

LETTERS

Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression

Arun Sreekumar^{1,2,3,4}, Laila M. Poisson^{5*}, Thekkelnaycke M. Rajendiran^{1,3*}, Amjad P. Khan^{1,3*}, Qi Cao^{1,3}, Jindan Yu^{1,3}, Bharathi Laxman^{1,3}, Rohit Mehra^{1,3}, Robert J. Lonigro^{1,4}, Yong Li^{1,3}, Mukesh K. Nyati^{4,6}, Aarif Ahsan⁶, Shanker Kalyana-Sundaram^{1,3}, Bo Han^{1,3}, Xuhong Cao^{1,3}, Jaeman Byun⁷, Gilbert S. Omenn^{2,7,8}, Debashis Ghosh^{4,5,11}, Subramaniam Pennathur^{2,4,7}, Danny C. Alexander¹², Alvin Berger¹², Jeffrey R. Shuster¹², John T. Wei^{4,9}, Sooryanarayana Varambally^{1,3,4}, Christopher Beecher^{1,2,3} & Arul M. Chinnaiyan^{1,2,3,4,9,10}

Multiple, complex molecular events characterize cancer development and progression^{1,2}. Deciphering the molecular networks that distinguish organ-confined disease from metastatic disease may lead to the identification of critical biomarkers for cancer invasion and disease aggressiveness. Although gene and protein expression have been extensively profiled in human tumours, little is known about the global metabolomic alterations that characterize neoplastic progression. Using a combination of high-throughput liquid-and-gas-chromatography-based mass spectrometry, we profiled more than 1,126 metabolites across 262 clinical samples related to prostate cancer (42 tissues and 110 each of urine and plasma). These unbiased metabolomic profiles were able to distinguish benign prostate, clinically localized prostate cancer and metastatic disease. Sarcosine, an *N*-methyl derivative of the amino acid glycine, was identified as a differential metabolite that was highly increased during prostate cancer progression to metastasis and can be detected non-invasively in urine. Sarcosine levels were also increased in invasive prostate cancer cell lines relative to benign prostate epithelial cells. Knockdown of glycine-*N*-methyl transferase, the enzyme that generates sarcosine from glycine, attenuated prostate cancer invasion. Addition of exogenous sarcosine or knockdown of the enzyme that leads to sarcosine degradation, sarcosine dehydrogenase, induced an invasive phenotype in benign prostate epithelial cells. Androgen receptor and the ERG gene fusion product coordinately regulate components of the sarcosine pathway. Here, by profiling the metabolomic alterations of prostate cancer progression, we reveal sarcosine as a potentially important metabolic intermediary of cancer cell invasion and aggressivity.

To profile the 'metabolome' during prostate cancer progression, we used both liquid and gas chromatography coupled with mass spectrometry³ to interrogate the relative levels of metabolites across 262 prostate-related biospecimens (outlined in Supplementary Fig. 1). Specifically, 42 tissue samples and 110 matched specimens of plasma and post-digital-rectal-exam urine from biopsy-positive cancer patients ($n = 59$) and biopsy-negative control individuals ($n = 51$) were assayed (Fig. 1a). A total of 1,126 metabolites were quantified and, as expected, only a small percentage of these metabolites (15.6%) were shared across the disparate biospecimen types (Fig. 1a).

Evaluation of the unbiased metabolomic profiles of plasma or urine did not identify robust differences between biopsy-positive and -negative individuals. For plasma, 20 out of 478 (4%) metabolites

were differential (Wilcoxon $P < 0.05$), with a false discovery rate (FDR) of 99%. Likewise, for urine, 36 out of 583 (6%) metabolites were differential (Wilcoxon $P < 0.05$), with an FDR of 67%. Thus, our initial focus was directed towards understanding the tissue metabolomic profiles as they showed more robust alterations.

Tissue samples were derived from benign adjacent prostate ($n = 16$), clinically localized prostate cancer ($n = 12$, PCA) and metastatic prostate cancer ($n = 14$) patients. Selection of metastatic tissue samples from different sites (see Supplementary Table 2) minimized characterization of analytes specific to cells of non-prostatic origin. In total, high-throughput profiling of the tissues quantitatively detected 626 metabolites (175 named, 19 isobars and 432 metabolites without identification), of which 82.3% (515 out of 626) were shared by the three diagnostic classes (Fig. 1b). Notably, there were 60 metabolites found in PCA and/or metastatic tumours but not in benign prostate. These profiles were displayed as a heat map (Fig. 1c) and *z*-score plot (Fig. 1d). In the latter, benign-based *z*-scores were plotted for each of the 626 metabolites. The plots revealed robust metabolic alterations in metastatic tumours (*z*-score range: -13.6 to 81.9) compared to fewer changes in clinically localized prostate cancer samples (*z*-score range: -7.7 to 45.8).

