Supplementary Methods

Materials

Materials were obtained from the following sources: antibodies to PHGDH (HPA021241) and PSPH (HPA020376) from Sigma; an antibody against PYCR1 (13108-1-AP) from Proteintech; an antibody against GMPS (A302-417A) from Bethyl Labs; an antibody against VDAC1 (ab16814) from Abcam; an antibody 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; a-15N-glutamine from Isotech/Sigma (486809); L-[3H(G)]-Serine from Perkin Elmer; Infinity Glucose Oxidase Liquid Stable Reagent (TR15221) from Thermo Electron; U-13C-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 (4). The TRC#s 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; VDAC1_1, TRCN0000029126; VDAC1_2, TRCN0000029127; 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 (15).

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 (17). 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 were 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/gene) and the available literature were used to identify known or putative gene function and to identify functional homologs. A gene was considered a metabolic enzyme if it modified a small molecule to generate another small molecule. Genes 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 favored a more liberal criterion of inclusion. A gene 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 analyzing 36 gene expression studies deposited in Oncomine (18) 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 log2 median centered 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 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 analyzing 15 gene expression studies from Oncomine that profiled ER-negative versus ER-positive 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 survived at 5 years. The 15 studies were analyzed as above to identify those genes which 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 analyzed gene expression studies comparing differentiated cells with stem cells (19), chromatin immunoprecipitation studies of stem cell-associated transcription factors (20, 21), and a previous meta-analysis of stemness-associated genes (22). Genes were considered to be

associated with stemness if their average expression was greater than 4-fold upregulated in the stem versus differentiated cells profiles analyzed by Mikkelsen et al. (19) 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 analyzed. 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

In order to undertake negative selection RNAi screening, a cell line which could form a tumour upon injection of the minimum number of cells was identified. To accomplish this, 11 breast cell lines which previously identified as capable of forming tumours were selected and 100,000 cells from each were injected into the 4^{th} 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 analyzed 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 under-represented 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 (23). 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 ug/mL 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 10cm 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 ug 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 analyzed 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 Log2 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 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 Log2 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 Log2 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 4th murine mammary fat pad as above. Once tumours were palpable in all animals (25 days post injection), doxycycline chow (600ppm) 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 ½*W*W*L. 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**'s indicate location of sample-specific barcode sequence):
AATGATACGGCGACCACCGAGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTG
GAA**NN**GACGAAAC

Common Reverse Primer:
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTGTGGATGAATACTGCCATTTG
TCTCGAGGTC

Illumina Sequencing Primer: AGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAA

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 (6) and using the data deposited at http://www.broadinstitute.org/tumourscape.

Determination of proportion of tumours with PHGDH over-expression

To determine the percentage of breast cancers with elevations in PHGDH mRNA levels, data deposited in Oncomine from van de Vijver et al (8) was utilized. An ER-negative tumour was considered to have elevated PHGDH mRNA if the expression level was higher than 1.5SD above the mean expression level in the ER-positive class (~91st percentile). For the percentage

of breast cancer exhibiting elevated PHGDH protein, data reported in Fig. 2c was utilized. An ER-negative 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 titers at a multiplicity of infection of 2.5 to 5. Cells were cultured in 12-well plates and infected via a 30-minute spin at 2,250 RPM 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 5-fold 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 1x serine. Dialyzed serum (3kDa MWCO) was utilized in serine depletion experiments except in the case of MCF-10A cells, where standard 5% serum was utilized.

Immunohistochemistry and immunoblotting

Immunoblotting was performed as described (24). 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 (25). 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 the Massachusetts Institute of Technology (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 25uM 3-bromopyruvate in a total of 200uL media. On day 4

after infection, media was collected from the wells and cells were washed once with phosphate buffered saline before lysis in 50mM 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 (ug 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 5uL of media to a solution containing 0.3M glycine/hydrazine solution (Sigma G5418), 2.4mM NAD+ (Fisher Scientific NC9877003), and 2uL/mL Lactate Dehydrogenase (5U/uL, Roche 10127230001) in a 200uL total volume in a 96-well microtiter plate. Plates were mixed briefly and incubated for 30 min at 37C before reading absorbance at 340nm. Lactate concentration was determined by comparison to a lactic acid standard (10mM-0mM, 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 2e6 cells per dish approximately 48 hours prior to metabolite extraction. 24 hours prior to metabolite extraction, culture media was replenished with media containing dialyzed FBS. For metabolite extraction, cells in the culture dish were rapidly washed three times with 37˚C PBS, and then metabolites extracted by addition of 80% aq. methanol (pre-cooled in dry-ice) followed by incubation of culture dishes on dry ice for 15 minutes. For quantification, a ^{13}C-labeled internal metabolite standard for each analyzed species was included in the extraction process. Cellular metabolite extracts were then collected by cell scraping and removal of the supernatant following centrifugation at 3750 RPM for 30 minutes (4˚C). The supernatants were then dried-down using N_2 gas and stored dry at -80 prior to mass spectroscopy analysis. Four biological replicate samples were generated and analyzed 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 analyzed by liquid chromatography-triple quadrapole 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 by Zamboni *et al* [Analytical Chemistry 82, 4403 (2010)]. Extracted metabolite concentrations were calculated from standard metabolite build-up curves using natural ^{12}C synthetic metabolites and normalized against cell number as well as the internal 13C-labeled metabolite standards added at the time of metabolite extraction.

