SUPPLEMENTARY METHODS

Clinical Samples. Benign prostate and localized prostate cancer tissues were obtained from a radical prostatectomy series at the University of Michigan Hospitals and the metastatic prostate cancer biospecimens were from the Rapid Autopsy Program, which are both part of University of Michigan Prostate Cancer Specialized Program of Research Excellence (S.P.O.R.E) Tissue Core. Samples were collected with informed consent and prior institutional review board approval at the University of Michigan. Detailed clinical information of the tissue samples and matched plasma and urine specimens used in the profiling phase of this study are provided in **Supplementary Tables 2 and 3 respectively.** Analogous information for tissues and urine samples used to validate sarcosine are given in **Supplementary Tables 6 and 7** respectively. All the samples were stripped of identifiers prior to metabolomic assessment. For the profiling studies, tissue samples were sent to Metabolon, Inc. without any accompanying clinical information. Upon receipt, each sample was accessioned by Metabolon into a LIMS system and assigned unique 10 digit identifier. The sample was bar coded and this anonymous identifier alone was used to track all sample handling, tasks, results etc. All samples were stored at -80 ºC until use.

Metabolomic Profiling. The metabolomic profiling analysis of all samples was carried out in collaboration with Metabolon using the following general protocol (outlined in **Supplementary Fig. 1)** and described in Lawton *et al.* (2008)[1]. This process involved: sample extraction, separation, detection, spectral analysis, data normalization, delineation of class-specific metabolites, pathway mapping, validation and functional characterization of candidate metabolites (refer to **Supplementary Fig. 4** for an outline of the data analysis strategy). All samples were randomized prior to mass spectrometric analyses to avoid any experimental drifts (**Supplementary Fig. 3**). A number of internal standards, including injection standards, process standards, and alignment standards were used to assure QA/QC targets were met and to control for experimental variability (see **Supplementary Table 1** for description of standards). The reproducibility of the profiling process was addressed at two levels; one by measuring only instrument variation, and the other by measuring overall process variation (refer **Supplementary Table 1** for a list metrics and standards used to assess reproducibility). Instrument variation was measured from a series of internal standards (n=14 in this study) added to each sample just prior to injection. The median coefficient of variation (CV) value for the internal standard compounds was 3.9%. To address overall process variability, metabolomic studies were augmented to include a set of nine experimental sample technical replicates (also called matrix, abbreviated as MTRX), which were spaced evenly among the injections for each day. Reproducibility analysis for the n=339 compounds detected in each of these nine replicate samples gave a measure of the combined variation for all process components including extraction, recovery, derivatization, injection, and instrument steps. The median CV value for the experimental sample technical replicates (tissue profiling part of this study) was 14.6%. **Supplementary Fig. 2** shows the reproducibility of these experimental-sample technical replicates; Spearman's rank correlation coefficient between pairs of technical replicates ranged from 0.93 to 0.97.

The above authenticated process was used to quantify the metabolomic alterations in prostate-derived tissues. Specifically, this study included 42 tissue samples and 110 matched plasma and urine specimens derived post digital rectal examination from biopsy proven prostate cancer patients and biopsy negative control individuals (**Fig.1 a**). Included in the tissues were benign adjacent prostate specimens (n=16), clinically localized prostate cancers (PCA, n=12), and metastatic prostate cancers (Mets, n=14) (**Fig. 1b**). Additionally, selection of metastatic tissue samples from different sites minimized the contribution from non-prostatic tissue (see **Supplementary Table 2** for clinical information). Notably, the benign adjacent prostate specimen and the non-neoplastic control tissues (used for sarcosine validation, see below) from the metastatic patients (both controls in this study) were subjected to handling procedures that were identical to those for localized prostate cancer and metastatic tumors respectively.Tissue specimens were processed in two batches of 21 samples each. Samples from each of the three tissue diagnostic classes—benign prostate, PCA, and metastatic tumor—were equally distributed across the two batches (**Supplementary Fig. 3**). In otherwords, each batch contained samples from 8 benign prostates, 6 PCAs, and 7 metastatic tumor samples (**Supplementary Fig. 3**).

Sample Preparation. Samples were kept frozen until assays were to be performed. The sample preparation was programmed and automated using the MicroLab STAR® liquid handler (Hamilton). Sample extraction consisted of sequential organic and aqueous extractions. A recovery standard was introduced at the start of the extraction process. The resulting pooled extract was equally divided into a liquid chromatography (LC) fraction and a gas chromatography (GC) fraction. Samples were dried on a TurboVap® (Zymark, Claiper Life Science) to remove the organic solvent and lyophilized to dryness. As discussed below, all samples were resuspended in identical volume of the injection solvent prior to injection. Injection standards were introduced during this final re-solvation. In addition to controls and blanks, an additional well-characterized sample (for quality control verification) was included multiple times into the randomization scheme such that sample preparation and analytical variability could be constantly assessed.

Liquid Chromatography/Mass Spectroscopy (LC/MS). The LC/MS portion of the platform is based on a Surveyor HPLC and a LTQ-FT mass spectrometer (Linear Ion Trap mass spectrometer with Fourier Transform, Thermo Fisher Corporation). The LTQ side data was used for compound quantitation. The FT side data, when collected, was used only to confirm the identity of specific compounds. The instrument was set for continuous monitoring of both positive and negative ions. Some compounds are redundantly visualized across more than one of these data-streams, however, not only is the sensitivity and linearity vastly different from interface to interface but these redundancies, in some instances, are actually used as part of the quality control program.

The vacuum-dried sample was re-solubilized in 100 µl of injection solvent that contains no less than five injection standards at fixed concentrations. The chromatography was standardized and was never allowed to vary. Internal standards were used both to assure injection and chromatographic consistency. The chromatographic system was operated

using a gradient of Acetonitrile (ACN): Water (both solvents were modified by the addition of 0.1% TFA) from 5% to 100% over an 8 minute period, followed by 100% ACN for 8 min. The column was then reconditioned back to starting conditions. The columns (Aquasil C-18, Thermo Fisher Corporation) were maintained in temperature-controlled chambers during use and were exchanged, washed and reconditioned after every 50 injections. As part of Metabolon's general practice, all columns and solvents were purchased from a single manufacturer's lot at the outset of these experiments.

Gas chromatography/Mass Spectrometry (GC/MS). For the metabolomic profiling studies, the samples destined for GC were re-dried under vacuum desiccation for a minimum of 24 hours prior to being derivatized under dried nitrogen using bistrimethyl-silyl-triflouroacetamide (BSTFA). Samples were analyzed on a Mat-95 XP GC/MS (Thermo Fisher Corporation) using electron impact ionization and high resolution. The column used for the assay was (5% phenyl)-methyl polysiloxane. During the course of the run, temperature was ramped from 40° to 300° C in a 16 minute period. The resulting spectra were compared against libraries of authentic compounds.
For the entire phase of unbiased profiling studies samples were bar-coded by LIMS and chromatographic runs were LIMS-scheduled tasks. The raw data files were tracked and processed by their LIMS identifiers and archived to DVD at regular intervals. This was processed as described later.

Quantification of target metabolites using isotope dilution GC/MS.
For isotope dilution GC/MS analysis of sarcosine and alanine (in case of urine sediments, **Fig. 3b**), residual water was removed from the samples by forming an azeotrope with 100 uL of dimethylformamide (DMF), and drying the suspension under vacuum. All of the samples were injected using an on column injector into an Agilent 6890N gas chromatograph equipped with a 15-m DB-5 capillary column (inner diameter, 0.2 mm; film thickness, 0.33 micron; J & W Scientific) interfaced with a Agilent 5975 MSD mass detector. The *t*-butyl dimethylsilyl derivatives of sarcosine were quantified by selected ion monitoring (SIM), using isotope dilution electron-impact ionization GC/MS. The levels of alanine and sarcosine that eluted at 3.8 and 4.07 minutes respectively, were quantified using their respective ratio between the ion of *m*/*z* 232 derived from native metabolite ([M-O-*t*-butyl-dimethylsilyl]$^-$) and the ions of *m*/*z* 233 and 235 respectively for alanine and sarcosine, derived from the isotopically labeled deuteriated internal standard [$^2H_3$] for the compounds. A similar strategy was used for assessment of sarcosine, cysteine, thymine, glycine and glutamic acid in the tissues. The m/z for native and labeled molecular peaks for these compounds 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. The area under the peak with m/z 591 was measured and normalized to ^{13}C glycine peak (m/z: 219, internal standard) from the same run followed by normalization to the tissue weight. The resulting value was scaled down by dividing by a constant value of 1 million and used to plot the relative box plots. In case of urine supernatants (**Supplementary Fig. 14a**), sarcosine was measured and normalized to creatinine. Relative area counts for each compound were obtained by manual integration of its chromatogram peaks using Xcalibur software (Thermo Fisher Corporation). The data is

presented as the log of the ratio, (sarcosine ion counts)/(creatinine ion counts). For metabolite validation, all the samples were assessed by single runs on the instrument except for sarcosine validation of tissues wherein each sample was run as 4 analytical replicates and the average ratio was used for calculate sarcosine levels. The limit of detection (signal/noise > 10) was ~ 10 femtomoles for sarcosine using isotope dilution GC/MS.

Metabolomic Libraries. These were used to search the mass spectral data. The library was created using approximately 800 commercially available compounds that were acquired and registered into the Metabolon LIMS. All compounds were analyzed at multiple concentrations under the conditions as the experimental samples, and the characteristics of each compound were registered into a LIMS-based library. The same library was used for both the LC and GC platforms for determination of their detectable characteristics. These were then analyzed using custom software packages. Initial data visualization used SAS and Spotfire.

