

# Functional genomics reveal that the serine synthesis pathway is essential in breast cancer

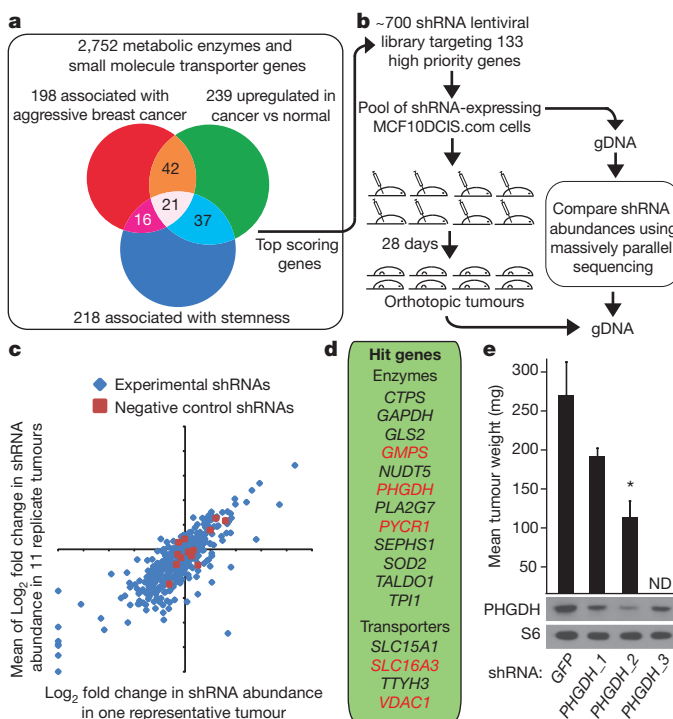
Richard Possemato<sup>1,2,3,4</sup>, Kevin M. Marks<sup>5</sup>, Yoav D. Shaul<sup>1,2,3,4</sup>, Michael E. Pacold<sup>1,2,3,4,6</sup>, Dohoon Kim<sup>1,2,3,4</sup>, Kıvanç Birsoy<sup>1,2,3,4</sup>, Shalini Sethumadhavan<sup>5</sup>, Hin-Koon Woo<sup>5</sup>, Hyun G. Jang<sup>5</sup>, Abhishek K. Jha<sup>5</sup>, Walter W. Chen<sup>1,2,3,4</sup>, Francesca G. Barrett<sup>1</sup>, Nicolas Stransky<sup>3</sup>, Zhi-Yang Tsun<sup>1,2,3,4</sup>, Glenn S. Cowley<sup>3</sup>, Jordi Barretina<sup>3,7</sup>, Nada Y. Kalaany<sup>1,2,3,4</sup>, Peggy P. Hsu<sup>1,2,3,4</sup>, Kathleen Ottina<sup>1,2,3,4</sup>, Albert M. Chan<sup>1,2,3,4</sup>, Bingbing Yuan<sup>1</sup>, Levi A. Garraway<sup>3,7</sup>, David E. Root<sup>3</sup>, Mari Mino-Kenudson<sup>8</sup>, Elena F. Brachtel<sup>8</sup>, Edward M. Driggers<sup>5</sup> & David M. Sabatini<sup>1,2,3,4</sup>

Cancer cells adapt their metabolic processes to drive macromolecular biosynthesis for rapid cell growth and proliferation<sup>1,2</sup>. RNA interference (RNAi)-based loss-of-function screening has proven powerful for the identification of new and interesting cancer targets, and recent studies have used this technology *in vivo* to identify novel tumour suppressor genes<sup>3</sup>. Here we developed a method for identifying novel cancer targets via negative-selection RNAi screening using a human breast cancer xenograft model at an orthotopic site in the mouse. Using this method, we screened a set of metabolic genes associated with aggressive breast cancer and stemness to identify those required for *in vivo* tumorigenesis. Among the genes identified, phosphoglycerate dehydrogenase (PHGDH) is in a genomic region of recurrent copy number gain in breast cancer and PHGDH protein levels are elevated in 70% of oestrogen receptor (ER)-negative breast cancers. PHGDH catalyses the first step in the serine biosynthesis pathway, and breast cancer cells with high PHGDH expression have increased serine synthesis flux. Suppression of PHGDH in cell lines with elevated PHGDH expression, but not in those without, causes a strong decrease in cell proliferation and a reduction in serine synthesis. We find that PHGDH suppression does not affect intracellular serine levels, but causes a drop in the levels of  $\alpha$ -ketoglutarate, another output of the pathway and a tricarboxylic acid (TCA) cycle intermediate. In cells with high PHGDH expression, the serine synthesis pathway contributes approximately 50% of the total anaplerotic flux of glutamine into the TCA cycle. These results reveal that certain breast cancers are dependent on increased serine pathway flux caused by PHGDH overexpression and demonstrate the utility of *in vivo* negative-selection RNAi screens for finding potential anticancer targets.

As a starting point for identifying metabolic genes required for tumorigenesis, we cross-referenced maps of metabolic pathways with the KEGG database to compile a comprehensive list of 2,752 genes encoding all known human metabolic enzymes and transporters (Supplementary Table 1). Public oncogenomic data were analysed to score genes based on three properties: (1) higher expression in tumours versus normal tissues; (2) high expression in aggressive breast cancer; or (3) association with the stem-cell state (Fig. 1a). Genes scoring in two of these three categories as well as those at the top of each category were selected to define a high-priority set of 133 metabolic enzyme and transporter genes (Supplementary Table 2). We assembled lentiviral short hairpin RNA (shRNA) vectors targeting these genes (median 5 shRNAs per gene) and used them to generate two libraries of shRNA-expressing lentiviruses, one containing 235 distinct shRNAs (targeting

transporters and control genes) and the other 516 distinct shRNAs (targeting metabolic enzymes and control genes)<sup>4</sup>.

To identify genes that may be essential for tumorigenesis, the libraries were screened for shRNAs that became depleted during breast tumour formation in mice. Human MCF10DCIS.com cells<sup>5</sup> were chosen for the screens because, of several breast cancer lines examined, these were capable of forming tumours upon injection of the fewest number of cells. One and a half million MCF10DCIS.com cells were infected with each library so that each cell carried one viral integrant, and



**Figure 1 | Outline of *in vivo* pooled screening strategy identifying PHGDH as essential for tumorigenesis.** **a**, Venn Diagram outlining meta-analysis. **b**, Outline of experimental design. gDNA, genomic DNA. **c**, Log<sub>2</sub> fold change in shRNAs abundance of experimental (blue) or neutral shRNAs (red) for a single tumour (*x*-axis) compared to an average of eleven tumours (*y*-axis). **d**, Genes scoring *in vivo*. **e**, Average weight of tumours from MCF10DCIS.com cells expressing shRNAs targeting PHGDH (*PHGDH\_1*, *PHGDH\_2* and *PHGDH\_3*) or control (*GFP*) and protein expression of PHGDH or RPS6 (S6). Error bars are s.e.m. (*n* = 4). \**P* value < 0.05. ND, not done.

<sup>1</sup>Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>2</sup>Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Broad Institute of Harvard and MIT, Seven Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>4</sup>The David H. Koch Institute for Integrative Cancer Research at MIT, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. <sup>5</sup>Agios Pharmaceuticals, 38 Sidney Street, Cambridge, Massachusetts 02139, USA. <sup>6</sup>Harvard Radiation Oncology Program, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02114, USA. <sup>7</sup>Department of Medical Oncology and Center for Cancer Genome Discovery, Dana-Farber Cancer Institute and Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. <sup>8</sup>Department of Pathology, Massachusetts General Hospital and Harvard Medical School, 55 Fruit Street, Boston, Massachusetts 02114, USA.

~500–1,000 cells per shRNA (100,000–1,000,000 cells total) were injected into mouse mammary fat pads at two sites per animal (Supplementary Discussion). Twenty-eight days later orthotopic tumours were harvested and massively parallel DNA sequencing was used to determine the abundance of each shRNA in genomic DNA from tumours and initially injected cells (Fig. 1b). shRNA abundances correlated well between replicate tumours (Fig. 1c) and 5 or 12 tumours per library were analysed to identify shRNAs that became significantly depleted during tumour formation. Sixteen genes were designated hits in the screen, with at least 75% of the shRNAs targeting these genes scoring (Fig. 1d and Supplementary Table 3).

