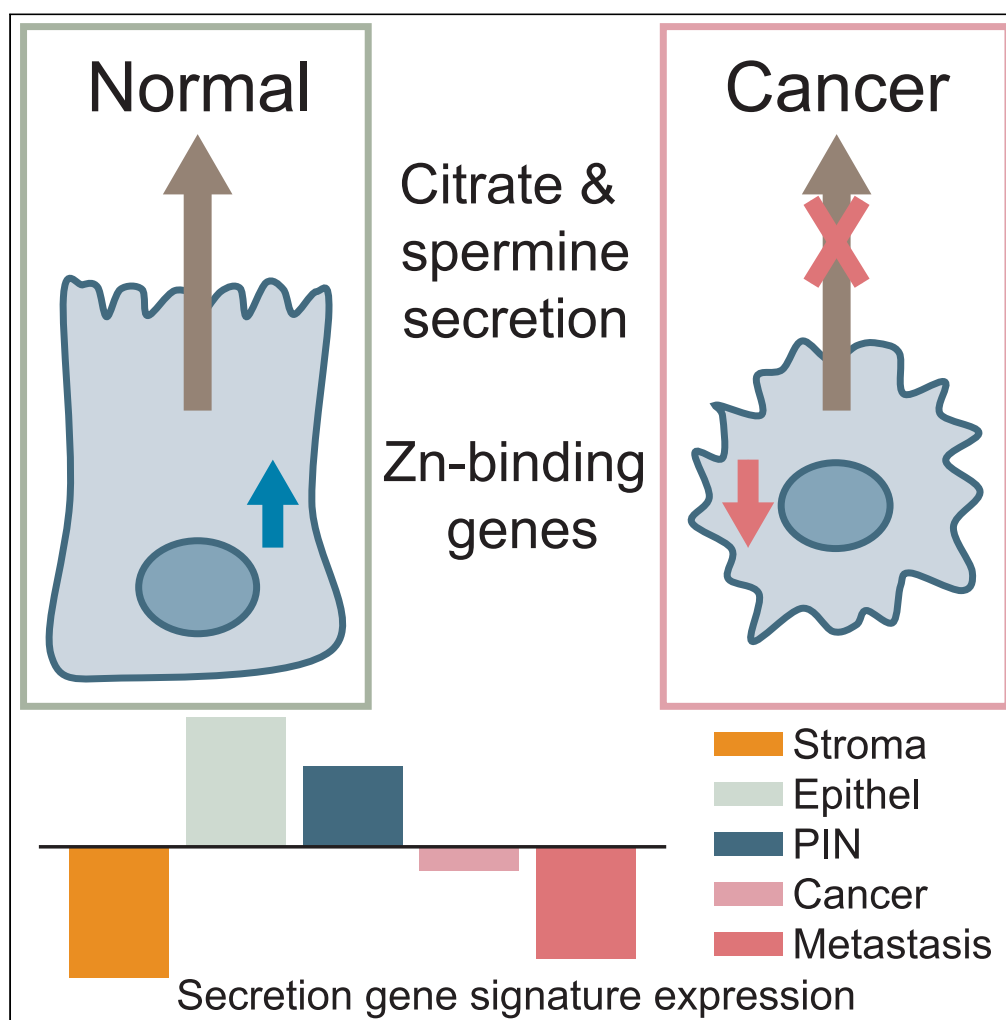


Article

The genes controlling normal function of citrate and spermine secretion are lost in aggressive prostate cancer and prostate model systems



Morten Beck Rye,
Sebastian Krossa,
Martina Hall, ...,
Finn Drabløs, May-
Britt Tessem,
Helena Bertilsson

morten.rye@ntnu.no

Highlights

Novel 150 gene signature
reflecting prostatic citrate
and spermine secretion

Identified several zinc-
binding proteins not
previously investigated in
the prostate

The signature is absent in
prostate model systems

Rye et al., iScience 25, 104451
June 17, 2022 © 2022 The
Author(s).
[https://doi.org/10.1016/
j.isci.2022.104451](https://doi.org/10.1016/j.isci.2022.104451)

Article

The genes controlling normal function of citrate and spermine secretion are lost in aggressive prostate cancer and prostate model systems

Morten Beck Rye,^{1,2,3,4,10,*} Sebastian Krossa,⁵ Martina Hall,^{6,7} Casper van Mourik,^{1,8} Tone F. Bathen,⁵ Finn Drabløs,¹ May-Britt Tessem,^{2,5,9} and Helena Bertilsson^{1,2,9}