Statistical Analysis
(refer to Supplementary Fig. 4 for outline)

a) Metabolomic Data

Data Imputation. The metabolic data is left censored due to thresholding of the mass spectrometer data. We imputed these missing values based on the average expression of the metabolite across all subjects. If the mean metabolite measure across samples was greater than 100,000, then zero was imputed, otherwise one half of the minimum measure for that sample was imputed. In this way, we hoped to distinguish which metabolites had missing data due to absence in the sample and which were missing due to instrument thresholds. We used sample minimums for the imputed values since the mass spectrometer threshold for detection may differ between samples and we wish to capture that threshold level.

Sample Normalization. To reduce between-sample variation we centered the imputed metabolic measures for each tissue sample on its median value and scaled it by its inter-quartile range (IQR).

Analysis:
z-score. This z-score analysis scaled each metabolite according to a reference distribution. Unless otherwise specified, the benign samples were designated as the reference distribution. Thus the mean and standard deviation of the benign samples was determined for each metabolite. Then each sample, regardless of diagnosis, was centered by the benign mean and scaled by the benign standard deviation, per metabolite. In this way, we can look at how the metabolite expressions deviate from the benign state.

Hierarchical Clustering. Hierarchical clustering was performed on the log transformed normalized data. A small value (unity) was added to each normalized value to allow log transformation. The log transformed data was median centered, per metabolite, prior to

clustering for better visualization. Pearson's correlation was used for the similarity metric. Clustering was performed using the Cluster program and visualized using Treeview [2]. A maize/blue color scheme was used in heat maps of the metabolites.

Comparative Tests.
To look at association of metabolite detection with diagnosis, we dichotomized the measures as present or absent (i.e., undetected). Chi-square tests were used to assess difference in rates of presence/absence of measurements for each metabolite between diagnosis groups. To assess the association between metabolite expression levels between diagnosis groups, two-tailed Wilcoxon rank sum tests were used for two-sample tests; benign vs. PCA, PCA vs. Mets. Kruskal-Wallis tests were used for three-way comparisons between all diagnosis groups; benign vs. PCA vs. Mets. Non-parametric tests were used reduce the influence of the imputed values. Tests were run per metabolite on those metabolites that had detectable expression in at least 20% of the samples. Significance was determined using permutation testing in which the sample labels were shuffled and the test was recomputed. This was repeated 1000 times. Tests in which the original statistic was more extreme than the permuted test statistic increased evidence against the null hypothesis of no difference between diagnosis groups. The threshold for significance was $P \leq 0.05$ for all tests. This threshold was not adjusted for multiple tests. False discovery rates (FDR) were determined from the permuted *P*-value using the q-value conversion algorithm of Storey et al[3] as implemented in the R package "q-value". Estimation of FDR does account for multiple testing.

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 analysis of variance (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*Treatment + E. In this way exp(B) is an estimate of (Y | Treatment = 1)/(Y | Treatment = 0) and the standard error of exp(B) can be estimated from SE(B) using the delta method.

Classification. Metabolites were added to classifiers based on increasing empirical *P*-value. Support vector machines (SVM) were used to determine an optimal classifier. Leave-one-out cross validation (LOOCV) was employed to estimate error rates among classifiers. To avoid bias, comparative tests to determine the empirical *P*-value ranking, were repeated for each leave-one-out sample set. SVM selected the optimal empirical *P*-value for inclusion in the classifier. Those metabolites that appeared in at least 80% of the LOOCV samples at or below the chosen empirical *P*-value were selected as the classification set. A principal components analysis was used to help visualize the separation provided by the resulting classification set of metabolites. Principal components one, two, and three were used for plotting.

Validation of Sarcosine in Urine. Urine sediment experiments were performed across three batches; batch-level variation was removed using two adjustments. First, two batches (n=15 and n=18) with available measurements on cell line controls DU145 and

RWPE were combined by estimating batch-level differences using only this cell line data in an ANOVA model with the log-transformed ratio of sarcosine to alanine as the response. The second adjustment put the resulting combined batches (n=33) together with the remaining third batch (n=60) by centering (by the median) and scaling (by the median absolute deviation) within each of these two batches.

Urine supernatant experiments measured sarcosine in relation to creatinine. Analysis was performed using a log base 2 scale to indicate fold change from creatinine. Urine sediments and supernatants were tested for differences between biopsy status using a two-tailed Wilcoxon rank-sum test. Associations with clinical parameters were assessed by Pearson correlation coefficients for continuous variables and two-tailed Wilcoxon rank-sum tests for categorical variables.

Mapping of "Omics" data to a common identifier. The metabolites profiled in this proposal were mapped to the metabolic maps in KEGG using their compound IDs, followed by identification of all the anabolic and catabolic enzymes in the mapped pathways. This was followed by retrieval of the official enzyme commission number (EC number) for the enzymes that were mapped to its official gene ID using KEGG's DBGET integrated data retrieval system **(eg:http://www.genome.jp/dbget-bin/www_bget?enzyme+2.4.1.1)**.

Enrichment of Molecular Concepts. In order to explore the network of inter-relationships among various molecular concepts and our integrated data (containing information from metabolome), we used the Oncomine Concepts Map (www.oncomine.org) bioinformatics tool developed by our group, [4,5]. In addition to being the largest collection of gene sets for association analysis, the Oncomine Concepts Map (OCM) is unique in that computes pair-wise associations among all gene sets in the database, allowing for the identification and visualization of "enrichment networks" of linked concepts. Integration with the OCM allows us to systematically link molecular signatures (i.e., in this case metabolomic signatures) to over 14,000 molecular concepts, confirming previous observations and generating novel hypotheses about the biological progression of prostate cancer. To study the enrichments resulting from the metabolomic data alone involved generation of a list of gene IDs from the metabolites that were significant with a *P*-value less than 0.05 for the comparisons being made. This signature was used to seed the analysis. On a similar note for gene expression-based enrichment analysis, we used gene IDs for transcripts that were significant ($p<0.05$) for the comparisons being made. Once seeded, each pair of molecular concepts was tested for association using Fisher's exact test. Each concept was then analyzed independently and the most significant concept reported. Results were stored if a given test had an odds ratio > 1.25 and *P*-value < 0.01. Adjustment for multiple comparisons was be made by computing q-values for all enrichment analyses. We are confident that the integrative analyses coupled to enrichment using OCM will generate a number of testable hypotheses on molecular events leading to development of cancer and its invasion in advanced disease. For the purpose of this study, all concepts that had a *P*-value less than $1X10^{-4}$ were considered significant. **Supplementary Fig. 9** shows the detailed outline of

the procedure followed for the process of molecular concept analyses for localized prostate cancer and metastatic cancer samples.

Chromatin Immunoprecipitation (ChIP) and ChIP-PCR. ChIP was carried out as previously described[6] using antibodies against AR (Millipore, #06-680), ERG (Santa Cruz, #sc354) and rabbit IgG (Santa Cruz, #sc-2027). For AR ChIP assays, VCaP cells were grown in phenol red-free medium with charcoal-stripped serum for hormone deprivation for 3 days, before treatment for 16 hr with 1% ethanol or 10nM of methyltrienolone (R1881, NEN Life Science Products) dissolved in ethanol. AR ChIP was performed in paired ethanol-treated and R1881-treated samples. ChIP-enriched chromatin as well as the whole-cell extract (WCE) was amplified by ligation-mediated PCR. When examining AR binding on target genomic regions, equal amount 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 analyzed by 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 **Supplemental Table 10.**

ChIP-Seq analysis. ChIP samples were prepared for sequencing using the Genomic DNA sample prep kit (Illumina) following manufacturers protocols. ChIP-Sequencing was performed using Illumina Genome Analyzer according to standard manufacturer's procedures. The raw sequencing image data were analyzed by the Illumina analysis pipeline, aligned to the unmasked human reference genome (NCBI v36, hg18) using the ELAND software (Illumina) to generate sequence reads of 25-32 bps. ChIP-Seq binding on target genes was visualized by UCSC genome browser.

Digital gene expression analysis by next-generation tag sequencing. LNCaP and VCaP prostate cancer cells were hormone deprived for 2 days prior to a time-course treatment of synthetic androgen (R1881) for 0, 3, 12, 24, 48 hrs. Total RNA was isolated using Trizol (Invitrogen). Samples with 0h and 48h androgen treatment were prepared for sequencing using the Digital Gene Expression-Tag Profiling with *NIaIII* kit and sequenced by Genome Analyzer (Illumina). 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 (TPM).

Quantitative RT-PCR. Q-PCR was performed using Power SYBR Green Mastermix (Applied Biosystems) on an Applied Biosystems 7300 Real Time PCR machine as previously described [6]. All primers were designed using Primer 3 and synthesized by Integrated DNA Technologies and are listed in **Supplementary Table 10**. All PCR experiments were performed in triplicates.

RNA interference. DU145 or RWPE cells were treated with non-targeting siRNA (D-001210-01, Dharmacon), or gene-specific siRNA to various targets as listed in **Supplementary Table 9.**

Cell invasion assay. Cell invasion was carried out using a modified basement membrane chamber assay as previously described[6]. Briefly, equal numbers of the indicated cells were seeded onto the basement membrane matrix (EC matrix, Chemicon) present in the insert of a 24-well culture plate, with fetal bovine serum added to the lower chamber as a chemo-attractant. After 48 h, non-invading cells and the EC matrix were removed by a cotton swab. Invaded cells were stained with crystal violet and photographed. The inserts were treated with 10% acetic acid and absorbance was measured at 560 nm.

Cell motility assay. For cell motility assay, RWPE cells were seeded in a 6-well plate and treated with either 50 μM sarcosine and alanine at 24 h intervals for a total duration of 96 hrs. Untreated RWPE cells were used as controls. Blue Flourescent microsphere beads[7] (Cellomics Cell Motility kit, Pierce Biotechnology, #K0800011) were added to 96-well collagen coated plate and allowed to attach for 1h at 37^{o}C. The unbound beads were removed by washing. Approximately, 200 untreated or amino acid treated cells in a total volume of 50 μl were carefully layered on the beads incubated for 12 hours at 37^{o}C in an atmosphere of 5% CO_2. Upon incubation, the cells were fixed and the cytoskeleton was stained with Rhodamine-conjugated Phalloidin as per manufacturer's instructions. Images of motility tracks were captured by fluorescent microscope using a DAPI filter.

