dehydrogenase (PHGDH), another enzyme in the pathway of generating glycine from glucose, was recently shown to be amplified in melanoma (32) and breast cancer (33) and the next enzyme in this pathway, phosphoserine aminotransferase (PSAT), is overexpressed in colorectal cancers, associated with chemoresistance (34) and required for serine pathway flux in breast cancer (33). More recently a delicate interdependency between glucose and amino acid metabolism has been elucidated, indicating serine is an allosteric regulator of pyruvate kinase M2 (35, 36). Our data show cell lines with high SHMT2 require the gene for survival, but exogenous expression of the gene does not alter the acinar development of MCF
10A cells grown in Matrigel (in contrast to PHGDH(32)), while it is sufficient to promote anchorage independent growth in colony formation assays but not to drive growth of tumors in vivo. Furthermore, SHMT2 is found overexpressed in many tumors and is associated with worse clinical outcome in breast and lung cancer. Therefore, our findings contribute to an evolving paradigm that serine/glycine metabolism is of critical importance to the development and maintenance of cancer.

Our studies also provided the first evidence for the requirement of a proteasomal subunit, PSMB4, as a potential driver oncogene in multiple tumor types. PSMB4 and five other proteasome subunits were found to be essential for the survival of a broad range of tumor types (breast, lung, skin and ovary (Table S3)). PSMB4 has also been identified as a gene required for the survival of human glioblastoma cells (37) although there is little evidence for elevated levels in brain cancer versus normal brain in the datasets we studied. Of the two proteasome subunits that underwent functional screening, PSMB4 was the only one which promoted both anchorage-independent growth and tumorigenesis and, to the authors’ knowledge, is the first proteasomal subunit shown to possess oncogenic properties.

The two FDA-approved proteasome inhibitors, bortezomib and carfilzomib, both target the core proteolytic subunits PSMB5, PSMB6 and PSMB7(38). The non-catalytic subunit PSMB4 represents a novel potential target as it plays an important role in regulating the assembly of the proteasome (28, 29), and thus by inhibiting its function and proteasome assembly one could potentially prevent the catalytic activity of all three proteolytic subunits. There is precedence that such an approach is feasible, as CRBN has been shown to directly interact with PSMB4, negatively regulating proteasome activity in vivo, possibly through interfering with the assembly of the 20S proteasome (39). Since proteasomal assembly requires the 15-amino acid C-terminal
tail of PSMB4 to intercalate into a groove between the PSMB6 and PSMB7 subunits, interfering with this interaction offers a potential therapeutic opportunity (40). Despite PSMB4 being amplified and over-expressed in cancer (a pre-requisite for being screened in this project), tumor cell sensitivity to loss of PSMB4 did not correlate with gene copy number or gene expression. Evidence suggests that regulation of proteins in the proteasomal complex appears to occur at the protein level, rather than at the copy number / mRNA level (21). Paradoxically, *in vitro* and patient-derived data suggests that elevated expression of certain proteasome subunits may contribute to bortezomib resistance (41, 42). While it is plausible that increased copy number and expression of *PSMB4* is required in certain tumors to maintain *PSMB4* transcript and/or protein levels, the exact mechanism of *PSMB4* deregulation remains to be evaluated and it is likely a complex mechanism tied to the intrinsic turnover rate of the proteasome.

Overall our approach has led to the discovery of potential therapeutic targets, but has also underscored the complexity of cancer biology. *HER2* is the paradigm for a gene driven by amplification of its locus generating high RNA and protein expression resulting in oncogenic activity. While our screen was predicated on this hypothesis, very few of the targets were selected on the basis of copy number, highlighting the biological complexity evident in a cancer cell. A number of studies have indicated that there is ever decreasing linearity in the relationship of gene copy number to RNA to protein (43) due to a number of regulatory mechanisms such as methylation, RNA editing and stability, and protein translation and ubiquitination. Amplicons may also contain more than one potential driver gene which co-operate in tumor etiology and maintenance (44, 45), while many genes within any given amplicon are likely to be passengers and do not contribute to the survival/homeostasis for a given cancer cell. Therefore, expanding the strategy used in this study by selecting genes based on multiple biologic parameters (copy
number, mutation, RNA, epigenetics and protein expression) may improve our ability to identify 
*bona fide* cancer targets. As technologies advance to generate accurate high density datasets, the 
use of high-content screening approaches to assess multiple hallmarks of cancer in relevant 
cancer models will greatly improve our ability to identify relevant targets (46, 47).

**Acknowledgments**

We thank Mamie Yu and Suresh Selvaraj for maintaining and providing cell lines for this study. We 
are grateful to Don Kirkpatrick for technical assistance and insightful discussions.
References

Tables

Table 1.