SUMMARY

High secretion of the metabolites citrate and spermine is a unique hallmark for normal prostate epithelial cells, and is reduced in aggressive prostate cancer. However, the identity of the genes controlling this biological process is mostly unknown. In this study, we have created a gene signature of 150 genes connected to citrate and spermine secretion in the prostate. We have computationally integrated metabolic measurements with multiple transcriptomics datasets from the public domain, including 3826 tissue samples from prostate and prostate cancer. The accuracy of the signature is validated by its unique enrichment in prostate samples and prostate epithelial tissue compartments. The signature highlights genes *AZGP1*, *ANPEP* and metallothioneins with zinc-binding properties not previously studied in the prostate, and the expression of these genes are reduced in more aggressive cancer lesions. However, the absence of signature enrichment in common prostate model systems can make it challenging to study these genes mechanistically.

INTRODUCTION

A unique hallmark for normal prostate epithelial cells is their ability to secrete large amounts of the metabolite citrate. While cells from other human organs convert citrate to isocitrate for oxidation and energy production in the TCA cycle, citrate conversion in the prostate epithelium is partly blocked, leading to accumulation of citrate (Costello and Franklin, 2000). It has been reported that citrate/isocitrate ratios in the prostate are 40/1, compared to 10/1 in other normal organs (Costello et al., 2000). In the prostate, excess citrate is secreted out of the cell into the prostate lumen of the prostatic glands, where it is a major constituent of the prostatic fluid and essential for normal prostate organ function.

Almost all types of prostate cancer originate from prostate epithelial cells, and nearly all prostate cancer cells lose their ability to secrete citrate during the development of malignant phenotypes (Costello and Franklin, 2000). A reduced level of citrate has therefore been suggested as a biomarker, both for cancer detection and for identification of cancers with more aggressive phenotypes (Braadland et al., 2018; Costello et al., 1999; Giskeodegard et al., 2013). Even with these demonstrations of importance for normal prostate function and prostate cancer progression, the research in this field has been markedly neglected, and thus the molecular mechanisms controlling these functions remain mostly unknown. Nevertheless, a few dedicated research groups have managed to gain some insights: First, high mitochondrial levels of zinc in the prostate block citrate conversion, and the regulation of zinc transporter genes in particular ZIP1 (SLC39A1), have been suggested to play a role in regulating prostate zinc levels (Costello and Franklin, 2016; Kolenko et al., 2013). Second, the metabolite aspartate is a possible source of increased citrate in prostate cells (Costello et al., 1988; Franklin and Costello, 1984). Aspartate is a precursor for oxaloacetate, which again is a precursor of citrate synthesis and works as a carrier of the carbon groups oxidized stepwise in the mitochondria during the TCA cycle. Increased levels of aspartate could thus lead to increased levels of citrate when the citrate/isocitrate conversion is blocked. Third, metabolites synthesized in the polyamine pathway are regulated differently in the prostate (Bettuzzi et al., 2000; Rhodes et al., 2002; Tessem et al., 2016). Some interest has been directed to the polyamine spermine because of its extremely high correlation with citrate (Lynch and Nicholson, 1997). Spermine and citrate have opposite polarity, and it has been speculated as to whether they may form a complex which is secreted into the prostate lumen (Cheng et al., 2001; Lynch and Nicholson, 1997).

¹Department of Clinical and Molecular Medicine, NTNU – Norwegian University of Science and Technology, P.O. Box 8905, 7491 Trondheim, Norway

²Clinic of Surgery, St.Olavs Hospital, Trondheim University Hospital, 7030 Trondheim, Norway

³Clinic of Laboratory Medicine, St.Olavs Hospital, Trondheim University Hospital, 7030 Trondheim, Norway

⁴BioCore - Bioinformatics Core Facility, NTNU – Norwegian University of Science and Technology, P.O. Box 8905, 7491 Trondheim, Norway

⁵Department of Circulation and Medical Imaging, NTNU - Norwegian University of Science and Technology, Trondheim, Norway

⁶Department of Biotechnology and Food Science, NTNU – Norwegian University of Science and Technology, Trondheim, Norway

⁷K. G. Jebsen Center for Genetic Epidemiology, NTNU – Norwegian University of Science and Technology, Trondheim, Norway

⁸Institute for Life Science & Technology, Hanze University of Applied Sciences, 9747 AS Groningen, the Netherlands