We identified the differential metabolites between the PCA and benign samples using a two-sided Wilcoxon rank-sum test coupled with a permutation test ($n = 1,000$). A total of 87 out of 518 metabolites were differential across these two classes ($P < 0.05$, corresponding to a 23% FDR). For visualizing the relationship between the 87 altered metabolites, hierarchical clustering was used to arrange the metabolites on the basis of their relative levels across samples (Fig. 2a). Among the perturbed metabolites, 50 were increased in PCA samples whereas 37 were downregulated. Figure 2b displays the relative levels of the 37 named metabolites that were differential between benign prostate and PCA samples. Similarly, 124 out of 518 metabolites were found to be increased in the metastatic samples compared to the localized tumours and 102 compounds were downregulated ($P < 0.05$, corresponding to a 4% FDR). Figure 2c displays the levels of the 91 named metabolites altered in metastatic samples. A subset of six metabolites including sarcosine, uracil, kynurenine, glycerol-3-phosphate, leucine and proline were significantly increased on disease progression from benign to PCA to metastatic prostate cancer. These metabolites could potentially serve as biomarkers for progressive disease, one of the factors that motivated us to examine sarcosine in greater detail.

¹The Michigan Center for Translational Pathology, ²Center for Computational Medicine and Biology, ³Department of Pathology, ⁴The Comprehensive Cancer Center, ⁵Department of Biostatistics, ⁶Department of Radiation Oncology, ⁷Department of Internal Medicine, ⁸Department of Human Genetics, ⁹Department of Urology, ¹⁰Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ¹¹Department of Statistics and Huck Institute of Life Sciences, Penn State University, Pennsylvania 16802, USA. ¹²Metabolon, Inc. 800 Capitola Drive, Suite 1 Durham, North Carolina 27713, USA.

*These authors contributed equally to this work.

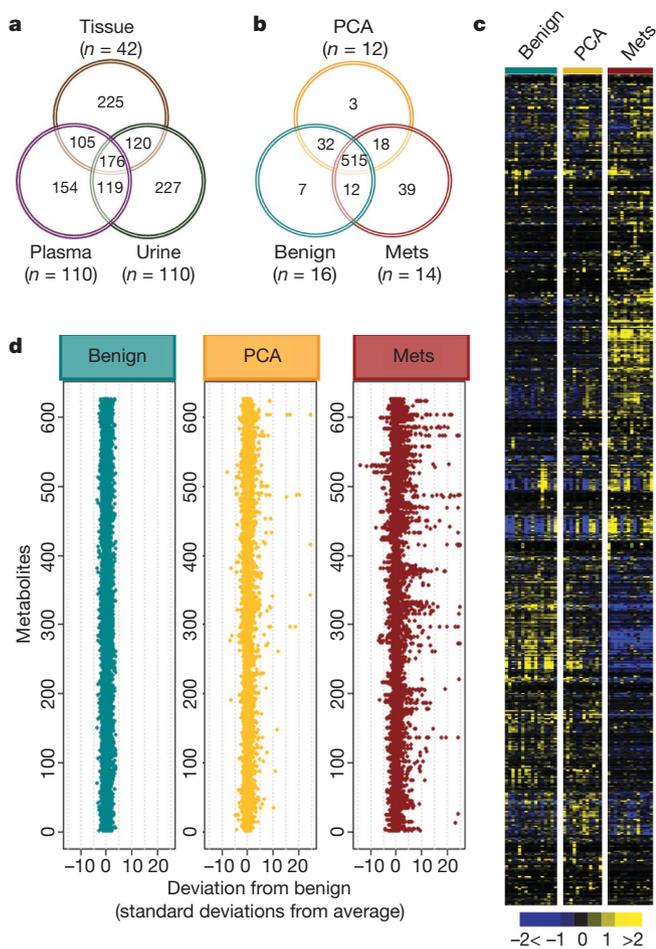


Figure 1 | Metabolomic profiling of prostate cancer. **a**, Venn diagram of the total metabolites detected across 42 prostate-related tissues and 110 matched plasma and urine samples. **b**, Venn diagram of 626 metabolites in tissues measured across 16 benign adjacent prostate tissues, 12 clinically localized prostate cancers (PCA) and 14 metastatic prostate cancers (Mets). **c**, Heat map representation of unsupervised hierarchical clustering of the data in **b** (rows) grouped by sample type (columns). Shades of yellow and blue represent an increase and decrease of a metabolite, respectively, relative to the median metabolite levels (see colour scale). **d**, z-score plots for the data in **b** normalized to the mean of the benign prostate samples (truncated at 25 s.d. for clarity, see Supplementary Methods for procedural details).

Mapping the differential metabolomic profiles to their respective biochemical pathways as outlined in the Kyoto Encyclopedia of Genes and Genomes (KEGG, release 41.1, <http://www.genome.jp/kegg>, Supplementary Fig. 8) revealed an increase in amino acid metabolism and nitrogen breakdown pathways during cancer progression to metastatic disease. A similar enrichment network of amino acid metabolism was also identified by the bioinformatics tool OncoPrint Concept Map^{4,5} (OCM, <http://www.oncoPrint.org>, $P = 6 \times 10^{-13}$, Supplementary Figs 9 and 10a), supporting our earlier gene-expression-based prediction of androgen-induced protein synthesis as an early event during prostate cancer development⁵. Additionally, OCM found strong enrichment for increased 'methyltransferase activity' (Supplementary Fig. 10b, $P = 7.7 \times 10^{-8}$) among metabolites upregulated in metastatic samples. This corroborates previous studies from our group and others showing an increase of the histone methyltransferase EZH2 in metastatic tumours^{6–11}.