SUPPLEMENTARY DISCUSSION

Background.
A number of groups have employed gene expression microarrays to profile prostate cancer tissues [8-16] as well as other tumors [17-20] at the transcriptome level. Similarly protein arrays and mass spectrometry-based proteomic profiling has been used to a more limited extent, to study alterations at the proteome level [21-27]. However, in contrast to genomics and proteomics, metabolomics (i.e., examining metabolites with a global, unbiased perspective) is an emerging science, and represents the distal read-out of the cellular state as well as associated pathophysiology. As part of a systems biology perspective, metabolomic profiling may be a useful complement to other "omics" approaches.

Technologies Used for Metabolomic Profiling. Multiple technologies have been used to profile metabolites in biological samples. These include high pressure liquid chromatography (HPLC), nuclear magnetic resonance (NMR)[28], mass spectrometry[29] (GC/MS and LC/MS) and Enzyme Linked Immuno Sorbent Assay (ELISA). Among these, HPLC identifies compounds based solely on their chromatographic retention time and is limited by the need for an external standard. NMR, although being sensitive and high throughput, is limited by the number of named compounds that can be defined. NMR has been used in a recent population-based study to link metabolic phenotypes to diet and blood pressure and delineate biomarkers for cardiovascular risk[30]. ELISA is limited by the need for specific antibodies, the generation of which is a challenge for small molecules. Mass spectrometry-based profiling, on the other hand, is sensitive and robust allowing for simultaneous identification of known metabolites as well as characterization of unknown compounds.

Metabolomic Profiling of Cancer
Using techniques described above and following a focused approach, most of the early studies on neoplastic metabolism have interrogated tumor adaptation to hypoxia [31,32]. These investigations revealed heterogeneity within the tumor constituted by varying gradients of metabolites (e.g., glucose or oxygen) and growth factors, which allow neoplastic cells to thrive under conditions of low oxygen tension [31]. Included among these targeted approaches are also studies that have implicated citrate and choline in the process of prostate cancer progression[33,34]. As an extension to these studies, multiple groups have addressed bioenergetic pathways associated with tumor progression and aggressivity[35,36,37]. More recently, interrogation of the metabolome using nuclear magnetic resonance [34,38-40] and gas chromatography, coupled with time-of-flight mass spectrometry, [41-43] have revealed the power of metabolomic signatures in classifying tumors. Despite this, the number of metabolites monitored in these studies is limited.

Metabolomic alterations in Plasma and Urine from biopsy-positive and biopsy-negative patients.
We sought to identify metabolites that were differential between biopsy positive and biopsy negative patients using their unbiased metabolomic profiles. To this end, two-sided Wilcoxon rank-sum tests were performed for each metabolite. One thousand permutations of the sample status labels were done to create a null distribution from

which an empirical p-value was determined. Moreover, for each of the metabolites that were missing in at least 80% of samples, we assessed whether missingness itself was differential through the use of chi-squared tests. A total of 478 and 583 metabolites were measured in plasma and urine respectively. Of the 478 well-measured plasma metabolites, only 20 (4.2%) had an empirical p-value <= 0.05 (FDR=98.8 %). Further, of the 76 poorly-measured plasma metabolites, only 3 (3.9%) had a chi-squared p-value <= 0.05 (FDR=87.7%, see **Supplementary Table 4** for list of metabolites). Similarly, Of the 583 well-measured urine metabolites, only 36 (6.2%) had an empirical p-value <= 0.05 (FDR=66.8 %). Moreover, of the 59 poorly-measured urine metabolites, only 1 (1.7%) had a chi-squared p-value <= 0.05 (FDR=69.9%, see **Supplementary Table 4** for list of metabolites).

Statistical Analysis of tissue-derived metabolomic profiles in prostate cancer progression. Three analyses were performed to provide a global perspective of the data. The first employed unsupervised hierarchical clustering on the normalized data (refer to the **Supplementary Fig. 4** and **Supplementary methods** for procedural details). This analysis separated the metastatic samples from both the benign and PCA tissues, but it did not accurately cluster the clinically localized prostate cancers from the benign prostates (**Supplementary Fig. 5a**). This indicated a higher degree of metabolomic alteration in the metastatic samples relative to benign and PCA specimens highlighted by the heat map representation of the data (**Fig. 1c,** see **Supplementary Fig. 5b** for the annotated heat map). This finding is consistent with our earlier observations based on gene expression analyses [5,8]. Further, this pattern of metabolomic alterations was shared across multiple metastatic samples derived from different sites of origin (**Supplementary Fig. 6**). Specifically the 28 metastatic samples used to validate sarcosine levels (**Fig. 3 a** and **Supplementary Fig. 12**) were derived from different tumor bearing organs to which prostate cancer metastasizes. These include liver, lung, mesentery, pancreas, lymph node etc. Further, with regards to the sarcosine levels in metastatic disease, we found elevated levels of this metabolite in differing metastatic sites in the patient irrespective of the site of tumor location. These findings indicate that the metastatic disease-specific metabolomic signature is independent of site of tumor origin.

In the second analysis, each metabolite was centered on the mean and scaled on the standard deviation of the normalized benign metabolite levels to create z-scores based on the distribution of the benign samples. **Fig. 1d** shows the 626 metabolites plotted on the vertical-axis, and the benign-based z-score for each sample plotted on the horizontal-axis for each class of sample. As illustrated by the figure, alterations in metabolites occur most robustly in metastatic tumors (z-score range: -13.6 to 81.9). In particular, there were 105 metabolites that had a z-score of two or greater in at least 33% of the metastatic samples analyzed. In contrast, the changes in clinically localized prostate cancer samples were less than in metastatic disease (z-score range: -7.7 to 45.8) such that only 38 metabolites had a z-score of two or greater in at least 33% of the samples.

To investigate the power of metabolomic profiles to discern the three classes namely benign, localized cancer and metastatic disease, we used a support vector machine (SVM) classification algorithm with leave-one-out cross-validation (see Supplementary

Methods). This predictor correctly identified all of the benign and metastatic samples, with misclassification of 2/12 PCA samples as benign. Interestingly, the two misclassified cancer samples had a low Gleason grade of 3+3, which suggested less aggressive tumors. In addition, we tabulated a list of 198 metabolites that were significant at a P=0.05 level in at least 80% of the leave-one-out cross-validated datasets. (See **Supplementary Table 5** for the list of 198 metabolites). For visualization, principal component analysis was employed on this data matrix of 198 metabolites (**Supplementary Fig.7**). Similar to the SVM results, principal component analysis delineated the three classes using only the first three components.

Enrichment analysis of class-specific metabolomic profiles using Oncomine Concept Maps (OCM)

Class-specific coordinated metabolite patterns were examined using the bioinformatics tool, Oncomine Concept Maps (OCM, www.oncomine.org,) that permitted systematic linkages of metabolomic signatures to molecular concepts, generating novel hypotheses about the biological progression of prostate cancer (refer to **Supplementary Fig. 9** for an outline of the analyses for localized prostate cancer and metastatic prostate cancer and to the supplemental methods for a description of OCM) [4]. Consistent with the KEGG analysis, the Oncomine analysis expanded upon this theme and (**Supplementary Fig. 10a**, blue node) identified an enrichment network of amino acid metabolism in these specimens (**Supplementary Fig. 10a** , red edges). These included the most enriched GO Biological processes; amino acid metabolism (P = 6 X 10^{-13}) and KEGG pathway for glutamate metabolism (P = 6.1 X 10^{-24}). KEGG pathways for glycine, serine and threonine metabolism (P = 2.8 X 10^{-14}), alanine and aspartate metabolism (P= 3.3 X 10^{-11}), arginine and proline metabolism (P = 2.3 X 10^{-10}) and urea cycle and metabolism of amino groups (P= 1.7 X 10^{-6}) also showed strong enrichment.

Importantly, the metabolomic profiles for compounds "over-expressed in metastatic samples" (**Supplementary Fig. 10b**, blue node) showed strong enrichment for elevated methyltransferase activity (**Supplementary Fig. 10b**, red edges). This increased methylation potential was supported by multiple enrichments of S-adenosyl methionine (SAM) mediated methyltransferase activity including: the enriched InterPro concept, SAM binding motif (P = 1.1 X 10^{-11}) and GO Molecular Function, methyltransferase activity (P = 7.7 X 10^{-8}). These enrichments were a result of significant elevation in the levels of S-adenosyl methionine (P = 0.004) in the metastatic samples compared to the PCA samples. The resulting enhancement in the methylation potential of the tumor was further supported by additional concepts that described increased chromatin modification (GO Biological Process, P = 2.9 X 10^{-6}), involvement of SET domain containing proteins (InterPro, P = 7.4 X 10^{-7}) and histone-lysine N-methyltransferase activity (GO Molecular Function, P = 6.3 X 10^{-6}) in the metastatic samples (**Supplementary Fig. 10b**, red edges).