⁹These authors contributed equally

¹⁰Lead contact

*Correspondence: morten.rye@ntnu.no

<https://doi.org/10.1016/j.isci.2022.104451>



These presented mechanisms, though intriguing, represent only small steps on the way to a widened perspective of this unique prostate function which is not completely understood. This includes the genes and proteins involved, as well as the connections between them. Most mechanisms suggested so far have been studied in cell-lines and animal models, which are not able to fully capture the complex interplay within human prostate tissue. However, a large amount of genome-wide –omics data on prostate tissue, as well as their model systems, has accumulated since the introduction of the genomic era more than 20 years ago. These data resources have yet to be fully exploited for research on citrate and spermine secretion in the prostate. In this study we address this hallmark property of the prostate by performing extensive bioinformatic analysis on data resources from the public domain. We do this by first creating a robust gene signature representing citrate and spermine secretion in the prostate. We then explore and validate this gene signature in multiple datasets, cohorts and public resources. In total we analyzed 31 datasets (Aryee et al., 2013; Aytes et al., 2014; Berglund et al., 2018; Bertilsson et al., 2012; Blattner et al., 2017; Cai et al., 2013; Chandran et al., 2007; Erho et al., 2013; Ghandi et al., 2019; Hsu et al., 2010; Jia et al., 2011; Klijn et al., 2015; Kuner et al., 2013; Lizio et al., 2019; Lizio et al., 2015; Mortensen et al., 2015; Penney et al., 2015; Poisson et al., 2012; Prensner et al., 2011; Richards et al., 2019; Ross-Adams et al., 2015; Sboner et al., 2010; Sogaard et al., 2018; Taylor et al., 2010; Tomlins et al., 2007; Wang et al., 2010; Yu et al., 2004; Zhang et al., 2017a, see also [Key resources table](#)) with 18,020 samples, of which 3826 were prostate samples including normal/normal-adjacent (epithelium and stroma) tissue, high/low-grade cancer tissue, metastatic samples and model systems. Our study identifies genes strongly associated with citrate and spermine secretion in the prostate and explores their behavior during cancer progression and in model systems. Our results validate previously suggested mechanisms, but also discover new central genes and mechanisms related to citrate and spermine with potential importance for this unique and intriguing prostate hallmark.

RESULTS

All datasets used in this study are listed in [Key resources table](#). It includes dataset ID and abbreviation used in the main text, sample description, dataset accession and references.

Metabolite concentrations between citrate and spermine are highly correlated across tissue samples

Our research group has previously generated a dataset with 129 normal and cancer tissue samples from the prostate, where concentrations from 23 metabolites and the expression of 14,161 unique genes were measured on the exact same samples (Bertilsson, dataset ID 1). Similar to a previous report (Tessem et al., 2016), we observed a strong correlation ($r = 0.95$) between concentrations of the metabolites citrate and spermine across the samples (Figures 1A and 1B). To simplify calculations, we will use the average citrate-spermine concentration profile for calculations throughout the rest of this study (Figure 1C).

A gene signature associated with the citrate-spermine concentrations is validated in 12 datasets of prostate cancer and normal samples

We identified genes associated with citrate-spermine levels in the prostate by calculating the Pearson correlations between the citrate-spermine concentration profile and each of the 14,161 unique genes measured in Bertilsson (dataset ID 1). The correlation analyses were performed on the 95 cancer samples in the dataset. We selected the 150 genes with the highest positive correlation to citrate-spermine concentration, which we termed the initial gene module. We further assessed the integrity of this gene module by creating a Correlation Module Score (CMS, see [STAR Methods](#)) where a high CMS means that strong intra-correlation is present between the genes in the module, resulting in a high module integrity. On the contrary, a low CMS indicates that there is low correlation between the genes in the module, and that the module integrity is lost. Moreover, if the gene module has a significantly high CMS in a dataset, it indicates that the biological process(es) which this module represents is functionally active for the samples in this dataset. For the gene modules studied here the biological process would be citrate-spermine secretion.

To test for significance, we generated 100 random CMSs by shuffling the citrate-spermine metabolite levels between the respective samples in Bertilsson (dataset ID 1). The resulting CMS was statistically significant when compared to CMSs from random gene modules (CMS = 0.34, $p = 0.007$, lognormal distribution) (Figure 2A).