Because amino acid metabolism and methylation were enriched during prostate cancer progression, we focused on differential metabolites that characterize these processes and additionally show a progressive increase from benign to PCA to metastatic disease. The amino acid metabolite sarcosine, an *N*-methyl derivative of glycine,

fits these criteria. Notably, metastatic samples showed markedly increased levels of sarcosine in 79% of the specimens analysed (chi-squared test, $P = 0.0538$), whereas 42% of the PCA samples showed an increase in the levels of this metabolite (Fig. 2a–c). Importantly, none of the benign samples had detectable levels of sarcosine. Taken together, this indicated the possible utility of sarcosine in monitoring disease progression and aggressiveness.

To confirm this pattern of sarcosine increase in cancer progression, we developed a highly sensitive and specific isotope dilution gas chromatography–mass spectrometry (GC–MS) method for accurately quantifying the metabolite from biospecimens (limit of detection = 10 femtomoles, Supplementary Fig. 11). In an independent set of 89 tissue samples (Supplementary Table 6), sarcosine levels were significantly increased in PCA specimens ($n = 36$) compared to benign adjacent prostate samples ($n = 25$, Wilcoxon $P = 4.34 \times 10^{-11}$, Fig. 3a). Additionally, there was an even greater increase of sarcosine in metastatic samples ($n = 28$) compared to organ-confined disease (Wilcoxon $P = 6.02 \times 10^{-11}$, Fig. 3a). In contrast, sarcosine was undetectable in adjacent non-neoplastic tissues from patients with metastatic disease (Supplementary Fig. 12a–c).

These findings led us to explore the potential of sarcosine as a candidate for future development in biomarker panels for early disease detection and aggressivity prediction. Towards this end, we monitored its levels in urine specimens from biopsy-positive and -negative individuals, most of whom have increased levels of prostate-specific antigen (PSA) ($>4.0 \text{ ng ml}^{-1}$) and in which prostate needle biopsy was used for diagnosis. This is a particularly challenging cohort as, in addition to these men being at high risk for prostate cancer, even a negative needle biopsy does not rule out the presence of cancer due to sampling issues. Sarcosine was found to be significantly higher in urine sediments ($n = 49$, Wilcoxon $P = 0.0004$, Fig. 3b) and supernatants ($n = 59$, Wilcoxon $P = 0.0025$, Supplementary Fig. 14a) derived from biopsy-positive prostate cancer patients compared to biopsy-negative controls ($n = 44$ and $n = 51$, respectively; Supplementary Tables 7 and 8). The overall receiver operating characteristic curves for sarcosine indicate that its predictive value is modest, with an area under the curve (AUC) of 0.71 for urine sediments and 0.67 for supernatants (Supplementary Fig. 14b, c). Notably, an AUC of 1.0 indicates perfect prediction and an AUC of 0.5 indicates prediction equivalent to random selection. When restricted to samples having PSA in the clinical grey zone of $2\text{--}10 \text{ ng ml}^{-1}$ ($n = 53$), sarcosine performed better than PSA in delineating the two diagnostic classes with an AUC of 0.69 (95% CI: 0.55, 0.84) compared to an AUC of 0.53 (95% CI: 0.37, 0.69) for PSA (Supplementary Fig. 15). Thus, sarcosine may have the potential to identify patients with modestly increased PSA that are likely to have a positive prostate biopsy.

To determine whether the sarcosine increase in prostate cancer has biological relevance, we measured its levels in prostate cancer cell lines VCaP, DU145, 22RV1 and LNCaP ($n = 3$ each) as well as in their benign epithelial counterparts, primary benign prostate epithelial cells (PrEC, $n = 2$) and immortalized benign RWPE prostate cells ($n = 3$). Significantly increased levels of the metabolite were found in prostate cancer cells compared to the benign cells (analysis of variance, ANOVA, $P = 0.0218$, Fig. 3c). Additionally, sarcosine levels correlated well with cell invasiveness (Spearman's correlation coefficient: 0.943, $P = 0.0048$). On the basis of our earlier findings that EZH2 overexpression in benign cells could mediate cell invasion and neoplastic progression^{6,10,11}, sarcosine levels were assessed on modulation of EZH2 expression. Interestingly, overexpression of EZH2 in benign prostate epithelial cells increased sarcosine levels (4.5-fold, Supplementary Fig. 16a), whereas its knockdown in DU145 prostate cancer cells diminished the levels of the metabolite (Supplementary Fig. 16a–c). To determine whether sarcosine has a more direct role in this process, we added the metabolite to non-invasive benign prostate epithelial cells. Alanine, an isomer of sarcosine, was used as a control for these experiments. Remarkably, the mere addition of exogenous sarcosine imparted an invasive phenotype to benign prostate epithelial cells