<table>
<thead>
<tr>
<th>Tumor Type</th>
<th>Gain # Peaks</th>
<th>Gain Genes</th>
<th>Gain OIC Genes</th>
<th>Loss # Peaks</th>
<th>Loss Genes</th>
<th>Loss UIC Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast (HER2)</td>
<td>27</td>
<td>250</td>
<td>69</td>
<td>25</td>
<td>327</td>
<td>193</td>
</tr>
<tr>
<td>Breast (TN)</td>
<td>22</td>
<td>372</td>
<td>103</td>
<td>21</td>
<td>375</td>
<td>237</td>
</tr>
<tr>
<td>Breast (HR)</td>
<td>20</td>
<td>308</td>
<td>85</td>
<td>18</td>
<td>180</td>
<td>131</td>
</tr>
<tr>
<td>Ovarian</td>
<td>26</td>
<td>486</td>
<td>130</td>
<td>25</td>
<td>330</td>
<td>233</td>
</tr>
<tr>
<td>Melanoma</td>
<td>36</td>
<td>1019</td>
<td>197</td>
<td>32</td>
<td>119</td>
<td>65</td>
</tr>
<tr>
<td>NSCLC</td>
<td>26</td>
<td>636</td>
<td>145</td>
<td>24</td>
<td>274</td>
<td>187</td>
</tr>
<tr>
<td><strong>Totals:</strong></td>
<td><strong>3071</strong></td>
<td><strong>729</strong></td>
<td><strong>1605</strong></td>
<td><strong>1046</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr</th>
<th>Cytoband</th>
<th>Description</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELOVL1</td>
<td>1</td>
<td>1p34.2</td>
<td>CGI-88 PROTEIN</td>
<td>CGH</td>
</tr>
<tr>
<td>C3ORF52</td>
<td>3</td>
<td>3p21.31</td>
<td>CHROMOSOME 3 OPEN READING FRAME 62</td>
<td>CGH</td>
</tr>
<tr>
<td>EN03</td>
<td>17</td>
<td>17pter-p11</td>
<td>ENOLASE 1, (ALPHA)</td>
<td>CGH</td>
</tr>
<tr>
<td>ERBB2</td>
<td>17</td>
<td>17q11.2-q12</td>
<td>ERYTHROBLASTIC LEUKEMIA VIRAL ONCOGENE HOMOLOG 2</td>
<td>EXPRESSION</td>
</tr>
<tr>
<td>ACTN4</td>
<td>19</td>
<td>19q13</td>
<td>ACTININ, ALPHA 4</td>
<td>EXPRESSION</td>
</tr>
<tr>
<td>MYC</td>
<td>8</td>
<td>8q24.21</td>
<td>V-MYC MYELOCYTOMATOSIS VIRAL ONCOGENE HOMOLOG (AVIAN)</td>
<td>EXPRESSION</td>
</tr>
<tr>
<td>EMLIN1</td>
<td>2</td>
<td>2p23.3-p23.2</td>
<td>ELASTIN MICROFIBRIL INTERFACE 1</td>
<td>EXPRESSION</td>
</tr>
<tr>
<td>WARS</td>
<td>14</td>
<td>14q32.31</td>
<td>INTERFERON INDUCED PROTEIN 53</td>
<td>EXPRESSION</td>
</tr>
<tr>
<td>EIF2S2</td>
<td>2</td>
<td>20pter-q12</td>
<td>EUKARYOTIC TRANSLATION INITIATION FACTOR 2, SUBUNIT 2 BETA, 38KD</td>
<td>EXPRESSION</td>
</tr>
<tr>
<td>BOP1</td>
<td>8</td>
<td>8q24.3</td>
<td>BLOCK OF PROLIFERATION 1</td>
<td>EXPRESSION</td>
</tr>
<tr>
<td>SHMT2</td>
<td>12</td>
<td>12q22-q14</td>
<td>SERINE HYDROXYMETHYLTRANSFERASE 2 (MITOCHONDRIAL)</td>
<td>EXPRESSION</td>
</tr>
<tr>
<td>EIF6</td>
<td>20</td>
<td>20q12</td>
<td>EUKARYOTIC TRANSLATION INITIATION FACTOR 6</td>
<td>EXPRESSION</td>
</tr>
<tr>
<td>EDEM2</td>
<td>20</td>
<td>20q11.22</td>
<td>CHROMOSOME 20 OPEN READING FRAME 31</td>
<td>EXPRESSION</td>
</tr>
<tr>
<td>ALDOA</td>
<td>16</td>
<td>16p11.2</td>
<td>ALDOLASE A, FRUCTOSE-BISPHOSPHATE</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>EIF3S1</td>
<td>16</td>
<td>16p11.2</td>
<td>EUKARYOTIC TRANSLATION INITIATION FACTOR 3, SUBUNIT 8, 110KDA</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>EIF4A3</td>
<td>17</td>
<td>17q25.3</td>
<td>DEAD (ASP,GLU,ALA, ASP) BOX POLYPEPTIDE 48</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>PHF5A</td>
<td>22</td>
<td>1q32.3</td>
<td>PHD FINGER PROTEIN 5A</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>PSMA6</td>
<td>14</td>
<td>14q13</td>
<td>PROTEASOME (PROSOME, MACROPAIN) SUBUNIT, ALPHA TYPE, 6</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>PSMB4</td>
<td>1</td>
<td>1q21</td>
<td>PROTEASOME (PROSOME, MACROPAIN) SUBUNIT, BETA TYPE, 4</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>PSMD1</td>
<td>17</td>
<td>17q11.2</td>
<td>PROTEASOME (PROSOME, MACROPAIN) 26S SUBUNIT, NON-ATPASE, 4</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>PSMD2</td>
<td>3</td>
<td>3q27.1</td>
<td>PROTEASOME (PROSOME, MACROPAIN) 26S SUBUNIT, NON-ATPASE, 1</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>PSMD3</td>
<td>17</td>
<td>17q12-q11.1</td>
<td>PROTEASOME (PROSOME, MACROPAIN) 26S SUBUNIT, NON-ATPASE, 3</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>PSMD8</td>
<td>19</td>
<td>19q13.2</td>
<td>PROTEASOME (PROSOME, MACROPAIN) 26S SUBUNIT, NON-ATPASE, 8</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>SRRF2D</td>
<td>15</td>
<td>19q13.2</td>
<td>SMALL NUCLEAR RIBONUCLEOPROTEIN D2 POLYPEPTIDE 16.9KDA</td>
<td>MULTIPLE</td>
</tr>
<tr>
<td>THOC4</td>
<td>17</td>
<td>17q25.3</td>
<td>THO COMPLEX 4</td>
<td>MULTIPLE</td>
</tr>
</tbody>
</table>
Tables and Figure Legends.