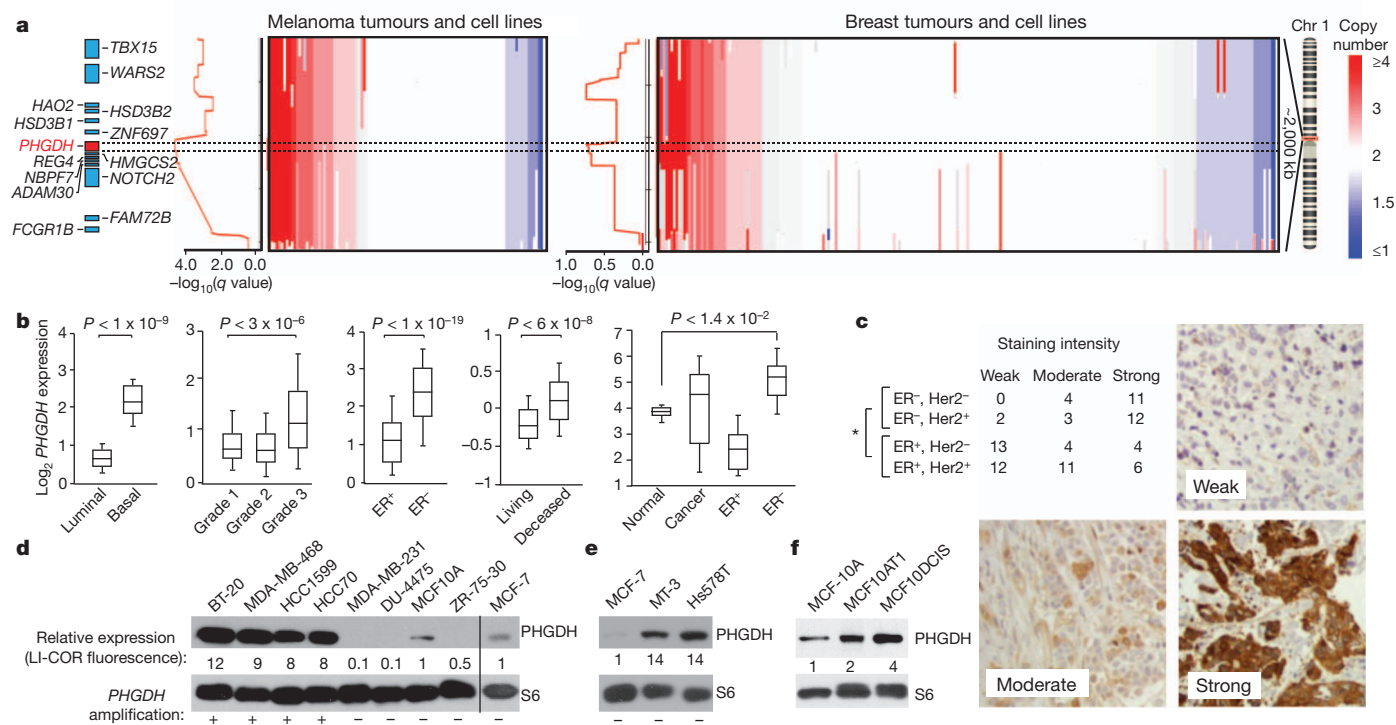
Several genes previously shown to have important roles in cancer emerged as hits, including the mitochondrial ATP transporter *VDAC1*; the lactic acid transporter *SLC16A3*; and the nucleotide synthesis genes *GMPS* and *CTPS*. The hit list also includes genes involved in the control of oxidative stress (*SOD2*, *GLS2*, *SEPHS1*), the pentose phosphate pathway (*TALDO1*), glycolysis (*GAPDH*, *TPI1*), and in the proline (*PYCR1*) and serine (*PHGDH*) biosynthetic pathways. An analogous pooled screen carried out in MCF10DCIS.com cells grown in culture rather than in tumour xenografts revealed that of 20 genes that scored in the *in vitro* screen, 10 also scored in the *in vivo* screen (Supplementary Fig. 2a, Supplementary Table 3 and Supplementary Discussion). Interestingly, *AK2*, which encodes an adenylate kinase that generates ADP from ATP and AMP, was required for *in vitro* but not *in vivo* growth (Supplementary Fig. 2b).

For five hit genes (*PHGDH*, *GMPS*, *SLC16A3*, *PYCR1* and *VDAC1*), two scoring shRNAs were tested for their effects on tumour formation. Each of these shRNAs suppressed expression of their targets in MCF10DCIS.com cells and reduced tumour-forming capacity.

(Fig. 1e and Supplementary Fig. 2c). For reasons discussed later, *PHGDH* was of particular interest. The three shRNAs that scored in the *in vivo* screen also decreased *PHGDH* protein expression, and two shRNAs of differing knockdown efficacies inhibited tumour growth, consistent with their capacity to suppress *PHGDH* expression (Fig. 1e). Moreover, tumours derived from cells that in culture had confirmed reductions in *PHGDH* levels had, in immunohistochemical (Supplementary Fig. 3a) and immunoblotting assays (Supplementary Fig. 3b), *PHGDH* staining or levels similar to control tumours, suggesting that tumorigenesis selected for cells that lost shRNA-mediated *PHGDH* suppression.

To prioritize genes for follow-up studies we consulted a recently available analysis of copy number alterations across cancer genomes<sup>6</sup>. This analysis showed that *PHGDH* exists in a region of chromosome 1p commonly amplified in breast cancer and melanoma (Fig. 2a), as well as in several other cancer types (not shown). In total, 18% of patient-derived breast cancer cell lines and 6% of primary tumours have amplifications in *PHGDH*. In the data sets examined, none of the other hit genes are in genomic regions of focal and recurrent copy number gain.

Our meta-analysis for genes associated with aggressive breast cancer is corroborated by a previous study that found elevated *PHGDH* messenger RNA levels in breast cancers that are ER negative, of the basal type, and associated with poor 5-year survival<sup>7</sup>. We confirmed these associations in distinct gene expression data sets (Fig. 2b) and additionally found that *PHGDH* is elevated in ER-negative breast cancer relative to normal breast tissue (Fig. 2b). Of all the genes identified as hits in our screen, *PHGDH* has the most significantly elevated expression in ER-negative breast cancer (Supplementary Fig. 4). Moreover, by analysing 82 human breast tumour samples with an



**Figure 2 | Genomic amplifications of *PHGDH* in cancer and association of *PHGDH* expression with aggressive breast cancer markers.** **a**, *PHGDH* vicinity copy number data for melanoma (left,  $n = 111$ ) and breast cancer (right,  $n = 243$ ) samples. Coloured bar indicates degree of copy number loss (blue) or gain (red). Samples sorted by copy number at *PHGDH* locus (dotted lines). Graphs at left of copy number data show amplification significance ( $-\log_{10}(q$  value),  $\sim 0.60$  is the significance threshold for amplification). **b**, Representative *PHGDH* gene expression data for indicated breast cancer groups. Whiskers indicate 91st and 9th percentile. **c**, Table reports numbers of

human breast cancer samples with 'weak', 'moderate', or 'strong' *PHGDH* staining from breast cancer subgroups indicated. Representative staining intensities shown in images. Magnification,  $\times 20$ . \* $P < 0.0001$  comparing ER<sup>+</sup> versus ER<sup>-</sup> classes (Fisher's exact test). **d–f**, *PHGDH* protein levels are shown for *PHGDH* amplified versus non-amplified (annotated with + or -) (**d**), *PHGDH* non-amplified, over-expressing (**e**), and MCF-10A-derived cell lines (**f**). Values below *PHGDH* immunoblots are normalized immunofluorescent quantification (LI-COR) of *PHGDH* levels relative to actin control and MCF-10A and MCF-7.

immunohistochemical assay for PHGDH, we found that PHGDH protein levels correlate significantly with ER-negative status (Fig. 2c). In total, compared to ER-positive breast tumours, ~68% and ~70% of ER-negative breast tumours have elevations of PHGDH at the mRNA and protein levels, respectively (Fig. 2b, c and Supplementary Methods). ER-negative breast cancer comprises approximately 20–25% of all breast cancer cases, but as many as 50% of all breast cancer deaths within 5 years of diagnosis<sup>8</sup>, underscoring the importance of identifying additional drug targets for this class of breast cancer.