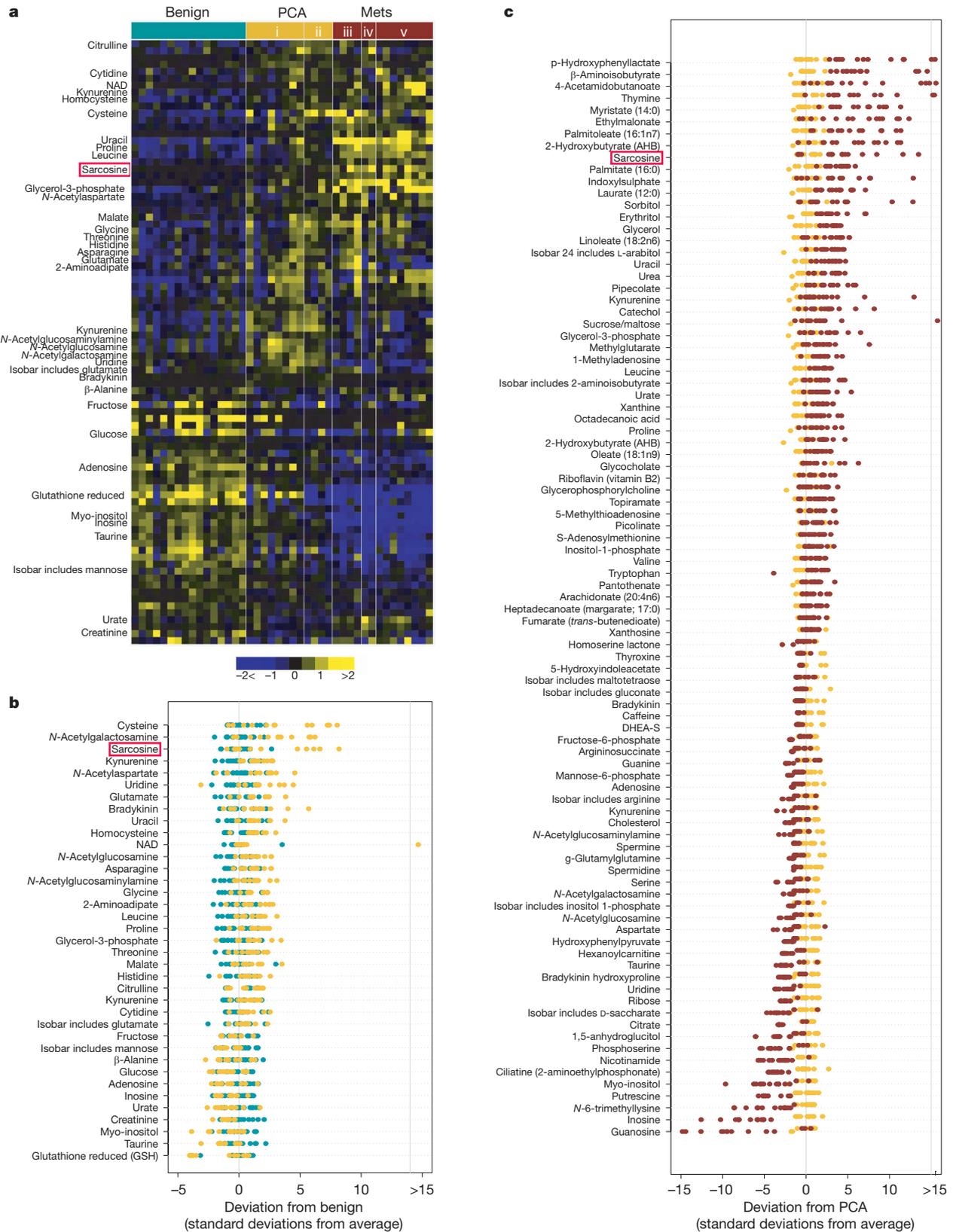


Figure 2 | Metabolomic alterations of prostate cancer progression. **a**, Heat map showing 87 differential metabolites in PCA relative to benign samples (Wilcoxon $P \leq 0.05$). Localized PCA samples are grouped as (i) low grade (Gleason 3+3, 3+4) or (ii) high grade (Gleason 4+3, 4+4). Metastatic samples are grouped by the site of tissue procurement, namely (iii) soft tissue, (iv) rib/diaphragm or (v) liver. **b**, Benign-based z-score plot of named

metabolites from **a**. Each point represents one metabolite in one sample, coloured by tissue type (blue, benign; yellow, PCA). **c**, As in **b** except for the comparison between Mets (red) and PCA (yellow), with data represented relative to the mean of the PCA samples. For clarity, the plots in **b** and **c** have been truncated at 15 standard deviations above the mean of the benign and PCA samples, respectively.