Sarcosine, a marker of cancer progression and change in metabolic activity. Based on our profiling data, we believe that changes in metabolic activity and cancer progression are highly interrelated events. Importantly, changes in the levels of sarcosine reflect the inherent changes in the biochemistry of the tumor as it develops and

progresses to a more advanced state. This is evident from our data wherein cancer progression has been shown to be associated with an increase in amino acid metabolism and methylation potential of the tumor. Furthermore, one of the factors leading to an increased methylation potential is the increase in levels of S-adenosyl methionine (SAM) and its pathway components during tumor progression. This translates into elevated levels of methylated metabolites like N-methyl-glycine (sarcosine), methyl-guanosine, methyl-adenosine (known markers of DNA methylation) etc. in tumors compared to their benign counterparts. Notably, one of the major pathways for sarcosine generation involves the transfer of the methyl group from SAM to glycine, a reaction catalyzed by glycine-N-methyl transferase (GNMT). Using siRNA directed against GNMT, we have shown that sarcosine generation is important for the cell invasion process. In contrast to GNMT, knock-down of SARDH in benign prostate epithelial cells (RWPE) resulted in induction of an invasive phenotype. Also, our data reveals androgen receptor and the ERG gene fusion product, two key elements in prostate cancer progression, coordinately regulate components of the sarcosine pathway. These support our notion that elevated levels of sarcosine are a result of a change in the tumor's metabolic activity that is closely associated with the process of tumor progression. It is worth noting that sarcosine produced from tumor progression-associated changes in metabolic activity, by itself promotes tumor invasion. This supports our hypothesis that the process of tumor progression and changes in metabolic activity are intimately connected.

Biological significance of sarcosine elevation during prostate cancer progression. Biologically, our data suggests that sarcosine levels are reflective of two important hallmarks associated with prostate cancer development; namely increased amino acid metabolism and enhanced methylation potential leading to epigenetic silencing. The former is evident from the metabolomic profiles of localized prostate cancer that show high levels of multiple amino acids. This is also well corroborated by gene expression studies [5] that describe increased protein biosynthesis in indolent tumors. Increased methylation has been known to play a major role in epigenetic silencing. We and others have shown increased levels of EZH2, a methyltransferase belonging to the polycomb complex, in aggressive prostate cancer and metastatic disease[44,45]. The current study expands our understanding in this realm by implicating tumor progression to be associated with elevated methylation potential. This is supported by the finding of elevated levels of S-adenosyl methionine (the major methylation currency of the cell) and its associated pathway components during prostate cancer progression. This is further reflected by elevated levels of methylated metabolites in our dataset. Included among these is the methylated derivative of glycine (i.e., sarcosine) that shows a progressive elevation in its levels from benign to localized tumor to metastatic disease. Notably, one of the major pathways for sarcosine generation involves the methylation reaction wherein the enzyme glycine-N-methyltransferase catalyses the transfer of methyl groups from SAM to glycine (an essential amino acid). Thus elevated levels of sarcosine can be attributed to an increase in both amino acid levels (in this case glycine) and an increase in methylation, both of which form the hallmarks of prostate cancer progression. Further, our finding of coordinate regulation of sarcosine pathway by both androgen receptor and the ERG gene fusion reveals the possible mechanism behind sarcosine elevation in this disease.

Assessment of androgen regulation of glycine-N-methyltransferae (GNMT) and Sarcosine dehydrogenase (SARDH) using Digital Gene Expression (DGE).

To determine whether GNMT and SARDH are differentially regulated by androgen, we examined the DGE (digital gene expression) dataset for LNCaP and VCaP cells following androgen stimulation for 48 h. The expression for these genes were measured as the number of transcripts per million of total sequencing reads (TPM). GNMT was highly elevated in both VCaP and LNCaP cells (7.2 and 102 TPM respectively) in response to androgen compared to alcohol-treated controls (0.9 and 2.1 TPM respectively). In contrast androgen stimulation weakly repressed SARDH levels in VCaP cells (androgen vs control: 0.6 vs 0.9 TPM) while having no effect in LNCaP cells (androgen vs control: 1.41 vs 1.42 TPM).

Validation of additional prostate-cancer specific metabolites in independent clinical specimens.

We also validated the increase of five additional metabolites in 52 prostate-derived samples that were a subset of the 89-sample cohort used for sarcosine validation studies described in **Fig 3a**. These metabolites were assessed using targeted mass-spectrometric assays. As shown in **Supplementary Fig. 13,** levels of cysteine, glutamic acid, glycine and thymine were all elevated upon progression from benign to localized prostate cancer and advancement to metastatic disease, while citrate levels were reduced upon disease progression.

SUPPLEMENTARY REFERENCES

1. Lawton, K.A., *et al.* Analysis of the adult human plasma metabolome. *Pharmacogenomics* **9**, 383-397 (2008).
2. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14863-14868 (1998).
3. Storey, J. *J Royal Stat Soc* **64**, 479 (2002).
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. *Nat Genet* **39**, 41-51 (2007).
6. Yu, J., *et al.* Integrative genomics analysis reveals silencing of beta-adrenergic signaling by polycomb in prostate cancer. *Cancer Cell* **12**, 419-431 (2007).
7. Klemke, R.L., *et al.* Regulation of cell motility by mitogen-activated protein kinase. *The Journal of cell biology* **137**, 481-492 (1997).
8. Dhanasekaran, S.M., *et al.* Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**, 822-826. (2001).
9. Lapointe, J., *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 811-816 (2004).
10. LaTulippe, E., *et al.* Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer research* **62**, 4499-4506 (2002).
11. Luo, J., *et al.* Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. *Cancer research* **61**, 4683-4688. (2001).
12. Luo, J.H., *et al.* Gene expression analysis of prostate cancers. *Mol Carcinog* **33**, 25-35. (2002).
13. Magee, J.A., *et al.* Expression Profiling Reveals Hepsin Overexpression in Prostate Cancer. *Cancer research* **61**, 5692-5696. (2001).
14. Singh, D., *et al.* Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1**, 203-209. (2002).
15. Welsh, J.B., *et al.* Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer research* **61**, 5974-5978. (2001).
16. Yu, Y.P., *et al.* Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J Clin Oncol* **22**, 2790-2799 (2004).
17. Golub, T.R., *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science (New York, N.Y* **286**, 531-537 (1999).
18. Hedenfalk, I., *et al.* Gene-expression profiles in Hereditary Breast Cancer. *The New England Journal of Medicine* **344**, 539-548 (2001).
19. Perou, C.M., *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747-752 (2000).

20. Alizadeh, A.A., *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling [see comments]. *Nature* **403**, 503-511 (2000).
21. Ahram, M., *et al.* Proteomic analysis of human prostate cancer. *Mol Carcinog* **33**, 9-15 (2002).
22. Hood, B.L., *et al.* Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol Cell Proteomics* **4**, 1741-1753 (2005).
23. Prieto, D.A., *et al.* Liquid Tissue: proteomic profiling of formalin-fixed tissues. *Biotechniques* **Suppl**, 32-35 (2005).
24. Varambally, S., *et al.* Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* **8**, 393-406 (2005).
25. Martin, D.B., *et al.* Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. *Cancer research* **64**, 347-355 (2004).
26. Wright, M.E., Han, D.K. & Aebersold, R. Mass spectrometry-based expression profiling of clinical prostate cancer. *Mol Cell Proteomics* **4**, 545-554 (2005).
27. Cheung, P.K., *et al.* Protein profiling of microdissected prostate tissue links growth differentiation factor 15 to prostate carcinogenesis. *Cancer research* **64**, 5929-5933 (2004).
28. Brindle, K.M., Fulton, S.M., Gillham, H. & Williams, S.P. Studies of metabolic control using NMR and molecular genetics. *J Mol Recognit* **10**, 182-187 (1997).
29. Gates, S.C. & Sweeley, C.C. Quantitative metabolic profiling based on gas chromatography. *Clin Chem* **24**, 1663-1673 (1978).
30. Holmes, E., *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **453**, 396-400 (2008).
31. Dang, C.V. & Semenza, G.L. Oncogenic alterations of metabolism. *Trends Biochem Sci* **24**, 68-72 (1999).
32. Kress, S., *et al.* Expression of hypoxia-inducible genes in tumor cells. *J Cancer Res Clin Oncol* **124**, 315-320 (1998).
33. Mueller-Lisse, U.G., Swanson, M.G., Vigneron, D.B. & Kurhanewicz, J. Magnetic resonance spectroscopy in patients with locally confined prostate cancer: association of prostatic citrate and metabolic atrophy with time on hormone deprivation therapy, PSA level, and biopsy Gleason score. *Eur Radiol* **17**, 371-378 (2007).
34. Wu, C.L., *et al.* Proton high-resolution magic angle spinning NMR analysis of fresh and previously frozen tissue of human prostate. *Magn Reson Med* **50**, 1307-1311 (2003).
35. Vizan, P., *et al.* K-ras codon-specific mutations produce distinctive metabolic phenotypes in NIH3T3 mice [corrected] fibroblasts. *Cancer research* **65**, 5512-5515 (2005).
36. Al-Saffar, N.M., *et al.* Noninvasive magnetic resonance spectroscopic pharmacodynamic markers of the choline kinase inhibitor MN58b in human carcinoma models. *Cancer research* **66**, 427-434 (2006).
37. Ramanathan, A., Wang, C. & Schreiber, S.L. Perturbational profiling of a cell-line model of tumorigenesis by using metabolic measurements. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5992-5997 (2005).

38. Cheng, L.L., *et al.* Metabolic characterization of human prostate cancer with tissue magnetic resonance spectroscopy. *Cancer research* **65**, 3030-3034 (2005).
39. Burns, M.A., *et al.* Reduction of spinning sidebands in proton NMR of human prostate tissue with slow high-resolution magic angle spinning. *Magn Reson Med* **54**, 34-42 (2005).
40. Kurhanewicz, J., Swanson, M.G., Nelson, S.J. & Vigneron, D.B. Combined magnetic resonance imaging and spectroscopic imaging approach to molecular imaging of prostate cancer. *J Magn Reson Imaging* **16**, 451-463 (2002).
41. Denkert, C., *et al.* Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer research* **66**, 10795-10804 (2006).
42. Ippolito, J.E., *et al.* An integrated functional genomics and metabolomics approach for defining poor prognosis in human neuroendocrine cancers. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9901-9906 (2005).
43. Denkert, C., *et al.* Metabolite profiling of human colon carcinoma - deregulation of TCA cycle and amino acid turnover. *Molecular cancer* **7**, 72 (2008).
44. Varambally, S., *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624-629. (2002).
45. Varambally, S., *et al.* Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science (New York, N.Y* **322**, 1695-1699 (2008).