Table 1. Summary of GISTIC analysis of copy number aberrations for each tumor type. Number of significant peaks (as determined by GISTIC), total gene and miRNA count within those regions are listed for both copy number gain and loss. Genes identified as over-expressed in cancer (OIC) and under-expressed in cancer (UIC) compared to normal controls are also listed. HER2: HER2-positive, TN: Triple-negative, HR: Hormone receptor-positive.

Table 2. Genes passing selection criteria after two rounds of RNAi screens. Analysis indicates the criteria by which the genes were selected; correlation of phenotype with copy number (CGH), correlation of phenotype with mRNA level (Expression) and effect on multiple lines (multiple). Genes highlighted in bold were selected for follow-up functional studies. Cell line copy number and gene mRNA baseline levels are shown in Table S5.

Figure 1. Recurrent gene copy number changes in breast, ovarian, lung, melanoma and prostate primary tumors. (A) Frequencies of gene copy gain (green) and loss (red) are plotted as a function of genomic location for each tumor types. Positive values indicate frequencies of samples showing copy number increases [Log2(copy number) > 0.3], and negative values indicate frequencies of samples showing copy number decreases [Log2(copy number) < 20.3]. Vertical solid lines indicate boundaries between chromosomes. Vertical dotted lines indicate positions of centromeres. Sample number is indicated to the right of each graph. (B) Analysis of significant regions of copy number gain by GISTIC. Q-scores for gene copy number gains are
overlaid on a single plot for breast, ovarian, lung and melanoma samples. (C) Averaged Q-values across all four tumor types identify most significant regions for the entire dataset.

**Figure 2.** RNAi screen design and results. (A) Schematic outline of the screening approach used in this study. (B) Summary of all data for primary RNAi screen (upper panel) and secondary RNAi screen (lower panel) are shown. Details of genes and cell lines can be found in Tables S3 and S7 respectively. For each graph individual targeted genes and controls are plotted on the horizontal axis, and cell growth represented as percent control plotted on the vertical axis. There are 32 data points for each gene or control representing a data point for each cell line in the screen. In the lower panel, genes are ordered by the criteria by which the genes were selected for the secondary screen (indicated above the data points). Examples of how genes were selected for further analysis are shown for individual genes; (C) correlation with copy number: C3ORF62 (Pearson’s correlation = 0.53), (D) correlation with expression: ERBB2 (Pearson’s correlation = 0.52) and (F) multiple hit: EIF4A3 (genes were classified as a multiple hit if six or more cell lines had an RNAi response of greater than 0.65 (65% cell death) indicated by dotted line) respectively. See Materials and Methods for detailed selection criteria.