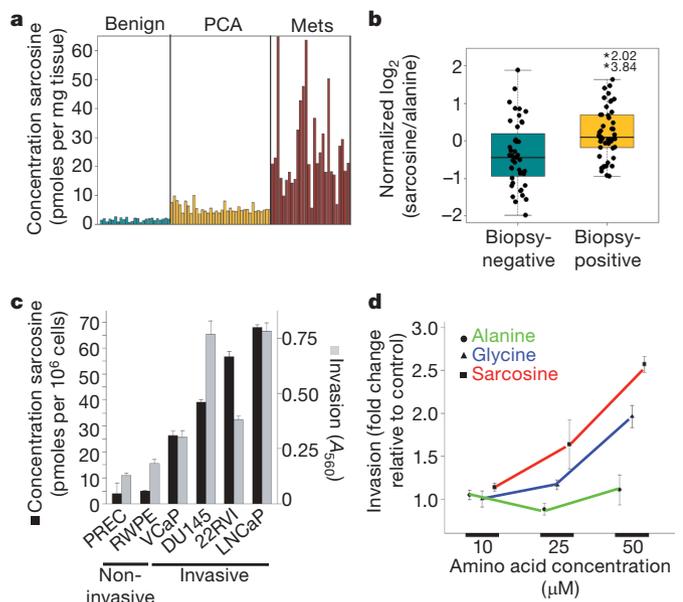


Figure 3 | Sarcosine levels in prostate cancer and its association with cell invasion. **a**, Sarcosine levels in prostate-cancer-related tissue specimens ($n = 89$). **b**, Sarcosine levels in post-digital-rectal-exam urine sediments from men with biopsy-proven prostate cancer ($n = 49$) and prostate-biopsy-negative controls ($n = 44$). Asterisks indicate truncated measures. **c**, Increased levels of sarcosine (black bars) were found in invasive prostate cancer cells compared to non-invasive benign prostate epithelial cell lines. Mean and s.e.m. of sarcosine levels ($n = 3$, except for PrEC cells where $n = 2$). Cell invasion (grey bars) was also measured (mean and s.e.m.). **d**, Assessment of cell invasiveness of prostate epithelial cells upon exogenous administration of alanine, glycine or sarcosine (mean and s.e.m., $n = 3$).

(Fig. 3d and Supplementary Fig. 17). Furthermore, the number of motile prostate epithelial cells was significantly higher on sarcosine treatment (t -test $P = 6.997 \times 10^{-6}$, $n = 10$) compared to alanine-treated controls (Supplementary Fig. 18). Sarcosine treatment, however, did not affect the ability of these cells to progress through the different stages of cell cycle (Supplementary Fig. 19a–d) or impair cell proliferation (Supplementary Fig. 19e). Notably, glycine, a precursor of sarcosine, induced invasion in these cells, although to a lesser degree than sarcosine (Fig. 3d). This invasion could result from the conversion of glycine to sarcosine by the enzyme glycine-*N*-methyltransferase (GNMT; Fig. 4a).

In addition to GNMT, sarcosine levels are regulated by sarcosine dehydrogenase (SARDH), the enzyme that converts sarcosine back to glycine, and dimethylglycine dehydrogenase (DMGDH), which generates sarcosine from dimethylglycine (Fig. 4a). By virtue of their ability to control sarcosine levels in cells, these enzymes may assume a critical role in modulating prostate cancer invasion. To test this hypothesis, a series of RNA-interference-mediated knockdown experiments were carried out. Attenuation of *GNMT* (Fig. 4b, Supplementary Fig. 20) in DU145 prostate cancer cells resulted in a significant reduction in cell invasion (t -test $P = 0.0073$, $n = 3$), with a concomitant threefold decrease in the intracellular sarcosine levels compared to control non-target short interfering RNA (siRNA)-transfected cells. Similar knockdown experiments performed in benign RWPE cells significantly hampered the ability of exogenous glycine (t -test $P = 0.0082$, $n = 3$), but not sarcosine, to induce invasion (Supplementary Fig. 21a, b). Comparable loss of cell invasion and reduction in sarcosine levels were also apparent in DU145 cancer cells on knockdown of *DMGDH* (Supplementary Fig. 22a, b). In contrast, knockdown of *SARDH* in benign prostate epithelial cells resulted in an ~ 3 -fold increase in endogenous sarcosine levels with a concomitant >3.5 -fold increase in invasion (Fig. 4c and Supplementary Fig. 23).

With the understanding that androgen signalling and ETS family of genes (*ERG*, *ETV1*) fusions are key factors for prostate cancer