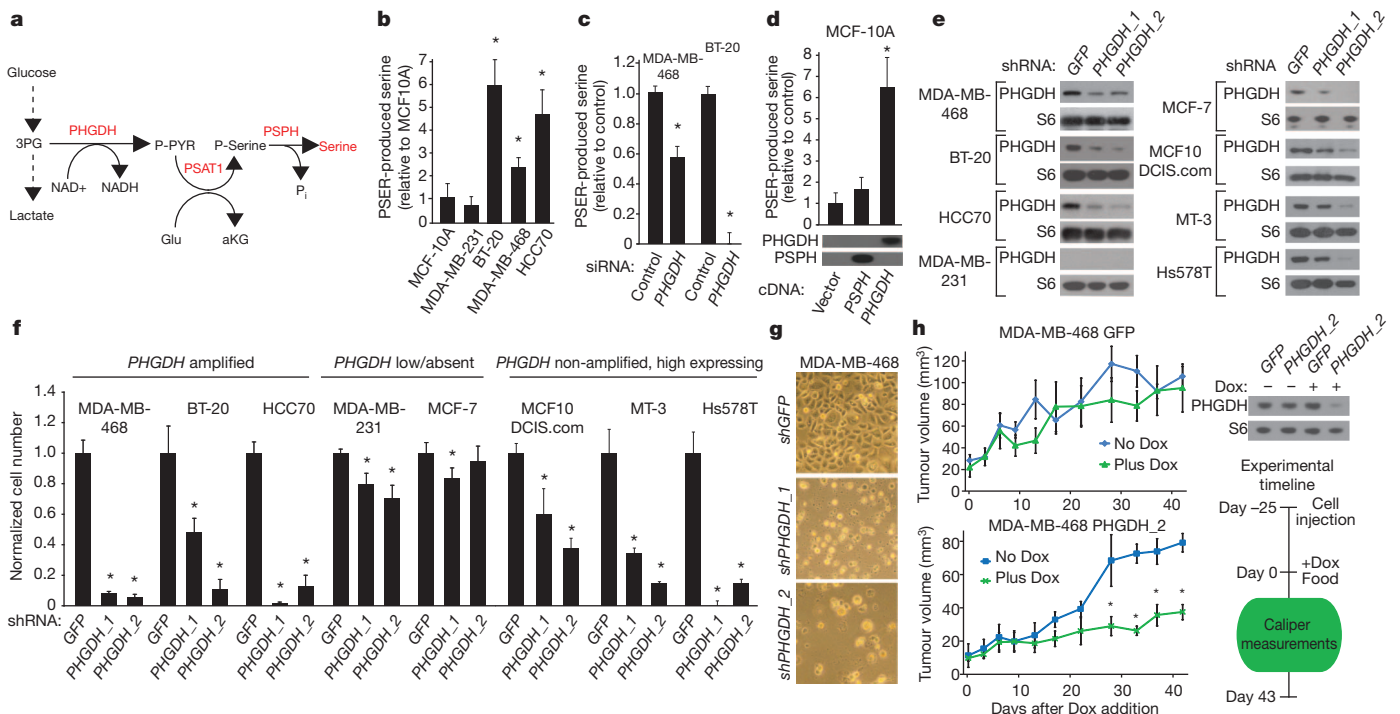
Across a set of breast cancer lines, four lines with *PHGDH* amplifications had 8–12-fold higher PHGDH protein expression compared to non-transformed MCF-10A and ER-positive MCF-7-cell lines, which do not have *PHGDH* amplifications (Fig. 2d). Mechanisms other than gene copy number increases must also exist for boosting PHGDH expression because PHGDH protein levels were also elevated in two ER-negative cell lines (MT-3, Hs578T) lacking the *PHGDH* amplification (Fig. 2e). This is consistent with the finding that PHGDH expression is upregulated at the mRNA and protein level in a higher fraction of ER-negative breast cancers than the fraction exhibiting amplification at the DNA level. Interestingly, PHGDH is also expressed fourfold more in the MCF10DCIS.com cells used in the *in vivo* screen than in two parental lines (MCF-10A and MCF10AT) that exhibit no or lower tumorigenicity<sup>9</sup> (Fig. 2f).

*PHGDH* encodes 3-phosphoglycerate dehydrogenase, the first enzyme branching from glycolysis in the three-step serine biosynthetic pathway<sup>10</sup> (Fig. 3a). PHGDH uses NAD as a cofactor to oxidize the glycolytic intermediate 3-phosphoglycerate into phosphohydroxypyruvate<sup>11,12</sup>, which subsequent enzymes in the pathway convert into serine via transamination (PSAT1) and phosphate ester hydrolysis (PSPH) reactions<sup>10</sup> (Fig. 3a). Serine is essential for synthesis of proteins and other biomolecules needed for cell proliferation, including nucleotides, phosphatidyl-serine and sphingosine

(Supplementary Fig. 1). Classic studies show elevated serine biosynthetic activity, as determined by enzyme assays, in rat tumour lysates<sup>10,13</sup>, and suggest that PSPH is the rate-limiting enzyme of this pathway in the liver<sup>14</sup>. Interestingly, we find that numerous genes that are expected to promote serine biosynthesis or are involved in the subsequent metabolism of serine for biosynthesis are elevated in ER-negative breast cancer (Supplementary Fig. 5), demonstrating that PHGDH elevation occurs in the context of upregulation of a broader pathway.

To understand the metabolic consequences of increased PHGDH expression we used metabolite profiling and serine synthesis pathway flux analysis to examine breast cancer cells with and without *PHGDH* amplifications. We found that cells with *PHGDH* amplifications (BT-20, MDA-MB-468 and HCC70), had increased flux through the serine synthesis pathway compared to those without *PHGDH* amplifications (MDA-MB-231, MCF7 and MCF10A) (Fig. 3b and Supplementary Fig. 6a). Cells with elevated PHGDH and high pathway flux were capable of robust proliferation in medium lacking serine, whereas in cells with low levels of PHGDH, the deprivation of serine caused a significant blunting or even cessation of proliferation (Supplementary Fig. 6b).

PHGDH is required for the increased serine pathway flux of cells with elevated PHGDH because RNAi-mediated PHGDH suppression significantly reduced flux in MDA-MB-468 and BT-20 cells (Fig. 3c). Conversely, in MCF-10A human mammary cells engineered to over-express PHGDH, serine pathway flux increased to levels similar to those in MDA-MB-468, BT-20 and HCC70 cells (Fig. 3d). Furthermore, MCF-10A cells overexpressing PHGDH had increased proliferation in the absence of serine, indicating that PHGDH overexpression is sufficient to drive flux through the pathway (Supplementary Fig. 6c). Interestingly, overexpression of PSPH, considered the rate-limiting serine biosynthetic enzyme in the liver, did not increase pathway flux in MCF-10A cells (Fig. 3d). The observation that PSPH is rate limiting



**Figure 3 | Cell lines with elevated PHGDH expression have increased serine biosynthetic pathway activity and are sensitive to PHGDH suppression.**

**a**, Serine biosynthesis (PSER) pathway. **b–d**, Serine production by serine biosynthesis pathway in indicated breast cell lines (**b**), after PHGDH suppression by siRNA (**c**), and MCF-10A cells expressing *PHGDH* or *PSPH* cDNAs with associated immunoblots (**d**). **e**, Immunoblots of indicated proteins for indicated cell lines expressing control shRNA (*GFP*) or shRNAs against *PHGDH* (*PHGDH\_1* and *PHGDH\_2*). **f**, Relative proliferation of cells transduced with

shRNA constructs after seven days. **g**, Images showing cellular morphology (magnification,  $\times 20$ ) of MDA-MB-468 at day seven of **f**. **h**, Tumour growth of MDA-MB-468 cells expressing doxycycline-inducible control shRNA (*GFP*) or shRNA against *PHGDH* (*shPHGDH\_2*) in mice fed doxycycline (Dox, 2 mg kg<sup>-1</sup>, green lines,  $n = 5$ ) or normal (blue lines,  $n = 4$ ) diet after initial tumour palpation (day 0). Immunoblots of PHGDH or RPS6 (S6) shown for cells *in vitro*. \* $P < 0.05$  relative to control. Error bars for metabolite measurements ( $n = 4$ ) and tumour size indicate s.e.m., and for cell number indicate s.d. ( $n = 3$ ).