SUPPLEMENTARY TABLES

Supplementary Table 1. List of standards used during process evaluation of metabolomic profiling.

Standard	Description	Purpose
MTRX	Large pool of human plasma maintained at Metabolon, characterized extensively	Assure all aspects of profiling process are within specifications
CMTRX	Pool created using a small aliquot from each customer sample	Assess effect of matrix on profiling process; distinguish biological- from process variability
PRCS	Aliquot of ultra-pure water	Process blank to assess contribution to compound signals from process
SOLV	Aliquot of extraction solvents	Solvent blank used to segregate contamination sources in extraction
DS	Derivatization Standard	Assess variability of derivatization for GC/MS samples
IS	Internal Standard	Assess variability/performance of instrument
RS	Recovery Standard	Assess variability; verify performance of extraction/instrumentation

Supplementary Table 2: Clinical information associated with tissue specimens used for metabolomic profiling

Characteristic	Value[+]
Benign: Benign adjacent prostate tissues from patients with prostate cancer	
No. of patients	16[*]
Age at biopsy (years)	56 ± 6.7 [40, 63]
Race	
White(non-Hispanic origin)	12 (92.3%)
Other	1 (7.7%)
PCA: Patients with clinically localized prostate cancer	
No. of patients	11[*]
Age at biopsy (years)	57 ± 7.7 [40, 63]
Sample Gleason Grade (minor + major)	
3 + 3	3 (25%)
3 + 4	5 (41.7%)
4 + 3	3 (25%)
4 + 4	1 (8.3%)
Baseline PSA	10.4 ± 8.1 [2.4, 24.6]
Stage	
T2a	3 (30%)
T2b	4 (40%)
T3a	2 (20%)
T3b	0 (0%)
T4	1 (10%)
Race	
White (non-Hispanic origin) (%)	8 (80%)
Other (%)	2 (20%)
Mets: Patients with metastatic prostate cancer.	
No. of patients	13[*]
Age at death (years)	66 ± 12.1 [40, 82]
Sample Location	
Soft tissue	4 (28.6%)
Liver	8 (57.1%)
Rib	1 (7.1%)
Diaphragm	1 (7.1%)
Race	
White (non-Hispanic origin) (%)	13 (100%)

[+] For continuous variables the Mean ± SD [range] is given. Count and percentage is given for categorical variables.

*There are 16 benign tissue samples from 16 men. Clinical information is available for 13 of these men. There are 12 local prostate cancer tumor samples from 11 men. Clinical information is available for 10 of these men. There are 14 metastatic tumor samples from 13 men. Clinical information is available on all 13 men.

Supplementary Table 3. Clinical information associated with biofluid specimens (urine, plasma) used for metabolomic profiling

Characteristic	Value[+]
Negative: Prostate biopsy found no evidence of cancer	
No. of patients	51
Age at biopsy (years)	61 ± 9.6 [40, 80]
Baseline PSA [*]	6.1 ± 3.4 [0.8, 20.8]
Race	
White (non-Hispanic origin)	25 (49.1%)
Other	3 (5.9%)
Unreported	23 (45.1%)
Positive: Prostate cancer was detected upon biopsy	
No. of patients	59
Age at prostatectomy (years)	57 ± 7.7 [40, 63]
Baseline PSA [*]	10.4 ± 8.1 [2.4, 24.6]
Sample Gleason Grade (minor + major)	
3 + 3	25 (42.4%)
3 + 4	14 (23.7%)
4 + 3	11 (18.6%)
4 + 4	3 (5.1%)
4 + 5	5 (8.5%)
5 + 5	1 (1.7%)
Race	
White (non-Hispanic origin) (%)	26 (44.1%)
Other (%)	4 (6.8%)
Unreported	29 (49.1%)

[+] For continuous variables the Mean ± SD [range] is given. Count and percentage is given for categorical variables.

* PSA data is available for 45 of the men with a negative biopsy and for 55 of the men with positive biopsy.

Supplementary Table 4. List of named metabolites and isobars measured in the three biospecimens (T=tissue, U=urine, P=plasma) using either liquid chromatography (LC) or gas phase chromatography (GC) coupled to mass spectrometry.

Biospecimen			MS	Biochemical
	U		LC	(1'R,1'S) Biopterin
T		P	GC	1,5-Anhydroglucitol (1,5-AG)
		P	LC	1-Methyladenine
T	U		LC	1-Methyladenosine (1mA)
	U		LC	2,3-Dihydroxybenzoate
T			GC	2-Aminoadipate
	U		LC	2-Deoxyadenosine
T			LC	2'-Deoxyuridine-5'-triphosphate (dUTP)
T	U		LC	2-Hydroxybutyrate (AHB)
T		P	GC	2-Hydroxybutyrate (AHB)
	U		GC	2-Isopropylmalate
	U		LC	2-Methylhippurate
T			GC	3,4-Dihydroxyphenylethyleneglycol (DOPEG)
	U		GC	3-Hydroxy-3-methylglutarate
	U	P	LC	3-Hydroxybenzoate
	U	P	LC	3-Hydroxyphenylacetate
	U	P	GC_U/LC_P	3-Methoxy-4-Hydroxyphenylacetate
T	U	P	LC	3-Methyl-2-oxopentanoate
T	U	P	LC	3-Methylhistidine (1-Methylhistidine)
	U		LC	3-Nitro-L-histidine
T		P	GC_T/LC_P	3-Phosphoglycerate
	U		LC	3-ureidopropionate
T	U	P	LC	4-Acetamidobutanoate
	U		LC	4-Acetominophen sulfate
		P	LC	4-Amino-5-methyl-2(1H)-pyrimidinone
T	U	P	LC	4-Guanidinobutanoate
	U		GC	4-Hydroxy-3-methoxymandelate
	U		LC	4-Hydroxybenzyl alcohol
	U		GC	4-Hydroxymandelate
T		P	GC_P/LC_T	4-Methyl-2-oxopentanoate
T	U		LC	5,6-Dihydrothymine
T	U		GC	5,6-Dihydrouracil
T			GC	5-Hydroxyindoleacetate (5-HIA)
T	U		LC	5-Hydroxytryptophan
T	U		LC	5-Methylthioadenosine (MTA)
T			LC	5-Sulfosalicylate
T			LC	6-Phosphogluconate
	U	P	LC	7,8-Dihydrofolate
	U		LC	7,8-Dihydroneopterin
T	U	P	LC	Acetylcarnitine (ALCAR; C2 AC)
T	U		GC	Aconitate
T	U	P	GC_{TU}/LC_P	Adenine
T	U		LC	Adenosine
	U		LC	Adenosine 3,5-cyclic monophosphate (cAMP)
		P	LC	Adenosine 5-monophosphate (AMP)

	U		LC	Agmatine
	U	P	LC	a-Hydroxybenzeneacetate
T	U	P	LC	a-Ketoglutarate
T	U	P	GC	Alanine
T	U	P	LC	Alanylalanine
	U	P	GC	Allantoin
	U		GC	Arabinose
	U		GC	Arabitol
T		P	GC	Arachidonate (20:4n6)
	U		GC	Arginine
T			LC	Argininosuccinate
T	U		GC	Ascorbate (Vitamin C)
	U		GC	a-Sorbopyranose
T	U		GC	Asparagine
T			GC	Aspartate
T	U	P	LC	Assymetric Dimethylarginine (ADMA)
T		P	GC	a-Tocopherol
T	U	P	LC	Azelate (Nonanedioate)
T	U		GC	b-Alanine
T	U		GC	b-Aminoisobutyrate
	U		LC	Benzoate
T	U	P	GC	b-Hydroxybutyrate (BHBA)
	U		GC	b-Hydroxyisovalerate
	U		LC	b-Hydroxyphenylethylamine
		P	LC	b-Hydroxypyruvate
T			LC	Bicine
T		P	LC	Biliverdin
T	U		LC	Biotin
T			LC	Bradykinin
T	U	P	LC	Bradykinin hydroxyproline
T			GC	Cadaverine
T	U	P	LC	Caffeine
T	U	P	LC	Carnitine
T	U	P	LC	Catechol
T		P	GC	Cholesterol
T	U		LC	Ciliatine (2-Aminoethylphosphonate)
T	U	P	GC	Citrate
T	U	P	GC	Citrulline
	U		LC	Cortodoxone
T	U	P	LC	Creatinine
T	U		GC	Cystathionine
T	U		GC	Cysteine
T			LC	Cytidine
T			LC	Cytidine monophosphate (CMP)
		P	LC	Cytosine
T	U		LC	Dehydroepiandrosterone sulfate (DHEA-S)
T	U		LC	Deoxyuridine
	U		LC	Dethiobiotin
	U		LC	Diaminopimelate
T			LC	Dihydroxyacetonephosphate (DHAP)