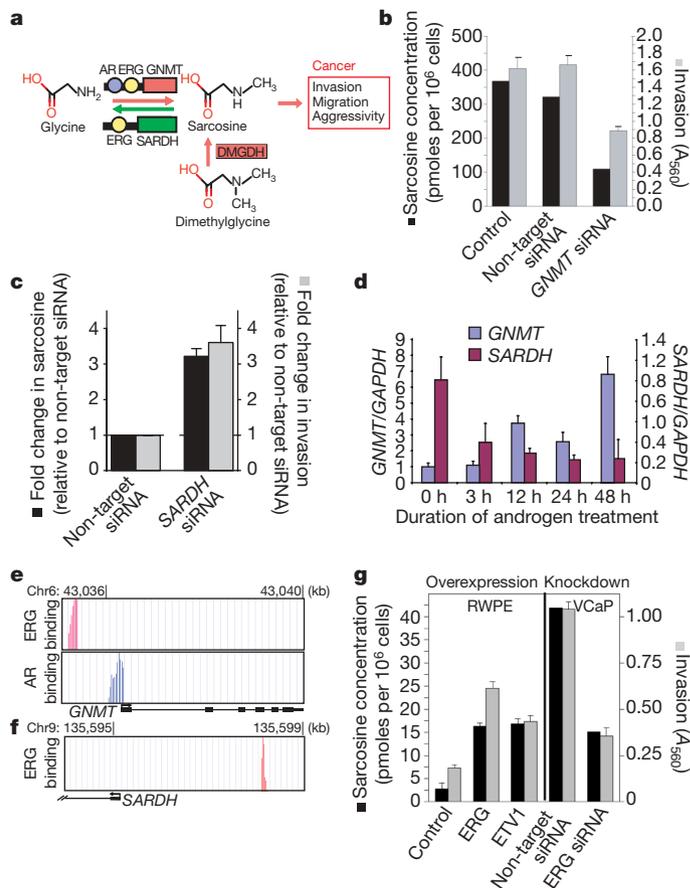


Figure 4 | A role for sarcosine in androgen signalling and prostate cancer cell invasion. **a**, Schematic of the sarcosine pathway and its potential link to prostate cancer. **b**, Assessment of sarcosine levels and cell invasiveness after knockdown of *GNMT* in DU145 cells by RNA interference. **c**, As in **b** except knockdown of *SARDH* in RWPE cells ($n = 6$). **d**, qRT-PCR analysis of *GNMT* and *SARDH* mRNA expression in androgen-stimulated VCaP cells. **e**, Androgen receptor (AR) and ERG binding sites on the promoter of *GNMT*, as determined by ChIP-seq. The y axes display the number of reads in a 25 bp sliding window. **f**, As in **e**, except ERG binding sites in the promoter of *SARDH*. **g**, Left, overexpression of ERG or ETV1 in RWPE cells and measurement of sarcosine levels and cell invasiveness. Right, as in left, except knockdown of *TMPRSS2-ERG* in VCaP cells by RNA interference. All error bars represent mean and s.e.m., $n = 3$ unless indicated otherwise.

progression¹², we investigated their role in regulating *GNMT* and *SARDH*. Treatment with androgen for 48 h in VCaP (*ERG*-positive) and LNCaP (*ETV1*-positive) prostate cancer cells resulted in a stepwise increase in *GNMT* expression and a concomitant decrease in *SARDH* levels, as assessed by digital gene expression and quantitative polymerase chain reaction (qPCR) (Fig. 4d and Supplementary Fig. 25e). This finding was supported by chromatin immunoprecipitation sequencing (ChIP-Seq), which revealed direct binding of androgen receptor and ERG to the promoter of *GNMT* in VCaP cells (Fig. 4e), whereas only ERG binding was seen on the *SARDH* promoter (Fig. 4f). In *ETV1*-positive LNCaP cells, androgen receptor, but not ERG as expected, was bound to both *GNMT* and *SARDH* promoters (Supplementary Figs 25a, b). The binding data were validated by ChIP-PCR (Supplementary Figs 24a, b, d and 25c, d), which additionally revealed weak binding of androgen receptor to the *SARDH* promoter in VCaP cells (Supplementary Fig. 24c). These findings together directly link activation of the sarcosine pathway to androgen receptor and ETS gene fusion regulation—two key mediators of prostate cancer progression. Remarkably, both ERG- and ETV1-induced invasion were associated with a threefold sarcosine increase in benign RWPE cells (Fig. 4g). Similarly, knockdown of the *TMPRSS2-ERG* gene fusion in VCaP cells (Supplementary Fig. 26)

resulted in a more than threefold decrease in sarcosine with a similar decrease in the invasive phenotype (Fig. 4g).

Taken together, we explored the metabolome of prostate cancer progression. This led to the characterization of metabolomic signatures, which in the context of other molecular alterations may lead to a more complete understanding of disease progression. Specifically, we identified sarcosine as a key metabolite increased most robustly in metastatic prostate cancer and detectable in the urine of men with organ-confined disease. Interestingly, sarcosine and its proximal regulatory enzymes seem to have an intermediary role in neoplastic progression modulating cell invasion and migration. The master transcriptional regulators of prostate cancer progression, androgen receptor and the ETS gene fusions, seem to regulate directly sarcosine levels by means of transcriptional control of its regulatory enzymes. Thus, components of the sarcosine pathway may have potential as biomarkers of prostate cancer progression and serve as new avenues for therapeutic intervention.

METHODS SUMMARY

Biospecimens and associated clinical data related to the study were collected with written consent from the University of Michigan and approved for use by the Internal Review Board. Unbiased metabolomic profiling using liquid/gas chromatography coupled to mass spectrometry (LC/GC-MS) was performed as described³ using a ThermoFisher linear ion-trap mass spectrometer with Fourier transform and Mat-95 XP mass spectrometers, respectively (Supplementary Fig. 1). Target metabolites were assessed in tissue and urine samples using isotope dilution GC-MS. Metabolomic data analysis is detailed in Supplementary Fig. 4. All Wilcoxon rank-sum tests and *t*-tests are two-sided using a threshold of $P < 0.05$ for significance. Repeated measures ANOVA is used for the cell line data with *P* values from the model F-test. Class-specific metabolomic patterns were visualized using *z*-score plots and heat maps. Unsupervised clustering of samples using metabolomic signatures was performed using Cluster¹³ and TreeView¹⁴, and visualized using heat maps. Analysis of network relationships among various molecular concepts and metabolomic data was performed using OCM^{4,5} (<http://www.oncomine.org>), as outlined in Supplementary Fig. 9. Invasion was measured using a modified Boyden chamber assay as described¹⁰. The cell motility assay was performed as reported previously using blue fluorescent microsphere beads¹⁵. Targeted knock-down of candidate genes¹⁶ using gene-specific siRNA sequences are listed in Supplementary Table 9. qPCR for enzymes regulating sarcosine levels, *EZH2* and *ETS*, was performed as described¹² using the indicated oligonucleotide primers (Supplementary Table 10). Chromatin immunoprecipitation to interrogate the regulatory role of androgen and ETS was performed using published protocols¹⁷. ChIP-Seq and digital gene expression were measured using the Genomic DNA Sample Prep Kit and the NIAIII kit on a Genome Analyser (Illumina) according to the manufacturer's instructions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 October 2008; accepted 6 January 2009.