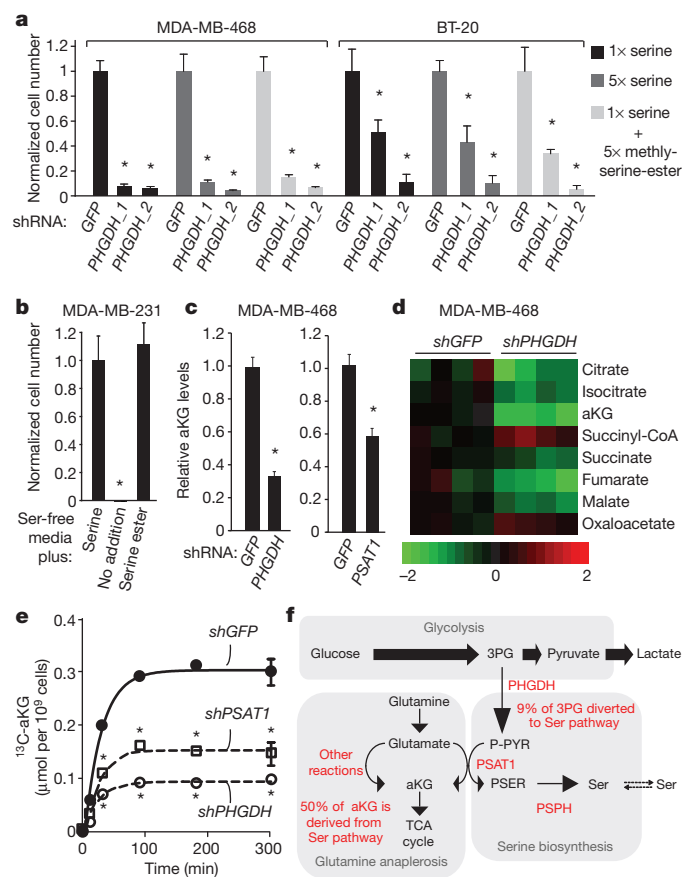
in the liver whereas PHGDH is rate limiting in MCF-10A cells can be reconciled by the observation that serine levels in the liver (2 mM) are well above the concentration at which PSPH is feedback-inhibited by serine (500  $\mu$ M), but low in cell lines in culture ( $\sim$ 100  $\mu$ M), a concentration at which PSPH should be active<sup>14</sup>. These data demonstrate that PHGDH is a key enzyme controlling flux through the serine biosynthetic pathway in cancer cells.

We next asked if cells with an increase in PHGDH expression require it for cell proliferation and survival. In cell lines with elevated PHGDH expression (BT-20, MDA-MB-468, HCC70, Hs578T and MT-3), but not without (MDA-MB-231 and MCF-7), RNAi-mediated suppression of PHGDH caused a marked decrease in cell number (Fig. 3e, f and Supplementary Fig. 6d) and cell death (Fig. 3g and Supplementary Fig. 6e) in the absence of apoptotic markers (Supplementary Fig. 6f). This sensitivity to PHGDH suppression was observed both in cells with *PHGDH* amplifications (BT-20, MDA-MB-468 and HCC70) and in those with high PHGDH expression but lacking *PHGDH* amplification (MT-3 and Hs578T). Consistent with flux through the serine synthesis pathway being important in cells with high PHGDH expression, suppression of the other two enzymes in the pathway (PSAT1 and PSPH) inhibited the proliferation of MDA-MB-468 and BT-20, but not MCF-7, cells (Supplementary Fig. 6g). Moreover, inhibition of PSPH inhibited tumour formation by MCF10DCIS.com cells (Supplementary Fig. 6h). Therefore, elevated PHGDH expression defines a set of breast cancer cell lines with increased serine pathway flux that are dependent upon PHGDH, PSAT1 and PSPH for proliferation. This finding suggests that many ER-negative breast cancers that express PHGDH at high levels ( $\sim$ 70% of all ER-negative disease in our data set; Fig. 2c) may be sensitive to inhibitors of the serine synthesis pathway.

To investigate whether PHGDH suppression can affect the growth of established tumours, we generated an inducible shRNA<sup>15</sup> that, upon doxycycline treatment, reduced PHGDH protein levels in MDA-MB-468 cells (Fig. 3h). MDA-MB-468 cells transduced with this shRNA were allowed to form murine mammary fat pad tumours for 25 days before introduction of doxycycline in a subset of mice (Fig. 3h). Compared to control mice, those given doxycycline exhibited substantially reduced tumour growth, whereas tumours made from cells transduced with a control inducible shRNA grew equally well in the presence or absence of doxycycline (Fig. 3h). These results indicate that PHGDH suppression can adversely affect growth in existing tumours (Supplementary Discussion).

Serine is a central metabolite for biosynthetic reactions, and we find that overexpression of PHGDH contributes significantly to biosynthetic flux to serine. However, PHGDH suppression inhibited proliferation even in cells growing in media containing normal levels of extracellular serine (Fig. 3f), and supplementation with additional serine or a cell-permeable methyl-serine-ester did not blunt the effects of PHGDH suppression (Fig. 4a, b). Intracellular and extracellular serine are in equilibrium (Supplementary Fig. 7a), and import of extracellular serine was not defective in the cell lines studied (Supplementary Fig. 7b). These findings suggest that serine production may not be the only important role of PHGDH in cell lines with high PHGDH expression. We considered three hypotheses to explain our observations: (1) serine produced via the PHGDH pathway is used in a different manner than exogenous serine; (2) suppression of PHGDH adversely affects glycolysis; or (3) the PHGDH, PSAT1 and PSPH reactions produce metabolites besides serine that are also critical for cell proliferation. The first hypothesis was deemed unlikely because serine synthesized intracellularly is in equilibrium with extracellular serine (Supplementary Fig. 7a). The second hypothesis was also unlikely because PHGDH suppression did not affect glucose uptake or lactate production (Supplementary Fig. 7c).

To pursue the third hypothesis, we considered which additional metabolites the serine synthesis pathway might produce in significant levels in cells with high PHGDH expression. The serine pathway



**Figure 4 | Suppression of PHGDH results in a deficiency in anaplerosis of glutamine to aKG.** **a**, Relative proliferation of cell lines indicated expressing control shRNA (*GFP*) or shRNAs against *PHGDH* (*PHGDH\_1* and *PHGDH\_2*) after seven days of growth under conditions indicated. **b**, Relative proliferation of MDA-MB-231 cells under conditions indicated. **c**, Intracellular aKG four days after treatment with shRNA against *PHGDH* or *PSAT1*; cell number normalized relative to control shRNA (*GFP*). **d**, TCA cycle intermediate levels four days after treatment with shRNA against *PHGDH* or *GFP* ( $n = 4$ ). Colour bar shows Log<sub>2</sub> scale. **e**, aKG isotopic labelling at indicated time points after treatment with isotopically labelled glutamine four days after treatment with shRNA against *PHGDH*, *PSAT1* or *GFP*. **f**, Model of relative metabolite fluxes for indicated pathways. \* $P < 0.05$  relative to control. Error bars indicate s.e.m. ( $n = 4$ ).

produces equimolar amounts of serine and  $\alpha$ -ketoglutarate (aKG; Supplementary Fig. 1). Proliferating cells use intermediates of the TCA cycle, such as aKG, as biosynthetic precursors, and upregulate anaplerotic reactions that drive glutamine-derived carbon into the TCA cycle, counterbalancing biosynthetic efflux<sup>16</sup> (Supplementary Discussion). We hypothesized that in cells with high PHGDH expression, the PSAT1 reaction might contribute a significant fraction of glutamate to aKG flux. If true, the serine biosynthesis pathway would have an important role in TCA anaplerosis of glutamine-derived carbon. Consistent with this possibility, suppression of PHGDH in MDA-MB-468 cells caused a large reduction in the levels of aKG (Fig. 4c and Supplementary Fig. 7d). In fact, of the major metabolites measured, aKG was the one with the most significant and largest change upon PHGDH suppression, whereas serine levels were not significantly changed (Supplementary Fig. 8). PHGDH suppression also caused a significant reduction in other TCA components (Fig. 4d and Supplementary Fig. 8). Like suppression of PHGDH, suppression of PSAT1 also caused a significant reduction in serine pathway flux and aKG levels (Fig. 4c and Supplementary Fig. 7d, e). Furthermore, labelling studies using U-<sup>13</sup>C-glutamine revealed that the absolute flux from glutamine to aKG and other TCA intermediates was significantly reduced in cells with RNAi-mediated suppression of PHGDH or