T			GC	Dimethylbenzimidazole
T	U	P	GC	Erythritol
T	U		LC	Ethylmalonate
		P	LC	Folate
T	U	P	GC	Fructose
T			GC	Fructose-6-phosphate
		P	GC	Fucose
T		P	GC	Fumarate (trans-Butenedioate)
	U		GC	Galactose
T	U		LC	g-Glutamylcysteine
		P	LC	g-Glutamylglutamate
T		P	LC	g-Glutamylglutamine
		P	LC	g-Glutamylleucine
	U	P	LC	g-Glutamyltyrosine
	U		GC	Gluconate
T	U	P	GC	Glucose
T	U	P	GC	Glutamate
T	U	P	GC	Glutamine
T			LC	Glutarate (Pentanedioate)
T			LC	Glutathione reduced (GSH)
T	U	P	GC	Glycerate
T		P	GC	Glycerol
T	U	P	GC	Glycerol-3-phosphate (G3P)
T		P	LC	Glycerophosphorylcholine (GPC)
T	U	P	GC	Glycine
T		P	LC	Glycocholate (GCA)
		P	LC	Glycoproline
	U	P	LC	Guanidineacetate
T	U		GC	Guanine
T	U		LC	Guanosine
		P	LC	Guanosine 5-diphosphate (GDP)
	U		GC	Guanosine 5-diphosphofucose
T		P	GC	Heptadecanoate (Margarate; 17:0)
	U	P	LC	Heptanedioate (Pimelate)
T	U	P	LC	Hexanoylcarnitine (C6 AC)
T	U	P	LC	Hippurate (Benzoylglycine)
T	U	P	LC	Histamine
T	U	P	GC	Histidine
T			LC	Histidinol
	U	P	LC	Homocitrulline
T	U	P	LC	Homocysteine
	U		GC	Homogentisate
T			LC	Homoserine lactone
	U		GC	Hydroxyacetate
	U		LC	Hydroxynicotinate
T			LC	Hydroxyphenylpyruvate
T	U	P	GC	Hydroxyproline
T			GC	Hypotaurine
T	U	P	LC	Hypoxanthine
T	U		GC	Imidazolelactate

	U	P	GC	Iminodiacetate
T	U	P	GC_P/LC_{TU}	Indolelactate
T	U	P	LC	Indoxylsulfate
T		P	LC	Inosine
T			GC	Inositol-1-phosphate (I1P)
T	U	P	GC	Isoleucine
T	U		GC_U/LC_T	Kynurenate
T	U	P	LC	Kynurenine
T	U	P	GC	Lactate
	U		GC	L-Allo-threonine
T		P	GC	Laurate (12:0)
T	U	P	GC	Leucine
T		P	GC	Linoleate (18:2n6)
T	U	P	GC	Lysine
T	U	P	GC_T/LC_{UP}	Malate
	U		GC	Maltose
	U		GC	Mannitol
T	U	P	GC	Mannose
T	U		GC	Mannose-6-phosphate
T	U	P	LC	Methionine
	U	P	LC	Methyl indole-3-acetate
	U		LC	Methyldopa
T	U	P	LC	Methylglutarate
T	U	P	GC	myo-Inositol
T		P	GC	Myristate (14:0)
T	U	P	LC	N-6-trimethyllysine
	U		LC	N-Acetylalanine
T	U	P	LC	N-Acetylaspartate (NAA)
T			GC	N-Acetylgalactosamine
T	U		GC	N-Acetylglucosamine
T			GC	N-Acetylglucosaminylamine
	U		GC	N-Acetylglutamate
	U	P	LC	N-Acetylleucine
T	U		GC_U/LC_T	N-Acetylneuraminate
	U		LC	N-Acetylserotonin
	U	P	LC	N-Acetylvaline
T			LC	N-Carbamoylaspartate
	U		LC	N-Formyl-methionine
	U		GC	Niacin (Vitamin B3)
T	U	P	LC	Nicotinamide
T			LC	Nicotinamide adenine dinucleotide (NAD+)
T	U		LC	Nicotinamide Ribonucleotide (NMN)
		P	GC	Nonate
	U		GC	Noradrenaline
	U		GC	Normetanephrine
	U		LC	N-Tigloylglycine
T		P	GC	Octadecanoate
T			LC	Ofloxacin
T		P	GC	Oleate (18:1n9)
T	U	P	GC	Ornithine

	U		GC	Orotate (Pyrimidinecarboxylate)
T			LC	Orotidine-5'-phosphate
T	U	P	GC	Orthophosphate (Pi)
T	U	P	LC	Oxalate (Ethanedioate)
T	U	P	GC	Oxoproline
	U		GC	p-Acetamidophenyl-beta-D-glucuronide
T		P	GC	Palmitate (16:0)
T		P	GC	Palmitoleate (16:1n7)
	U		LC	p-Aminobenzoate
T	U	P	LC	Pantothenate
T	U	P	LC	Paraxanthine
T	U	P	LC	Phenylalanine
		P	LC	Phenyllactate
T	U	P	GC_T/LC_{UP}	Phosphoenolpyruvate (PEP)
T	U		GC	Phosphoethanolamine
T		P	LC	Phosphoserine
		P	LC	P-Hydroxybenzaldehyde
T	U	P	GC_{TU}/LC_P	p-Hydroxyphenylacetate (HPA)
T	U	P	GC	p-Hydroxyphenyllactate (HPLA)
T			LC	Picolinate
T	U	P	LC	Pipecolate
	U		LC	Porpobilinogen
T	U	P	GC	Proline
T			GC	Putrescine
	U	P	LC	Pyridoxal
	U	P	LC	Pyridoxal Phosphate (PLP, Vitamin B6)
T	U	P	LC	Pyridoxamine
T	U	P	GC_{TU}/LC_P	Pyrophosphate (PPi)
		P	GC	Quinate
T	U		GC_U/LC_T	Quinolinate
T	U	P	LC	Riboflavin (Vitamin B2)
T			GC	Ribose
T	U		LC	S-Adenosylhomocysteine (SAH)
T	U		LC	S-Adenosylmethionine (SAM)
		P	GC	Salicylicate
	U	P	LC	Salicylurate
T			GC	Sarcosine (N-Methylglycine)
T	U	P	GC	Serine
	U		LC	Serotonin
	U		LC	Shikimate
T		P	GC_T/LC_P	Sorbitol
T			GC	Spermidine
T	U		GC	Spermine
T	U		LC	Suberate (Octanedioate)
T	U	P	GC	Succinate
	U	P	GC	Sucrose
T			GC	Sucrose/Maltose
T	U	P	LC	Tartarate
T			LC	Taurine
	U		LC	Taurocholate

	U		LC	Thiaminie
T	U	P	GC	Threonine
	U		LC	Thymidine
T	U		GC_T/LC_U	Thymine
T			LC	Thyroxine
T			LC	Topiramate
T		P	LC	trans-2, 3, 4-Trimethoxycinnamate
		P	LC	Trhomoxane B2
T	U	P	GC_P/LC_{TU}	Tryptophan
T	U	P	GC_{UP}/LC_T	Tyrosine
T			LC	UDP-N-acetylmuraminate (UDP-MurNAc)
T	U		GC	Uracil
T	U	P	LC	Urate
T	U	P	GC	Urea
T			LC	Uridine
	U		LC	Urocanate
T	U	P	GC	Valine
T	U	P	LC	Xanthine
T			LC	Xanthosine
	U		GC	Xanthurenate
T	U	P	GC_{TU}/LC_P	Xylitol
	U		GC	Xylose
T		P	LC	Isobar Glycochenodeoxycholate, Glycodeoxycholate
T			LC	Isobar includes 1-Kestose, Maltotriose, Melezitose
T	U	P	LC	Isobar includes 2-Aminoisobutyrate,3-Amino-isobutyrate
	U	P	LC	Isobar includes 5-Keto-D-gluconate, 2-Keto-L-gluconate
T	U	P	LC	Isobar includes Arginine, N-Acetyl-ornithine
		P	LC	Isobar includes Asparagine, Ornithine, Glycyl-glycine
T	U		LC	Isobar includes D-Arabinose 5-phosphate, D-Ribulose 5-phosphate
T		P	LC	Isobar includes D-Fructose 1-phosphate, b-Fructose 6-phosphate
T	U	P	LC	Isobar includes D-saccharate,1,5-anhydro-D-glucitol
		P	LC	Isobar includes D-Sorbitol 6-phosphate, Mannitol-1-phosphate
		P	LC	Isobar includes Fumarate, 3-Methyl-2-oxobutanoate
		P	LC	Isobar includes Galactinol dihydrate, Turanose, Kojibiose
T		P	LC	Isobar includes g-Aminobutyryl-L-histidine, L-Anserine
T	U	P	LC	Isobar includes Gluconate, DL-Arabinose, D-Ribose, L-Xylose
T		P	LC	Isobar includes Glutamate, O-Acetyl-L-serine
		P	LC	Isobar includes Hydrocinnamate, 2-Phenylpropionate
T			LC	Isobar includes Inositol 1-phosphate, Mannose 6-phosphate
T			LC	Isobar includes L-Arabitol, Adonitol
		P	LC	Isobar includes L-Gulono-1,4-lactone, Glucono-g-lactone
T		P	LC	Isobar includes L-Threonine, L-Allothreonine, L-Homoserine
T		P	LC	Isobar includes Maltotetraose, Stachyose
T			LC	Isobar includes Maltotetraose, Stachyose
		P	LC	Isobar includes Mannitol, Dulcitol
T	U	P	LC	Isobar includes Mannose, Fructose, Glucose, Galactose
T			LC	Isobar includes N-Acetyl-D-glucosamine, N-Acetyl-D-mannosamin
	U	P	LC	Isobar includes N-Acetyl-L-methionine
T	U		LC	Isobar includes R,S-Hydroorotate, 5,6-Dihydroorotate
T	U	P	LC	Isobar includes Valine, Betaine

Supplementary Table 5: List of 198 metabolites that make up the three-class-predictor derived from LOOCV. Metabolite contributions to the first three principal components (PC) as shown in **Supplementary Fig. 6.**