1. Abate-Shen, C. & Shen, M. M. Molecular genetics of prostate cancer. *Genes Dev.* **14**, 2410–2434 (2000).
2. Ruijter, E. *et al.* Molecular genetics and epidemiology of prostate carcinoma. *Endocr. Rev.* **20**, 22–45 (1999).
3. Lawton, K. A. *et al.* Analysis of the adult human plasma metabolome. *Pharmacogenomics* **9**, 383–397 (2008).

4. Rhodes, D. R. *et al.* Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia* **9**, 443–454 (2007).
5. Tomlins, S. A. *et al.* Integrative molecular concept modeling of prostate cancer progression. *Nature Genet.* **39**, 41–51 (2007).
6. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629 (2002).
7. van der Vlag, J. & Otte, A. P. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nature Genet.* **23**, 474–478 (1999).
8. Laible, G. *et al.* Mammalian homologues of the Polycomb-group gene *Enhancer of zeste* mediate gene silencing in *Drosophila* heterochromatin and at *S. cerevisiae* telomeres. *EMBO J.* **16**, 3219–3232 (1997).
9. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
10. Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl Acad. Sci. USA* **100**, 11606–11611 (2003).
11. Varambally, S. *et al.* Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science* **322**, 1695–1699 (2008).
12. Tomlins, S. A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648 (2005).
13. Eisen, M. B. & Brown, P. O. DNA arrays for analysis of gene expression. *Methods Enzymol.* **303**, 179–205 (1999).
14. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
15. Klemke, R. L. *et al.* Regulation of cell motility by mitogen-activated protein kinase. *J. Cell Biol.* **137**, 481–492 (1997).
16. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629 (2002).
17. Yu, J. *et al.* A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Res.* **67**, 10657–10663 (2007).
18. Storey, J. D. A direct approach to false discovery rates. *J. R. Stat. Soc. [Ser B]* **64**, 479–498 (2002).
19. Yu, J. *et al.* Integrative genomics analysis reveals silencing of beta-adrenergic signaling by polycomb in prostate cancer. *Cancer Cell* **12**, 419–431 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Granger for help in manuscript preparation, J. Siddiqui and R. Varambally for help with the clinical database, and A. Vellaichamy and S. Pallela for technical assistance. We thank K. Pienta for access to metastatic prostate cancer samples from the University of Michigan Prostate SPOR rapid autopsy programme. This work is supported in part by the Early Detection Research Network (A.M.C., J.T.W.), National Institutes of Health (A.S., S.P., J.B., T.M.R., D.G., G.S.O. and A.M.C.) and an MTTC grant (G.S.O. and A.S.). A.M.C. is supported by a Clinical Translational Science Award from the Burroughs Wellcome Foundation. A.S. is supported by a grant from the Fund for Discovery of the University of Michigan Comprehensive Cancer Center. L.M.P. is supported by the University of Michigan Cancer Biostatistics Training Grant. A.M.C. and S.P. are supported by the Doris Duke Charitable Foundation.

Author Contributions A.S., L.M.P. and A.M.C. wrote the manuscript. A.S. and A.M.C. conceptualized, designed and interpreted the data. L.M.P., R.J.L., S.K.-S., D.G. and D.C.A. performed data analysis. T.M.R., G.S.O., J.B. S.P., J.R.S., A.B. and C.B. carried out the mass spectrometry studies. A.P.K., J.Y., Q.C., B.L., Y.L., M.K.N., A.A., X.C. and S.V. performed biochemical experiments. R.M., B.H., A.M.C. and J.T.W. coordinated the clinical and pathology components of the study.

Author Information The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.M.C. (arul@umich.edu).

METHODS

Biospecimens and cell lines. Prostate tissues, urine and plasma were obtained from the University of Michigan SPORE and EDNRN Tissue Core. All samples were collected with informed consent as per the approval of the Institutional Review Board.

RWPE, DU145, LnCAP and PC3 cells were obtained from ATCC, PrEC cells from Cambrex BioScience, 22-RV1 was provided by J. Macoska and VCaP by K. Pienta. VCaP and LnCAP were grown in charcoal-stripped-serum-containing media for 24 h, before treatment for a further 24 h with vehicle or 1 nM methyl-trienolone (R1881, NEN) dissolved in ethanol.

Metabolomic profiling. Metabolomic profiling was performed using the platform described previously³ and outlined in Supplementary Fig. 1. The LC-MS portion of the platform is based on a Surveyor HPLC and a Thermo-Finnigan LTQ-FT mass spectrometer (Thermo Fisher Corporation) with the instrument set for continuous monitoring of both positive and negative ions. Samples that were analysed by GC-MS were derivatized under dried nitrogen using bistrimethylsilyl-trifluoroacetamide (BSTFA) and analysed on a Thermo-Finnigan Mat-95 XP using electron impact ionization and high resolution. For both LC-MS and GC-MS, spectral files were searched using metabolomic libraries created by Metabolon that contain about 800 commercially available compounds.