PSAT1 (Fig. 4e and Supplementary Fig. 9a, b). These data indicate that in cell lines with high PHGDH expression, the serine synthesis pathway is responsible for approximately 50% of the net conversion of glutamate to aKG and that suppression of PHGDH results in a significant loss of TCA intermediate flux and steady-state levels of TCA intermediates (Fig. 4f and Supplementary Fig. 9a, b). Furthermore, labelling studies using U-<sup>13</sup>C-glucose in cell lines with PHGDH amplification (MDA-MB-468) and without (MDA-MB-231) revealed that in cells with high PHGDH expression, flux through the serine biosynthesis pathway shunts 8–9% of the glycolytic flux towards serine production, compared to 1–2% in the cell line with low PHGDH expression (Fig. 4f and Supplementary Fig. 9a). Therefore, increased flux through the serine biosynthesis pathway has a major impact on aKG production, but a smaller effect on glycolysis or serine availability in these cells (Supplementary Discussion). In contrast, another prominent aKG-producing transaminase, alanine aminotransferase, does not contribute significantly to aKG production in PHGDH-amplified cells (Supplementary Fig. 10).

We find that PHGDH expression is a critical part of a cellular program promoting serine pathway flux (Supplementary Fig. 5) and is responsible for a considerable portion of anaplerosis of glutamate into the TCA cycle as aKG (Supplementary Fig. 1). As ~70% of ER-negative breast cancers exhibit elevated PHGDH (Fig. 2c), our work suggests that targeting the serine synthesis pathway may be therapeutically valuable in breast cancers with elevated PHGDH expression or PHGDH amplifications (Supplementary Discussion). Lastly, we anticipate that the screening approach described here may be applicable to other cancer types or gene sets, enabling the identification of novel cancer targets directly in an *in vivo* context.

## METHODS SUMMARY

To undertake negative-selection RNAi screening in solid tumours, pools of MCF10DCIS.com cells expressing an shRNA library were injected into the fourth mammary fat pad of immunocompromised mice and allowed to form tumours. Abundances of shRNAs in the tumours were determined using massively parallel sequencing and compared to shRNA abundance in the injected cells. Genes targeted by shRNAs that were significantly depleted during tumour growth were considered hits and prioritized by analysing gene copy number data from human tumours and cancer cell lines. Lentiviral shRNAs were used to suppress PHGDH expression in breast cancer cell lines with and without PHGDH genomic amplification. Serine synthesis pathway activity and anaplerosis were measured via flux analyses using isotopically labelled molecules.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 19 October 2010; accepted 8 July 2011.

Published online 14 July 2011.

1. Jones, R. G. & Thompson, C. B. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev.* **23**, 537–548 (2009).

2. Hsu, P. P. & Sabatini, D. M. Cancer cell metabolism: Warburg and beyond. *Cell* **134**, 703–707 (2008).
3. Bric, A. *et al.* Functional identification of tumor-suppressor genes through an *in vivo* RNA interference screen in a mouse lymphoma model. *Cancer Cell* **16**, 324–335 (2009).
4. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
5. Miller, F. R., Santner, S. J., Tait, L. & Dawson, P. J. MCF10DCIS.com xenograft model of human comedo ductal carcinoma *in situ*. *J. Natl. Cancer Inst.* **92**, 1185–1186 (2000).
6. Beroukhi, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
7. Pollari, S. *et al.* Enhanced serine production by bone metastatic breast cancer cells stimulates osteoclastogenesis. *Breast Cancer Res. Treat.* **125**, 421–430 (2011).
8. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
9. Miller, F. R. *et al.* Xenograft model of progressive human proliferative breast disease. *J. Natl. Cancer Inst.* **85**, 1725–1732 (1993).
10. Snell, K. Enzymes of serine metabolism in normal, developing and neoplastic rat tissues. *Adv. Enzyme Regul.* **22**, 325–400 (1984).
11. Achouri, Y., Rider, M. H., Schaftingen, E. V. & Robbi, M. Cloning, sequencing and expression of rat liver 3-phosphoglycerate dehydrogenase. *Biochem. J.* **323**, 365–370 (1997).
12. Walsh, D. A. & Sallach, H. J. Purification and properties of chicken liver D-3-phosphoglycerate dehydrogenase. *Biochemistry* **4**, 1076–1085 (1965).
13. Knox, W. E., Herzfeld, A. & Hudson, J. Phosphoserine phosphatase distribution in normal and neoplastic rat tissues. *Arch. Biochem. Biophys.* **132**, 397–403 (1969).
14. Lund, K., Merrill, D. K. & Guynn, R. W. The reactions of the phosphorylated pathway of L-serine biosynthesis: thermodynamic relationships in rabbit liver *in vivo*. *Arch. Biochem. Biophys.* **237**, 186–196 (1985).
15. Wiederschain, D. *et al.* Single-vector inducible lentiviral RNAi system for oncology target validation. *Cell Cycle* **8**, 498–504 (2009).
16. DeBerardinis, R. J. *et al.* Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc. Natl Acad. Sci. USA* **104**, 19345–19350 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of the D.M.S. laboratory, M. Kwon, B. Luo and F. Reinhardt for assistance. This research is supported by fellowships from S. G. Komen for the Cure to R.P. and the Life Science Research Foundation to Y.D.S. and grants from the Keck Foundation, the David H. Koch Institute for Integrative Cancer Research at MIT, The Alexander and Margaret Stewart Trust Fund, and NIH Grant CA103866 to D.M.S. D.M.S. is an investigator of the Howard Hughes Medical Institute.

**Author Contributions** R.P. and D.M.S. conceived the project and designed the experiments and K.M.M. and E.M.D. designed metabolic analysis and flux experiments. R.P. performed the screening, knockdown, cell proliferation and tumour formation experiments. K.B., Y.D.S., F.G.B., M.E.P., P.P.H., K.O., Z.-Y.T., N.Y.K. and W.W.C. assisted with experiments. E.F.B. collected breast cancer patient samples and scored PHGDH immunohistochemistry. D.K. performed and M.M.-K. assisted with interpretation of PHGDH immunohistochemistry. K.M.M., S.S., H.-K.W., A.K.J. and E.M.D. performed metabolite profiling, flux experiments and metabolite modelling. N.S., J.B. and L.A.G. assisted with amplification data. G.S.C. and D.E.R. assisted with screening technology. B.Y. and A.M.C. provided bioinformatic support. R.P. wrote and D.M.S. edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.M.S. ([sabatini@wi.mit.edu](mailto:sabatini@wi.mit.edu)).