Biochemical	Permuted P-value	LOOCV Freq	PC1	PC2	PC3
1,5-anhydroglucitol (1,5-AG)	<0.001	100.0%	0.40	1.29	-0.10
1-Methyladenosine (1mA)	<0.001	100.0%	-3.45	-1.06	0.06
2-Hydroxybutyrate (AHB)	<0.001	100.0%	-3.89	-0.54	0.10
2-Hydroxybutyrate (AHB)	<0.001	100.0%	0.59	-1.61	0.25
4-Acetamidobutanoate	<0.001	100.0%	-4.52	-0.44	0.09
5-Hydroxyindoleacetate (5-HIA)	0.002	100.0%	-3.90	0.26	-0.58
Adenosine	<0.001	100.0%	-1.02	1.36	0.55
Arachidonate (20:4n6)	0.005	100.0%	-0.24	-0.85	-0.12
Aspartate	0.001	100.0%	10.42	0.38	-0.25
Assymetric Dimethylarginine (ADMA)	0.001	100.0%	-3.39	-0.65	-0.36
b-aminoisobutyrate	<0.001	100.0%	0.33	-3.29	0.62
Bicine	<0.001	100.0%	-3.03	-2.03	0.05
Biliverdin	0.003	83.3%	6.31	1.38	1.05
Bradykinin hydroxyproline	<0.001	100.0%	-2.97	0.46	-0.02
Caffeine	0.007	97.6%	-3.74	0.68	0.07
Catechol	<0.001	100.0%	-4.36	-0.46	-0.02
Ciliatine (2-Aminoethylphosphonate)	<0.001	100.0%	-1.12	1.65	0.05
Citrate	<0.001	100.0%	14.72	10.48	-4.24
Creatinine	0.008	85.7%	3.68	0.11	0.17
Cysteine	<0.001	100.0%	-2.01	-1.56	-1.86
Dehydroepiandrosterone sulfate (DHEA-S)	<0.001	100.0%	-4.07	0.40	-0.21
Erythritol	<0.001	100.0%	-3.34	-0.72	0.09
Ethylmalonate	<0.001	100.0%	-4.31	-0.47	-0.01
Fumarate (trans-Butenedioate)	0.004	100.0%	-1.48	-0.76	0.28
g-Glutamylglutamine	<0.001	100.0%	-3.16	0.82	0.04
Glutamate	0.01	85.7%	20.83	-0.92	0.21
Glutathione reduced (GSH)	<0.001	100.0%	7.31	5.70	4.66
Glycerol	<0.001	100.0%	14.20	-2.69	0.72
Glycerol-3-phosphate (G3P)	<0.001	100.0%	3.97	-2.53	0.10
Glycine	0.008	97.6%	17.79	-1.01	-0.08
Glycocholate (GCA)	0.002	100.0%	-3.99	-0.70	0.00
Guanosine	<0.001	100.0%	3.77	2.15	0.27
Heptadecanoate (Margarate; 17:0)	<0.001	100.0%	-3.96	-0.30	-0.11
Hexanoylcarnitine (C6 AC)	<0.001	100.0%	-2.24	0.86	-0.11
Histamine	0.003	100.0%	-2.24	0.48	0.16
Histidine	0.002	100.0%	6.40	-0.94	-0.12
Homocysteine	<0.001	100.0%	-2.68	-0.56	-0.15
Homoserine lactone	0.001	100.0%	1.36	0.53	-0.08
Hydroxyphenylpyruvate	<0.001	100.0%	-4.31	0.27	-0.03
Inosine	<0.001	100.0%	8.03	2.66	0.10
Inositol-1-phosphate (I1P)	<0.001	100.0%	-2.86	-0.44	0.18
Kynurenate	0.001	100.0%	3.16	-0.32	-0.61

Kynurenine	<0.001	100.0%	-1.37	-1.38	0.08
Laurate (12:0)	<0.001	100.0%	-4.02	-0.51	0.07
Leucine	<0.001	100.0%	15.31	-2.00	0.46
Linoleate (18:2n6)	<0.001	100.0%	0.36	-2.07	0.16
Methylglutarate	0.002	100.0%	-4.54	-0.16	-0.03
myo-Inositol	<0.001	100.0%	21.56	2.52	0.89
Myristate (14:0)	<0.001	100.0%	-1.53	-1.93	0.27
N-6-trimethyllysine	0.001	100.0%	-1.38	0.89	-0.17
N-Acetylaspartate (NAA)	0.003	100.0%	-1.04	-0.76	-0.14
N-Acetylgalactosamine	<0.001	100.0%	-3.15	0.17	-0.37
N-Acetylglucosamine	<0.001	100.0%	-0.84	0.32	-0.58
N-Acetylglucosaminylamine	0.002	100.0%	-0.60	0.32	-0.74
Nicotinamide	<0.001	100.0%	11.61	1.26	-0.04
Nicotinamide adenine dinucleotide (NAD+)	0.002	100.0%	-3.34	-1.22	0.56
Octadecanoic acid	<0.001	100.0%	6.02	-1.21	0.42
Oleate (18:1n9)	<0.001	100.0%	4.07	-2.43	0.05
Orthophosphate (Pi)	<0.001	100.0%	25.61	-1.86	0.98
Palmitate (16:0)	<0.001	100.0%	8.30	-2.55	0.54
Palmitoleate (16:1n7)	<0.001	100.0%	-3.46	-1.33	0.20
Pantothenate	0.004	92.9%	0.64	-0.69	-0.01
Phosphoserine	<0.001	100.0%	-2.15	0.43	-0.03
p-Hydroxyphenyllactate (HPLA)	<0.001	100.0%	-4.00	-1.09	0.20
Pipecolate	<0.001	100.0%	-2.46	-1.02	0.42
Proline	<0.001	100.0%	15.66	-1.81	0.37
Putrescine	<0.001	100.0%	2.61	4.24	-1.25
Pyridoxamine	0.001	95.2%	-3.93	0.38	-0.11
Riboflavin (Vitamin B2)	<0.001	100.0%	-3.12	-0.47	-0.17
Ribose	<0.001	100.0%	1.37	2.35	-0.47
S-Adenosylmethionine (SAM)	0.001	100.0%	-4.45	-0.23	0.01
Sarcosine (N-Methylglycine)	<0.001	100.0%	-3.77	-1.26	0.25
Sorbitol	0.001	100.0%	-3.73	-0.89	0.22
Spermidine	<0.001	100.0%	-2.50	1.89	-0.70
Spermine	<0.001	100.0%	0.03	3.89	-0.91
Taurine	<0.001	100.0%	-1.39	1.16	0.16
Thymine	<0.001	100.0%	-4.51	-0.44	0.06
Tryptophan	<0.001	100.0%	10.69	-0.91	0.30
Uracil	<0.001	100.0%	2.03	-3.16	0.26
Urate	<0.001	100.0%	-1.18	-0.61	0.45
Urea	<0.001	100.0%	11.33	-2.00	0.97
Uridine	<0.001	100.0%	7.16	0.77	-0.25
Valine	<0.001	100.0%	12.84	-1.02	0.17
Xanthine	<0.001	100.0%	1.38	-3.27	0.41
Xanthosine	<0.001	100.0%	-3.93	-0.74	-0.27
Isobars and Un-named					
Isobar 1 includes mannose, fructose, glucose, galactose	0.001	100.0%	-2.35	0.77	0.28
Isobar 17 includes arginine, N-alpha-acetyl-ornithine	0.005	83.3%	3.62	0.73	0.42
Isobar 19 includes D-saccharic acid,1,5-anhydro-D-glucitol	<0.001	100.0%	-3.76	0.42	0.00
Isobar 2 includes 2-aminoisobutyric acid,3-	<0.001	100.0%	8.98	-1.91	0.13

amino-isobutyrate					
Isobar 24 includes L-arabitol, adonitol	<0.001	100.0%	-2.90	-0.82	0.06
Isobar 3 includes inositol 1-phosphate, mannose 6-phosphate	<0.001	100.0%	-3.29	0.37	-0.38
Isobar 40 includes Maltotetraose, stachyose	0.003	100.0%	-2.87	0.85	-0.39
X-1104	<0.001	100.0%	-4.08	0.26	-0.07
X-1111	<0.001	100.0%	10.83	-0.10	-0.56
X-1114	0.002	100.0%	12.88	-0.93	0.15
X-1142	0.004	100.0%	-2.98	-0.72	0.30
X-1186	0.001	97.6%	-3.21	1.06	0.60
X-1329	<0.001	100.0%	-4.45	-0.52	0.12
X-1333	0.002	100.0%	-3.09	0.89	-0.86
X-1342	0.003	100.0%	-2.18	-0.96	-0.06
X-1349	<0.001	100.0%	2.20	4.20	-1.81
X-1351	<0.001	100.0%	-2.18	-1.08	0.37
X-1465	<0.001	100.0%	-3.71	-0.66	0.06
X-1575	0.01	100.0%	-3.69	0.61	-0.06
X-1576	<0.001	100.0%	-3.62	-0.34	-0.21
X-1593	0.003	100.0%	-4.52	0.21	0.03
X-1595	<0.001	100.0%	-1.78	2.09	0.98
X-1597	0.001	100.0%	-1.03	-0.49	0.34
X-1608	0.005	100.0%	-4.17	0.43	0.17
X-1609	0.002	100.0%	-4.24	0.28	-0.05
X-1679	<0.001	100.0%	-3.66	-0.59	-0.16
X-1843	<0.001	100.0%	-4.47	-0.44	-0.04
X-1963	<0.001	100.0%	-1.95	0.73	0.32
X-1977	<0.001	100.0%	-0.79	-1.29	-0.23
X-1979	0.005	92.9%	-3.55	0.20	-0.07
X-2055	0.008	83.3%	-3.99	0.22	0.10
X-2074	<0.001	100.0%	-3.67	0.24	0.07
X-2105	0.005	90.5%	-3.66	0.11	-0.32
X-2108	0.005	100.0%	-0.98	-0.65	0.04
X-2118	<0.001	100.0%	8.38	1.08	-0.17
X-2141	0.007	88.1%	-2.92	-0.69	0.05
X-2143	0.002	100.0%	-3.44	-0.67	0.11
X-2181	<0.001	100.0%	-1.72	-0.84	-0.18
X-2237	0.001	100.0%	-2.93	-0.88	0.05
X-2272	<0.001	100.0%	-4.39	-0.30	-0.08
X-2292	<0.001	100.0%	-3.77	0.79	0.35
X-2466	<0.001	100.0%	-4.30	0.17	-0.17
X-2548	0.003	97.6%	-3.31	-0.19	0.14
X-2607	0.005	100.0%	-4.15	-0.24	0.05
X-2688	0.001	100.0%	1.16	0.59	-0.06
X-2690	<0.001	100.0%	-4.05	-0.86	0.27
X-2697	0.001	100.0%	-1.91	-1.43	-0.23
X-2766	<0.001	100.0%	-3.84	0.78	-0.34
X-2806	<0.001	100.0%	-0.96	1.00	-0.07
X-2867	<0.001	100.0%	0.66	2.99	-1.21
X-2973	<0.001	100.0%	-1.91	0.92	0.08
X-3003	0.001	100.0%	-4.48	-0.20	-0.03
X-3044	0.001	100.0%	1.53	-0.92	0.47