**Figure 3.** Effect of *PSMB4* and *SHMT2* on anchorage independent growth and cell cycle.

A. Western blot of FLAG-tagged target genes stably expressed in NIH/3T3 cells. Dotted line denotes lanes placed next to each other from different regions of the same gel. B. Soft agar assays to assess effect of target genes on anchorage independent growth. Total number of colonies and average colony diameter are shown in bar graphs below. Scale bar; 1 mm. C. Effect of target gene expression on MCF 10A acinar morphogenesis. Selected target genes were
expressed in the non-tumorigenic breast epithelial cell line to assay gross acinar morphogenesis. 
Oncogenic HRAS.G12V is shown as a positive control at a 4 day timepoint. Scale bar, 100 µm. 
**(D)** Effect of PSMB4 and SHMT2 on tumor growth in vivo. Nude mice bearing NIH/3T3 cell 
lines stably expressing PSMB4, SHMT2 and HRAS.G12V (positive control) were monitored for 
xenograft growth. The mean tumor volume (+ SEM) for each group (n = 5) is shown.

**Figure 4.** Expression of *SHMT2* and *PSMB4* in normal and cancer tissues. (A) Expression of 
*SHMT2* is shown for breast, lung, ovary, skin and prostate tissues from Genelogic (left panel) 
and TCGA (right panel) data sources. (B) Expression of *PSMB4* is shown for breast, lung, ovary, 
skin and prostate tissues from Genelogic (left panel) and TCGA (right panel) data sources. N = 
normal; C = Cancer; numbers above normal/cancer pairings represent p-values of a non-
parametric t-Test (Mann-Whitney test). (C) Kaplan-Meier survival plots showing the prognostic 
effect on Relapse-Free Survival (RFS) in breast cancer for *SHMT2* (left panel) and *PSMB4* (right 
panel). (D) Kaplan-Meier survival plots showing the prognostic effect on Overall Survival (OS) 
in lung cancer for *SHMT2* (left panel) and *PSMB4* (right panel). Data were generated with 
KMPlotter(25) using a median cut-off for gene expression.

**Figure 5.** Effects of *PSMB4* loss on proteasome activity and protein expression of the 
proteasome catalytic β subunits. **A.** siRNA knockdown of *PSMB4* in NCI-H1299 and ES-2 cells 
over a 72h time course measuring cell viability (upper left) and chymotrypsin-like proteasome 
activity normalized to cell number (upper right). Effects of the proteasome inhibitor, MG132, on 
cell viability and chymotrypsin-like proteasome activity are shown in lower panels. **B.** 
Measurement of RNA levels of the proteasome β-ring subunits *PSMB4*, *PSMB5*, *PSMB6*, and
PSMB7 in cells treated with PSMB4 RNAi for the stated times or MG132 for 24h. C. Protein levels (western blots) of the proteasome β-ring subunits PSMB4, PSMB5, PSMB6, and PSMB7 in cell lines transfected with PSMB4 RNAi for the stated times or MG132 for 24h. D. Comparison of cell line sensitivity to bortezomib and PSMB4 RNAi in a panel of cell lines. RNAi knockdown of PSMB4 was confirmed in Figure S5b.
Figure 1.

A

Breast

Ovarian

Lung

Melanoma

Prostate

Distance along chromosome

B

- Breast
- Ovary
- Lung
- Melanoma

Distance along chromosome

C

Q-Score

Distance along chromosome
Figure 2.

A

Breast, Melanoma, Lung, Ovary, Prostate
Genes
Amplified Regions (GISTIC)
overexpressed in cancer vs normal

620 Genes
siGenome RNAi Screen
32 cell lines

Expression Correlation Copy number Multiple Hits

Secondary
105 Genes
OTP RNAi Screen
32 cell lines

Expression Correlation Copy number Multiple Hits

B

620 Genes

Copy number Expression Multiple

C

C3ORF62

Copy Number

Percent Inhibition

D

ERBB2

mRNA Expression

Percent Inhibition

E

EIF4A3

Percent Inhibition

cell line
Figure 3.
Figure 4.

A

B

C

HR = 1.53 (1.35 - 1.73)
log rank P = 3.9e-11

low SHMT2

high SHMT2

D

HR = 1.4 (1.2-1.03)
log rank P = 1.4e-05

low SHMT2

high SHMT2

HR = 1.56 (1.37 - 1.77)
log rank P = 5.4e-12

low PSMB4

high PSMB4

HR = 0.93 (0.8-1.08)
log rank P = 0.33

low PSMB4

high PSMB4
Figure 5.
Comparative oncogenomics identifies PSMB4 and SHMT2 as potential cancer driver genes


Cancer Res Published OnlineFirst April 22, 2014.