Quantification of target metabolites was performed by isotope dilution GC-MS using selected ion monitoring (SIM). The samples were modified to their *t*-butyl dimethylsilyl derivatives and analysed with an Agilent 5975 MSD mass detector using electron impact ionization. For SIM analysis, the *m/z* for native and labelled molecular peaks for various target metabolites quantified were: 158 and 161 (sarcosine), 406 and 407 (cysteine), 432 and 437 (glutamic acid), 297 and 301 (thymine) and 218 and 219 (glycine), respectively. Assessment of citric acid was performed on the GC-MS in the full scan mode.

Statistical analysis. Missing metabolite measurements were replaced (imputed) with zero for metabolites where the mean relative standardized intensity measure was over 100,000 across the samples, that is, we assume missingness was probably due to absence of the metabolite in the sample. Otherwise one half of the sample minimum was used to replace the missing measurement, that is, we assume missingness was probably due to detection limits. Imputed data were median-centred and inter-quartile-range-scaled per sample. Plotted *z*-scores were calculated based on the mean and standard deviation of a reference set (benign samples, unless otherwise stated). Hierarchical clustering¹⁴ based on Pearson's correlation was performed on the log-transformed normalized data after median centring per metabolite. A small value (unity) was added to each normalized value to allow log transformation. Per-metabolite chi-squared tests were used to assess class-specific metabolite patterns of present and absent (undetected) measurements. Per-metabolite two-tailed Wilcoxon rank sum tests were used for two-sample tests of association between classes. Kruskal-Wallis tests were used for three-way comparisons between all diagnosis groups. Non-parametric tests were chosen to reduce the influence of the imputed values. Tests were run on those metabolites with detectable expression in at least 20% of the samples. Significance was determined

using 1,000 sample permutations. FDRs were calculated using the *q*-value conversion algorithm of ref. 18. Pairwise differences in expression in the cell line data and small-scale tissue data were tested using two-tailed *t*-tests with Satterthwaite variance estimation. Comparisons involving multiple cell lines used repeated measures ANOVA to adjust for the multiple measures per cell line. Fold change was estimated using ANOVA on a log scale, following the model $\log(Y) = A + B^* \text{treatment} + E$. In this way $\exp(B)$ is an estimate of $(Y | \text{treatment} = 1) / (Y | \text{treatment} = 0)$, and the standard error of $\exp(B)$ can be estimated from $\text{SE}(B)$ using the delta method. The threshold for significance was $P < 0.05$ for all tests.

ChIP-PCR. ChIP was carried out as described previously¹⁹ using antibodies against androgen receptor (Millipore), ERG (Santa Cruz) and rabbit IgG (Santa Cruz). Androgen receptor ChIP was performed in paired ethanol-treated and R1881-treated samples. ChIP-enriched chromatin as well as the whole-cell extract was amplified by ligation-mediated PCR. When examining androgen receptor binding on target genomic regions, equal amounts of ethanol-treated and R1881-treated ChIP amplicons were subjected to qPCR, and the fold enrichment (R1881/ethanol) was determined based on the cycle differences after normalization to input DNA. For ERG ChIP assays, VCaP cells grown in regular medium were used for ChIP using antibodies against ERG and rabbit IgG control. ChIP products were directly analysed by the qPCR assay, and ERG binding was evaluated based on the cycle difference between ChIP-enriched chromatin by ERG and corresponding IgG. The primers used are listed in Supplementary Table 9.

ChIP-Seq. ChIP samples were prepared for sequencing using the Genomic DNA Sample Prep kit following the manufacturer's protocols. ChIP-sequencing was performed using Illumina Genome Analyser according to standard manufacturer's procedures. The raw sequencing image data were analysed by our analysis pipeline and aligned to the unmasked human reference genome (NCBI v36, hg18) using ELAND software to generate sequence reads of 25–32 base pairs.

Digital gene expression analysis. Trizol-extracted RNA from samples with 0 h and 48 h androgen treatment was prepared for sequencing using the Digital Gene Expression-Tag Profiling with NIAIII kit (Illumina) and sequenced by the Genome Analyser. Sequencing reads were mapped back to the human reference genome using the ELAND software. The number of sequencing reads for genes of interest was counted. The expression level of each gene was measured as the number of transcripts per million of total sequencing reads.

Quantitative RT-PCR. qPCR was performed using SYBR Green Mastermix on an Applied Biosystems 7300 PCR machine as described previously¹⁹. All primers were designed using Primer 3 and synthesized by Integrated DNA Technologies, and are listed in Supplementary Table 9.

RNA interference. DU145 or RWPE cells were treated with non-targeting siRNA (Dharmacon) and gene-specific siRNA sequences as listed in Supplementary Table 10.

Cell invasion assay. Cell invasion assays were carried out using a modified basement membrane chamber assay as described previously¹⁹.

Cell motility assay. For cell motility assays, we used the Cellomics cell motility kit as per manufacturer's instructions.