X-3056	<0.001	100.0%	-3.34	-1.47	-0.01
X-3102	<0.001	100.0%	-2.14	-1.84	0.24
X-3129	<0.001	100.0%	0.05	0.62	-0.14
X-3138	<0.001	100.0%	-3.59	-0.48	0.05
X-3139	<0.001	100.0%	-0.51	0.07	-0.57
X-3176	<0.001	100.0%	17.50	0.96	0.63
X-3220	0.001	100.0%	-4.47	-0.38	-0.04
X-3238	<0.001	100.0%	-4.56	-0.37	0.02
X-3379	<0.001	100.0%	-3.62	0.33	0.03
X-3390	<0.001	100.0%	-0.95	1.62	0.10
X-3489	0.001	100.0%	-2.34	1.53	-0.39
X-3771	<0.001	100.0%	1.04	1.31	0.46
X-3778	<0.001	100.0%	8.04	5.11	5.78
X-3807	<0.001	100.0%	-3.96	-1.00	0.01
X-3808	<0.001	100.0%	-2.21	1.36	-0.24
X-3810	<0.001	100.0%	-3.69	0.51	0.16
X-3816	<0.001	100.0%	1.36	4.37	-1.46
X-3833	0.002	100.0%	-4.60	-0.10	-0.03
X-3893	<0.001	100.0%	-0.45	2.82	-0.72
X-3952	0.001	100.0%	-3.90	0.27	0.06
X-3955	<0.001	100.0%	-4.33	0.13	-0.04
X-3960	<0.001	100.0%	-3.54	0.33	0.07
X-3992	<0.001	100.0%	0.41	0.78	-0.24
X-3997	0.002	100.0%	-4.48	-0.43	0.04
X-4013	<0.001	100.0%	-4.61	-0.14	-0.06
X-4015	<0.001	100.0%	-0.42	-1.01	-0.21
X-4018	<0.001	100.0%	-4.44	-0.52	0.12
X-4027	<0.001	100.0%	-3.19	-0.75	-0.18
X-4051	<0.001	100.0%	-3.55	-0.77	0.04
X-4075	<0.001	100.0%	-3.53	-1.48	-0.07
X-4084	<0.001	100.0%	-1.68	0.35	0.15
X-4096	<0.001	100.0%	-2.64	0.26	-0.12
X-4117	0.003	100.0%	-1.96	-0.76	-0.08
X-4365	<0.001	100.0%	3.68	2.08	1.43
X-4428	0.002	100.0%	-4.02	-0.41	-0.02
X-4514	<0.001	100.0%	-3.58	0.24	0.05
X-4567	0.003	95.2%	-4.03	-0.92	0.09
X-4611	<0.001	100.0%	-3.19	-0.81	0.26
X-4615	<0.001	100.0%	-4.47	-0.36	-0.11
X-4616	0.005	95.2%	-4.39	-0.07	-0.20
X-4617	0.001	100.0%	-3.09	-0.48	0.18
X-4620	<0.001	100.0%	-4.28	-0.74	0.12
X-4624	0.003	85.7%	-2.24	0.14	0.14
X-4649	<0.001	100.0%	-4.50	-0.48	0.06
X-4866	0.001	100.0%	-4.28	0.36	0.06
X-4869	<0.001	100.0%	-3.98	-0.38	0.14
X-5107	0.001	100.0%	14.27	-3.16	-2.39
X-5109	0.004	100.0%	18.56	-3.55	-2.56
X-5110	0.004	81.0%	19.53	-3.20	-2.26
X-5128	<0.001	100.0%	-2.53	2.15	0.05

X-5187	<0.001	100.0%	1.20	-1.60	1.03
X-5207	<0.001	100.0%	3.09	1.09	-0.11
X-5208	<0.001	100.0%	5.13	2.13	0.49
X-5209	<0.001	100.0%	-3.15	1.04	-0.34
X-5210	<0.001	100.0%	0.56	1.63	0.39
X-5212	<0.001	100.0%	-3.50	0.78	-0.39
X-5214	0.003	100.0%	-2.78	1.52	-0.41
X-5215	<0.001	100.0%	-2.09	0.53	0.20
X-5229	0.003	100.0%	-2.51	0.48	0.50
X-5232	0.002	97.6%	0.36	-2.08	-1.55

Supplementary Table 6. Clinical information associated with tissue specimens used for sarcosine validation

Tissue type	Number of samples	Number of patients
Benign adjacent prostate tissue	25	20
Local tumor (PCA) tissue	36	36
Metastatic tumor tissue	28	19
Metastasis site: adrenal	1	1
Liver	14	12
Lung	1	1
Mesentary	2	1
Pancreas	1	1
Periaortic lymph	3	2
Soft tissue	2	2
Unknown	4	4

Supplementary Table 7. Clinical information of urine supernatants used for sarcosine assessment

Characteristic	Urine Supernatant Samples (n=110)	Urine Sediment Samples (n=93)
Biopsy Negative		
No. of patients	51*	44 **
Age at biopsy (years)	63.4 ± 9.9 [42, 82]	60.7 + 9.6 [40, 77]
Baseline PSA (ng/ml)	6.1 ± 3.8 [0.8, 20.8]	5.3 + 2.3 [1.1, 10.0]
Biopsy Positive		
No. of patients	59 #	49 ##
Age at biopsy (years)	68.0 ± 8.9 [51, 85]	63.8 + 9.3 [47, 81]
Baseline PSA (ng/ml)	11.9 ± 19.6 [2.7, 111]	11.4 + 23.5 [2.7, 111.0]
Gleason Sum		
6	25 (42.4%)	19 (41.3%)
7	25 (42.4%)	20 (43.5%)
8	3 (5.1%)	2 (4.4%)
9	5 (8.5%)	5 (10.9%)
10	1 (1.7%)	0 (0%)
Maximum tumor diameter	1.7 ± 1.0 [0.5, 4.3]	
Gland weight	49.1 + 12.2 [28.2, 75.1]	49.9 + 14.6 [28.2, 77.6]

[+] For continuous variables the Mean ± SD [range] is given. Count and percentage is given for categorical variables. 43 men contributed both urine sediment and supernatant.

* There are 51 benign samples from 51 men. Age is available only for 49 individuals and PSA levels are available only for 45 of these men.

** There are 44 benign samples from 44 men. Age at biopsy is available for only 38 individuals and PSA levels are available for only 19 of these men.

There are 59 biopsy-proven localized prostate cancer patients. PSA levels are available only for 55 of these men. Maximum tumor diameter is available for 19 patients. Gland weight is available for 25 men.

There are 49 biopsy-proved localized prostate cancer patients. Biopsy age is available for 46 of these men and PSA levels are available for 20 of these men. Gland weight is available for 23 patients.

Supplementary Table 8. Association of urine derived sarcosine measures with common clinical parameters

Characteristic[+]	Urine Supernatant Samples	Urine Sediment Samples
Correlation with Sarcosine (log_2)		
Age	0.18	0.19
PSA (log)	0.22	-0.06
Gland weight	-0.09	-0.17
Two-tailed Wilcoxon rank-sum test of sarcosine (log_2)		
Diagnosis (neg v pos)	P=0.0025	P=0.0004
Gleason (6 v 7+)	P=0.5756	P=0.6880

+ Refer to Supplementary Table 8, above, for sample size information.

Supplementary Table 9: RNA interference sequence used in the study

Target Gene	NCBI Locus ID	siRNA sequences
GNMT	NM_018960	ACAAGUGGGUCAUCGAAGA
GNMT	NM_018960	GAGUCUGGCUUUCGCAUU
ERG	NM_004440	CGACATCCTTCTCTCACAT
EZH2	NM_004456	GAGGTTCAGACGAGCTGAT
DMGDH	NM_013391	AAGCUGGACUGGAAUAUUU
DMGDH	NM_013391	GGUUUUAGCUGGAUUGUAU
SARDH	NM-007101	GGAGCGACCGGGAUGGUUU
SARDH	NM-007101	GGACAAAGUACCCAUGUUU

Supplementary Table 10: Sequence of gene-specific PCR primers used in the study

Gene-specific primer	NCBI Locus ID	Sequence
GNMT F1	NM_018960	CTTCATCCACGTGCTCAAGA
GNMT R1	NM_018960	TCCCCATCTTCCAGACAGAG
GNMT pF2	NM_018960	ACCGTGTTACCGTATTCCAG
GNMT pR2	NM_018960	GGAAGACAGGGGGAGTCTCT
SARDH F1	NM_007101	CTGATGAATGTGGACGACCT
SARDH R1	NM_007101	GTTCTCAATGACCTGTGCTC
DMGDH F1	NM_013391	ACAGGGACATATGCGAAAGC
DMGDH R1	NM_013391	GAAGACTGGTCTGCCTCACC

* Denotes second set of primers used to assess GNMT levels
F and R stand for forward and reverse primers respectively.