## METHODS

**Materials.** Materials were obtained from the following sources: antibodies to PHGDH (HPA021241) and PSPH (HPA020376) from Sigma; an antibody against PYCRI (13108-1-AP) from Proteintech; an antibody against GMPS (A302-417A) from Bethyl Labs; an antibody against VDACL1 (ab16814) from Abcam; antibodies to RPS6 (2217), PARP (9532) and Caspase-3 (9662) from Cell Signaling Technologies; an antibody against PSAT1 (H00029968-A01) from Novus Biologicals; an antibody against SLC16A3 (AB3316P) from Millipore; and HRP-conjugated anti-mouse, anti-rabbit secondary antibodies from Santa-Cruz Biotechnology; lactate dehydrogenase from Roche (10127230001); lactic acid from Acros; RPMI-1640 media, 3-bromopyruvate and glycine/hydrazine solution (G5418) from Sigma;  $\alpha$ - $^{15}\text{N}$ -glutamine from Isotech/Sigma (486809); L-[3H(G)]-serine from Perkin Elmer; Infinity Glucose Oxidase Liquid Stable Reagent (TR15221) from Thermo Electron; U- $^{13}\text{C}$ -glutamine from Isotech/Sigma (605166); MT-3 cells from DSMZ; Hs578T, MDA-MB-468, MDA-MB-231, BT-20, HCC1599, HCC70, DU4475, MCF-7 and ZR-75-30 cells from ATCC; MCF-10A, MCF-10AT1 and MCF10DCIS.com cells from the Karmanos Cancer Center, Michigan; matrigel from BD Biosciences; Phusion DNA polymerase from New England Biolabs; BCA Protein Assay from Pierce; siRNAs from Dharmacon; and amino-acid-free, glucose-free RPMI-1640 from US Biological. Lentiviral shRNAs were obtained from the The RNAi Consortium (TRC) collection of the Broad Institute<sup>4</sup>. The TRC numbers for the shRNAs used are: *GFP*, TRCN0000072186; *PHGDH\_1*, TRCN0000221861; *PHGDH\_2*, TRCN0000221865; *PSPH\_1*, TRCN0000002796; *PSPH\_2*, TRCN0000315168; *PSAT1\_1*, TRCN0000035266; *PSAT1\_2*, TRCN0000035268; *SLC16A3\_1*, TRCN0000038477; *SLC16A3\_2*, TRCN0000038478; *VDACL1\_1*, TRCN000029126; *VDACL1\_2*, TRCN000029127; *GMPS\_1*, TRCN0000045938; *GMPS\_2*, TRCN0000045941; *PYCR1\_1*, TRCN0000038979; *PYCR1\_2*, TRCN0000038980. The TRC website is <http://www.broadinstitute.org/rnai/trc/lib>. The doxycycline-inducible shRNA vector used was previously described<sup>15</sup>.

**Cell culture.** MDA-MB-468, MDA-MB-231, BT-20, HCC1599, HCC70, DU4475, ZR-75-30, MT-3, Hs578T and MCF-7 were cultured in RPMI supplemented with 10% IFS and penicillin/streptomycin. MCF-10A and MCF10AT1 cells were cultured as described previously<sup>17</sup>. MCF10DCIS.com cells were cultured in 50:50 DMEM and F12 media with 5% horse serum and penicillin/streptomycin.

**Compilation of metabolic gene list.** A list of all human metabolic enzymes and small molecule transporters was generated by cross-referencing maps of metabolic pathways (Roche) with the KEGG database (<http://www.genome.jp/kegg/kegg1.html>). NCBI resources including Entrez Gene (<http://www.ncbi.nlm.nih.gov/entrez>) and the available literature were used to identify known or putative gene function and to identify functional homologues. A protein was considered a metabolic enzyme if it modified a small molecule to generate another small molecule. Proteins which modified polymerized DNA or RNA or which modified proteins were excluded. In cases where an enzyme could modify both a small molecule and a macromolecule, we favoured a more liberal criterion of inclusion. A protein was considered a small molecule transporter if it formed a pore or channel through which a small molecule could traverse a lipid bilayer. Accessory or regulatory subunits of larger protein complexes were generally excluded.

**Meta-analysis of oncogenomic data.** To generate a cancer-relevant 'high priority' subset of metabolic genes (out of the 2,752 genes we classified as metabolic enzymes or small molecule transporters), we first identified those genes whose expression is significantly associated with the transformed state, advanced breast cancer, or stemness. Genes associated with the transformed state were obtained by analysing 36 gene expression studies deposited in Oncomine<sup>18</sup> that profiled normal human tissue and primary tumours derived from them. The gene expression profiles in each study were classified as normal or tumour and for each group the log<sub>2</sub> median centred intensity for each gene was determined. A *P* value associated with the significance of the difference between the two groups was calculated with the Student's *t*-test. After ranking the genes based on the *P* values, the top 10% of the genes with lowest *P* values were selected from each of the 36 studies. From these genes we identified those that are in the top 10% of the most upregulated metabolic genes across the all 36 studies at a *P* value < 0.05. Genes associated with aggressive breast cancer were obtained by analysing 15 gene expression studies from Oncomine that profiled ER<sup>-</sup> versus ER<sup>+</sup> tumours, grade 3 versus grade 1 or 2 tumours, tumours of basal versus epithelial morphology, or tumours from patients who failed to survive after 5 years of follow-up versus those who did survive at 5 years. The 15 studies were analysed as above to identify those genes that are in the top 10% of the most upregulated metabolic genes across the studies at a *P* value < 0.05. To identify genes associated with stemness, we analysed gene expression studies comparing differentiated cells with stem cells<sup>19</sup>, chromatin immunoprecipitation studies of stem-cell-associated transcription factors<sup>20,21</sup>, and a previous meta-analysis of stemness-associated genes<sup>22</sup>. Genes were considered to be associated with stemness if their average expression was greater than fourfold upregulated in the stem versus differentiated cells profiles analysed previously<sup>19</sup> or if

their promoters were bound by at least two stem-cell-specific transcription factors (Oct4, Nanog, Sox2, Tcf3, Dax1, Nac1 or Klf4) in both studies analysed. To generate the final high priority set of 133 genes that was screened (Supplementary Table 2), three categories of genes were selected: (1) genes scoring in all three analyses; (2) the most significantly scoring ~5% of genes in any one category; and (3) the most significantly scoring ~10% of genes in any two categories.

**Identification of cell lines for use in pooled screening.** To undertake negative-selection RNAi screening, a cell line that could form a tumour upon injection of the minimum number of cells was identified. To accomplish this, 11 breast cell lines that previously identified as capable of forming tumours were selected and 100,000 cells from each were injected into the fourth murine mammary fat pad. The cell lines tested included BT-20, BT-474, MCF10DCIS.com, HBL100, MCF7, MDA-MB-157, MDA-MB-231, MDA-MB-361, MDA-MB-453, T47D and ZR-75-1. After one month, tumours were scored by size and number scoring per site, and tumours or injection sites were analysed histologically to verify the presence of a tumour, or to identify microscopic tumours. In the timeframe of the experiment, MDA-MB-231, MDA-MB-361, MDA-MB-453, MCF7 and T47D cells formed microscopic tumours, whereas MCF10DCIS.com formed large tumours and ZR-75-1 formed small macroscopic tumours reproducibly. MCF10DCIS.com cells were then injected into murine mammary fat pads at 100,000, 10,000, 1,000 and 100 cells per site. All of these injections were capable of forming tumours, and tumour size correlated with the number of cells injected. The MCF10DCIS.com cell line was finally shown to be suitable for *in vivo* screening upon performing a screen using 180 shRNAs and demonstrating that nearly all of the shRNAs introduced initially could be recovered from the tumour and that replicate tumours exhibited significant correlation in those shRNAs over- or underrepresented compared to the injected pool. These experiments should not be construed to indicate that the excluded cell lines would not also be suitable for *in vivo* screening, as they were not tested using an shRNA pool.