Flux analysis

For alpha-ketoglutarate flux studies, cells were plated at 250,000 cells per well in 6-well culture dishes in typical culture media (see above). 24 hours prior to the flux study timecourse, media was replenished with fresh RPMI media containing dialyzed FBS. For the flux study timecourse, standard RPMI culture media with dialyzed FBS was used and the glutamine was replaced with U-^{13}C glutamine (2 mM final concentration, matching the glutamine concentration in standard RPMI culture media). At the relevant timepoints, metabolites were harvested as noted above in the methods for 'Metabolite Measurements.'

Serine pathway flux was measured using extracellular α-^{15}N-glutamine, which is taken up by cells and becomes intracellular α-^{15}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-^{15}N-amino nitrogen of glutamate to phospho-hydroxypyruvate, generating alpha-ketoglutarate and α-^{15}N-phosphoserine. As extracellular serine is in equilibrium with the intracellular pool, the rate of accumulation of extracellular α-^{15}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 hours post-cell plating), media was replenished with fresh media containing dialyzed FBS and α-^{15}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 Figure 6b, RPMI medium was used. Therefore, these data are not comparable between these two panels. Samples of media were collected from four biological replicates, at this initial timepoint and following 24 hours of additional culture. α-^{15}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 minutes at 3750 rpm. Supernatant was then dried-down using N_2 gas and the dry samples stored at -80C until mass spectrometry. In

parallel with the metabolite extracts, two replicate wells were trypsinized and counted at the initial timepoint as well as the 24 hour timepoint. 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-^{15}N-serine. LC-MS analysis of ^{15}N-Serine was performed using SRM in positive mode; separation was accomplished using an Atalantis HILIC Silica 5µ (2.1 X 100mm) column and a gradient of 10mM of Ammonium Formate in Water (mobile phase A: aq. 0.1% Formic acid) and Acetonitrile (mobile phase B, 0.1% Formic acid). running from 5% of mobile phase A to 60% mobile phase A increasing linearly over 4min. Following a 2 minute isocratic period, the system was returned to initial conditions for a total cycle time of 9min at a flow rate of 200µl /min. For flux studies, ^{13}C-labeled internal standards were omitted in both sample extracts and standard metabolite build-up curves.

Flux Modeling

Ordinary differential equation models were constructed for two relevant portions of central carbon metabolism, based on the schematics shown in Suppl. Fig 8B (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 unlabeled forms of metabolites after feeding of 100% U-^{13}C glucose or U-^{13}C glutamine containing media.

The fluxes were identified by minimization of an objective function to the empirical data. The choice of objective function was χ^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 σ_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 prior to 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 (Supplemental Fig. 8b (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 unlabeled metabolites (3PG, PEP and Lactate) obtained using LCMS 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 χ^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 χ^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. Since 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/million cells/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 (Supplemental Fig. 8b (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 unlabeled metabolites (Glutamine, Glutamate and aKG) obtained using LCMS for MDA-MB468 cells with and without PHGDH suppression via RNAi. Identical statistical thresholds were applied as for Model (i) (top 10% and $p>0.05$) to chose solutions for the analysis. Unlike model (i), the parameters converged very well without need for further constraint, confirmed by generating the χ^2 landscapes for all possible pairs of parameters (data not shown).

References

[17] Debnath, J. & Brugge, J.S., Modelling glandular epithelial cancers in three-dimensional cultures. *Nat Rev Cancer* 5 (9), 675-688 (2005).

[18] Rhodes, D.R. *et al.*, Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 9 (2), 166-180 (2007).
[19] Mikkelsen, T.S. *et al.*, Dissecting direct reprogramming through integrative genomic analysis. *Nature* 454 (7200), 49-55 (2008).
[20] Kim, J., Chu, J., Shen, X., Wang, J., & Orkin, S.H., An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132 (6), 1049-1061 (2008).
[21] Marson, A. *et al.*, Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134 (3), 521-533 (2008).
[22] Ben-Porath, I. *et al.*, An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* 40 (5), 499-507 (2008).
[23] Luo, B. *et al.*, Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci U S A* 105 (51), 20380-20385 (2008).
[24] Sancak, Y. *et al.*, Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell* 141 (2), 290-303.
[25] Kalaany, N.Y. & Sabatini, D.M., Tumours with PI3K activation are resistant to dietary restriction. *Nature* 458 (7239), 725-731 (2009).