**Pooled shRNA screening.** pLKO.1 lentiviral plasmids encoding shRNAs targeting the 133 transporters and metabolic enzymes listed in Supplementary Table 2 were obtained and combined to generate two plasmid pools. One contained the plasmids encoding shRNAs targeting all 47 transporters and another the plasmids encoding shRNAs targeting all 86 metabolic enzymes as well as control shRNAs designed not to target any gene. These plasmid pools were used to generate lentivirus-containing supernatants as described<sup>23</sup>. MCF10DCIS.com cells were infected with the pooled virus so as to ensure that each cell contained only one viral integrant. Cells were selected for 3 days with 0.5  $\mu\text{g ml}^{-1}$  puromycin. For the *in vivo* screen, cells were injected in 33% growth factor reduced matrigel into the fourth mammary fat pad of NOD.CB17 Scid/J mice (Jackson Labs) at 100,000 to 1,000,000 cells per injection site and tumours were harvested 4 weeks after implantation. For the *in vitro* screen, cells were plated in replicates of four at 1,000,000 per 10-cm plate and split at 1:8 once confluent (every 3–5 days) for 25–28 days. Genomic DNA was isolated from tumours or cells by digestion with proteinase K followed by isopropanol precipitation. To amplify the shRNAs encoded in the genomic DNA, PCR was performed for 33 cycles at an annealing temperature of 66 °C using 2–6  $\mu\text{g}$  of genomic DNA, the primer pair indicated below, and DNA polymerase. So that PCR products obtained from many different tumours could be sequenced together, forward primers containing unique 2-nucleotide barcodes were used (see below). After purification, the PCR products from each tumour were quantified by ethidium bromide staining after gel electrophoresis, pooled at equal proportions, and analysed by high-throughput sequencing (Illumina) using the primer indicated below. shRNAs from up to 16 genomic DNA samples were sequenced together. Sequencing reads were deconvoluted using GNU Octave software by segregating the sequencing data by barcode and matching the shRNA stem sequences to those expected to be present in the shRNA pool, allowing for mismatches of up to 3 nucleotides. The Log<sub>2</sub> values reported are the average Log base 2 of the fold change in the abundance of each shRNA in the pre-injection cells compared to tumours, for *n* = 5 tumours for the transporter pool and *n* = 12 tumours for the metabolic enzyme pool, or to cells at day 25–28 for *n* = 4 *in vitro* cultures. *P* values were determined by two-sided homoscedastic unpaired *t*-test comparing each shRNA to a basket of negative-control shRNAs contained within the shRNA pools. Individual shRNAs were identified as scoring in the screens using a *P* value cutoff of 0.05 and Log<sub>2</sub> fold-change cutoff of -1. Genes for which >75% of the shRNAs targeting the gene scored were considered hits. Individual shRNAs were considered to be differentially required *in vitro* versus *in vivo* using a *P*-value cutoff of 0.05 by a two-sided homoscedastic unpaired *t*-test comparing the *in vitro* and *in vivo* shRNA Log<sub>2</sub> fold change scores. For the transporter pool screen, this required normalization to the median of the two distributions. shRNAs present at less than 30 reads in the pre-injection cell sample were eliminated from further analysis.

Follow-up tumour growth studies of individual genes followed a similar timeline as above, except that during PHGDH and PSPH follow-up (Fig. 1e and

Supplementary Fig. 6), 10 days elapsed between infection and injections, whereas 5 days elapsed for all other validated genes (Supplementary Fig. 2). For doxycycline-inducible constructs, MDA-MB-468 cells were infected with *GFP*- or *PHGDH*-targeting shRNAs, puromycin selected and injected into the fourth murine mammary fat pad as above. Once tumours were palpable in all animals (25 days post-injection), doxycycline chow (600 p.p.m.) was provided to a randomly assigned set of animals for the duration of the experiment. Caliper measurements were taken every 4–6 days and tumour volume was estimated by  $0.5 \times W \times W \times L$ , where  $W$  is width and  $L$  is length. All experiments involving mice were carried out with approval from the Committee for Animal Care at MIT and under supervision of the Department of Comparative Medicine at MIT.

**Primers for amplifying shRNAs encoded in genomic DNA:** Barcoded forward primer (N indicates location of sample-specific barcode sequence): AATGATACGGCGACCACCGAGAAAGTATTTTCGATTTCTTGGCTTTATATATCTTGTGGAANNAGCAAAC. Common reverse primer: CAAGCAGAAGACGGCATA CGAGCTCTCCGATCTTGTGGATGAATACTGCCATTTGTCTCGAGGTC. Illumina sequencing primer: AGTATTTTCGATTTCTTGGCTTTATATATCTTGTGGAA.

**Analysis of gene copy number data.** The significance of copy number alteration across multiple data sets was determined using the GISTIC algorithm with methods described in ref. 6 and using the data deposited at <http://www.broadinstitute.org/tumorscape>.

**Determination of proportion of tumours with *PHGDH* overexpression.** To determine the percentage of breast cancers with elevations in *PHGDH* mRNA levels, data deposited in OncoPrint from ref. 8 were used. An ER<sup>-</sup> tumour was considered to have elevated *PHGDH* mRNA if the expression level was higher than 1.5 s.d. above the mean expression level in the ER<sup>+</sup> class (~91st percentile). For the percentage of breast cancer exhibiting elevated *PHGDH* protein, data reported in Fig. 2c were used. An ER<sup>-</sup> tumour was considered to have elevated *PHGDH* protein if the immunohistochemical staining signal was classified as 'high'.

**Cell proliferation assays.** For *PHGDH* or *PSPH* knockdown experiments, 10,000–20,000 MDA-MB-468, BT-20, HCC70, MCF-7, or MDA-MB-231 cells were infected with shRNA-expressing lentiviruses of known titres at a multiplicity of infection of 2.5 to 5. Cells were cultured in 12-well plates and infected via a 30-min spin at 2,250 r.p.m. in a Beckman Coulter Allegra X-12R centrifuge with an SX4750 rotor and uPlate Carrier attachment followed by an overnight incubation in media containing polybrene. Eight days after infection the number of cells was determined using a Coulter Counter (Beckman) and used to calculate relative cell proliferation. Where indicated, standard RPMI media was supplemented with serine to concentrations fivefold that of the serine already in the media. Where indicated, supplementation occurred at one and four days after lentiviral infection. For serine depletion experiments, cells were plated out as described above and the following day the standard culture medium was replaced with medium lacking serine or reconstituted with 1× serine. Dialysed serum (3 kDa MWCO) was used in serine depletion experiments except in the case of MCF-10A cells, where standard 5% serum was used.

**Immunohistochemistry and immunoblotting.** Immunoblotting was performed as described<sup>24</sup>. *PHGDH* protein levels were quantified using an Odyssey Infrared Imager (Li-Cor). For each measurement, the *PHGDH* signal obtained was normalized to the RPS6 signal from the same lane after accounting for background fluorescence. Immunohistochemistry was performed on formalin-fixed paraffin-embedded sections using a boiling Dako antigen retrieval method, as described<sup>25</sup>. A 1:250 dilution of the *PHGDH* antibody was used. A pathologist scored, in a blinded fashion, the intensity of the *PHGDH* staining in the breast tumour samples using a scale of 0–3 that represents none/weak, moderate and strong staining. Use of the tumour samples for *PHGDH* staining was approved by Institutional Review Boards at MIT (Protocol Number 1005003872) and Massachusetts General Hospital (Protocol Number 2010-P-001505/1).

**Glucose and lactate measurements.** Cells infected with shRNAs were plated on the day after infection at 5,000 cells per well of a 96-well plate in RPMI-10 alone or with 25 μM 3-bromopyruvate in a total of 200 μl media. On day 4 after infection, media was collected from the wells and cells were washed once with phosphate buffered saline before lysis in 50 mM NaOH. Lysate was mixed well and protein measured by BCA protein assay (Pierce). To determine the integrated protein content over the course of the assay (μg protein × days), a model was constructed with the following assumptions: control cells underwent two population doublings, cells proliferated exponentially to the final protein content, and the initial protein content for all samples was equivalent. Glucose concentration in the media was measured by glucose oxidase and peroxidase assay (Thermo Electron) and compared to control wells containing media with no cells to determine the quantity consumed. Lactate was measured by adding 5 μl of media to a solution containing 0.3 M glycine/hydrazine solution (Sigma G5418), 2.4 mM NAD<sup>+</sup> (Fisher Scientific NC9877003), and 2 μl ml<sup>-1</sup> lactate dehydrogenase (5 U μl<sup>-1</sup>,

Roche 10127230001) in a 200 μl total volume in a 96-well microtitre plate. Plates were mixed briefly and incubated for 30 min at 37 °C before reading absorbance at 340 nm. Lactate concentration was determined by comparison to a lactic acid standard (10 mM–0 mM, Fisher Scientific AC18987-0050) and compared to control wells containing media with no cells to determine the quantity produced. **Metabolite measurements.** For metabolite measurements, cells were cultured in cell-line-appropriate culture media (see above) in 10-cm dishes to approximately 70% confluence, typically by plating at  $2 \times 10^6$  cells per dish approximately 48 h before metabolite extraction. Twenty-four hours before metabolite extraction, culture media was replenished with media containing dialysed FBS. For metabolite extraction, cells in the culture dish were rapidly washed three times with 37 °C PBS, and then metabolites were extracted by addition of 80% aqueous methanol (pre-cooled in dry-ice) followed by incubation of culture dishes on dry ice for 15 min. For quantification, a <sup>13</sup>C-labelled internal metabolite standard for each analysed species was included in the extraction process. Cellular metabolite extracts were then collected by cell scraping and removal of the supernatant following centrifugation at 3,750 r.p.m. for 30 min (4 °C). The supernatants were then dried down using N<sub>2</sub> gas and stored dry at –80 °C before mass spectroscopy analysis. Four biological replicate samples were generated and analysed for each cell line. In addition, two parallel dishes of cells were trypsinized and counted using a Nexcelom cell counter; subsequent metabolite measurements were normalized to cell count.

All cell extracts were analysed by liquid chromatography-triple quadrupole mass spectrometry (LC-MS) using scheduled selective reaction monitoring (SRM) for each metabolite of interest, with the detector set to negative mode. Prior to injection, dried extracts were reconstituted in LCMS grade water. LC separation was achieved by the method reported<sup>26</sup>. Extracted metabolite concentrations were calculated from standard metabolite build-up curves using natural <sup>12</sup>C synthetic metabolites and normalized against cell number as well as the internal <sup>13</sup>C-labelled metabolite standards added at the time of metabolite extraction.

**Flux analysis.** For aKG flux studies, cells were plated at 250,000 cells per well in 6-well culture dishes in typical culture media (see above). Twenty-four hours before the flux study timecourse, media was replenished with fresh RPMI media containing dialysed FBS. For the flux study timecourse, standard RPMI culture media with dialysed FBS was used and the glutamine was replaced with U-<sup>13</sup>C glutamine (2 mM final concentration, matching the glutamine concentration in standard RPMI culture media). At the relevant time points, metabolites were harvested as noted above.

Serine pathway flux was measured using extracellular α-<sup>15</sup>N-glutamine, which is taken up by cells and becomes intracellular α-<sup>15</sup>N-glutamate at a very high rate. The activity of PSAT1 (conversion of phospho-hydroxypyruvate to phosphoserine) is coupled to the transfer of the α-<sup>15</sup>N-amino nitrogen of glutamate to phospho-hydroxypyruvate, generating aKG and α-<sup>15</sup>N-phosphoserine. As extracellular serine is in equilibrium with the intracellular pool, the rate of accumulation of extracellular α-<sup>15</sup>N-serine can be used to assess the activity of the serine biosynthetic pathway, and is proportional to the overall serine biosynthetic flux. For these flux studies, cells were plated at 250,000 cells per well in 6-well culture dishes in typical culture media (see above). When cells reached 60–70% confluence (typically 24–48 h post-cell plating), media was replenished with fresh media containing dialysed FBS and α-<sup>15</sup>N-glutamine (2 mM final concentration). For the data presented in Fig. 3b, MFC10A medium was used (see above) to permit the inclusion of the MCF10A cell line, whereas for the data presented in Supplementary Fig. 6a, RPMI medium was used. Therefore, these data are not directly comparable between these two panels. Samples of media were collected from four biological replicates, at this initial time point and following 24 h of additional culture. α-<sup>15</sup>N-serine was extracted from 300 μl of the sample media by addition of 3 volumes of acetonitrile, followed by collection of the supernatant following centrifugation for 30 min at 3,750 r.p.m. Supernatant was then dried down using N<sub>2</sub> gas and the dry samples stored at –80 °C until mass spectrometry. In parallel with the metabolite extracts, two replicate wells were trypsinized and counted at the initial time point as well as the 24 h time point. The average of these four wells was used for subsequent cell number normalization. Prior studies established the linearity of production of serine over this timecourse, and demonstrated that the intracellular and extracellular serine pools are at steady-state equilibrium, enabling measurement of a lower-bound phosphoserine pathway flux by sampling extracellular α-<sup>15</sup>N-serine. LC-MS analysis of <sup>15</sup>N-serine was performed using SRM in positive mode; separation was accomplished using an Atantis HILIC Silica 5 μm (2.1 × 100 mm) column and a gradient of 10 mM of ammonium formate in Water (mobile phase A: aqueous 0.1% formic acid) and acetonitrile (mobile phase B, 0.1% formic acid) with mobile phase A linearly increasing from 5% to 60% over 4 minutes. Following a 2 min isocratic period, the system was returned to initial conditions for a total cycle time of 9 min at a flow rate of 200 μl min<sup>-1</sup>. For flux studies, <sup>13</sup>C-labelled internal standards were omitted in both sample extracts and standard metabolite build-up curves.

**Flux modelling.** Ordinary differential equation models were constructed for two relevant portions of central carbon metabolism, based on the schematics shown in Supplementary Fig. 9a (models (i) and (ii)). Each model consisted of 3 differential equations with the constraints of balanced flux imposed on them. These equations describe the rates of loss of unlabelled forms of metabolites after feeding of 100% U-<sup>13</sup>C glucose or U-<sup>13</sup>C glutamine containing media.

The fluxes were identified by minimization of an objective function to the empirical data. The choice of objective function was  $\chi^2$ , defined as

$$\chi^2 = \sum_{k=1}^n \left( \frac{y_k - y(t_k; F)}{\sigma_k} \right)^2$$

where  $y_k$  is data point  $k$  with standard deviation  $\sigma_k$ , and  $y(t_k; F)$  is the value estimated by the model value at time point  $k$  for the set of fluxes  $F$ . Initial fluxes before the first optimization were arbitrarily chosen as 0.1. Three independent runs of 400 fits with the trust region approach were performed, each starting from the parameter values of the currently best fit randomly disturbed by up to 4 orders of magnitude.

**Model (i).** The schematic of the upper part of glycolysis (Supplementary Fig. 9a (i)) shows that  $F_2$  is the upper bound of the glycolytic flux that can be diverted to the pSer pathway. We estimated  $F_2$  by fitting the model to the time course of unlabelled metabolites (3PG, PEP and lactate) obtained using LC-MS of extracts from MDA-MB-468 and MDA-MB-231 cell lines, amplified and non-amplified PHGDH cell lines respectively. Three independent simulations of 400 fits were run for both the cell lines. The quality of fit was characterized by  $\chi^2$  values. The best 10% of the fits that also had  $P$  value above a significance threshold (0.05) were chosen for the analysis. The values of the parameter had a high variability suggesting that the parameter search space resembled a shallow basin. This was confirmed by generating the  $\chi^2$  landscapes for all possible pairs of parameters (data not shown). This observation suggested that additional constraints would greatly improve the predictive power of our model. Because each molecule of glucose that proceeds through glycolysis is broken into two molecules of 3PG, we imposed the requirement that  $F_1$  cannot be greater than twice the measured glucose

consumption rates (82 nmol per million cells per min). This additional constraint narrows down the solution of fluxes significantly, providing the results reported in the tables.

**Model (ii).** The schematic of the upstream reactions in glutaminolysis (Supplementary Fig. 9a (ii)) shows that  $F_1 + F_2$  is the glutamate to aKG flux. We estimated the fluxes as described above by fitting the model to the time course of unlabelled metabolites (glutamine, glutamate and aKG) obtained using LC-MS for MDA-MB-468 cells with and without PHGDH suppression via RNAi. Identical statistical thresholds were applied as for model (i) (top 10% and  $P > 0.05$ ) to choose solutions for the analysis. Unlike model (i), the parameters converged very well without need for further constraint, confirmed by generating the  $\chi^2$  landscapes for all possible pairs of parameters (data not shown).

- Debnath, J. & Brugge, J. S. Modelling glandular epithelial cancers in three-dimensional cultures. *Nature Rev. Cancer* **5**, 675–688 (2005).
- Rhodes, D. R. *et al.* Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166–180 (2007).
- Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
- Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–1061 (2008).
- Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).
- Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature Genet.* **40**, 499–507 (2008).
- Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proc. Natl Acad. Sci. USA* **105**, 20380–20385 (2008).
- Sancak, Y. *et al.* Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell* **141**, 290–303 (2010).
- Kalaany, N. Y. & Sabatini, D. M. Tumours with PI3K activation are resistant to dietary restriction. *Nature* **458**, 725–731 (2009).
- Buescher, J. M., Moco, S., Sauer, U. & Zamboni, N. Ultrahigh performance liquid chromatography–tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Anal. Chem.* **82**, 4403–4412 (2010).