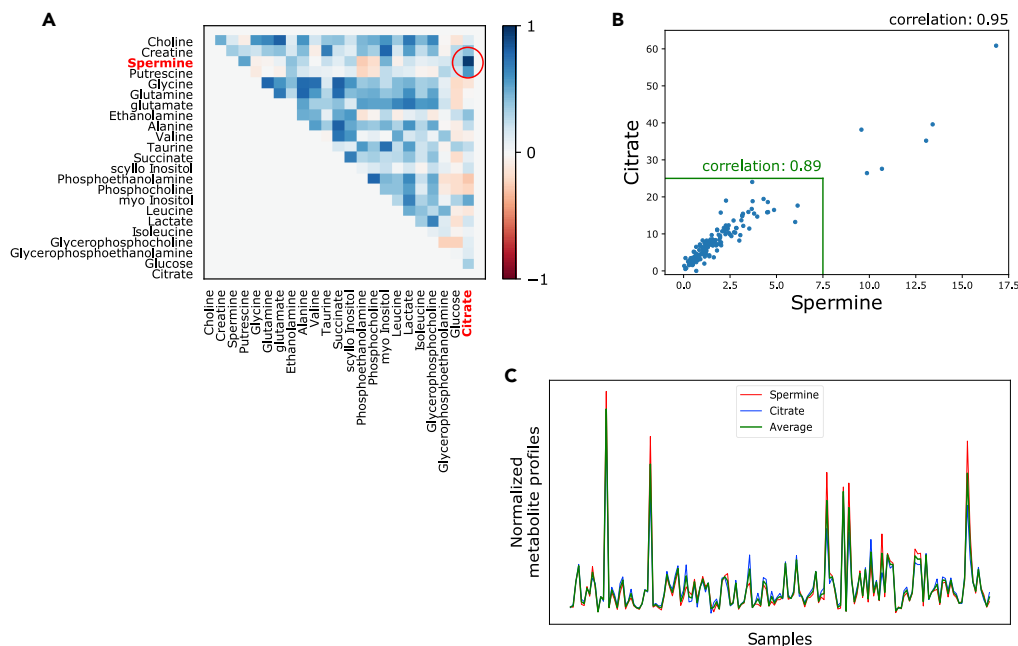


Figure 1. Correlations between metabolites citrate and spermine

(A) Correlations between all 23 metabolites measurements across 129 prostate normal and cancer samples from the Bertilsson dataset (Dataset ID 1). The correlation is particularly strong for metabolites citrate and spermine.

(B) Correlations between citrate and spermine across 129 prostate normal and cancer samples from Bertilsson. The correlation is high also when extreme values are excluded (highlighted inside the green boundary).

(C) Similarity between the citrate, spermine and the average citrate-spermine metabolite profiles across 129 prostate normal and cancer samples from Bertilsson.

We then validated whether the integrity of the initial gene module was significantly preserved in gene expression data from 11 additional datasets (dataset ID 2–12), including a total of 2638 tissue samples from normal prostate and prostate cancer (see [Key resources table](#)). We calculated and compared the CMSs for the initial module, as well as the 100 random modules in all 11 datasets. The CMS from the initial gene module was significant in all 11 cancer datasets, and 7 out of 8 datasets with normal samples (lognormal test) (Figure 2B, [Tables S1](#) and [S2](#)). We thus defined the initial module as our initial gene signature associated with citrate-spermine concentrations in the prostate. We used this initial citrate-spermine gene signature (abbreviated to the CS gene signature) as a starting point for further refinement.

We also performed the above procedure using the 40 normal samples in *Bertilsson*, however the resulting gene module was not significant (CMS = 0.37, $p = 0.14$, lognormal distribution), and could not be validated in any of the additional datasets (Figure S1, [Tables S3](#) and [S4](#)). The reason for this is probably because of an insufficient number of normal samples in *Bertilsson* to perform a robust correlation analysis.

Refinement across 12 datasets produces a citrate-spermine gene signature with improved integrity

Corresponding gene expression and metabolite measurements are available only for *Bertilsson* (dataset ID 1). We thus implemented a bioinformatics strategy to identify a refined CS gene signature with improved integrity across all 12 datasets (2794 samples). We used the initial CS signature from *Bertilsson* as a starting point, and then used results from cancer samples across the other 11 datasets to nominate better gene candidates to replace genes from the initial CS signature (see [STAR Methods](#)). When evaluating the new top 150 candidate genes, we found that 74 of the 150 genes in the initial CS signature had been replaced by new genes to improve integrity across all datasets (Figure 2C). Of note, the refined CS signature improved CMS for in both cancer and normal samples in all 11 additional datasets. This consistent improvement in normal samples (including normal samples from *Bertilsson*) affirms the relevance of the signature because these samples were not used to generate or refine the CS signature. In addition, only a marginal reduction in CMS was observed for the cancer samples in *Bertilsson* (Figure 2D). We thus conclude that the refined CS

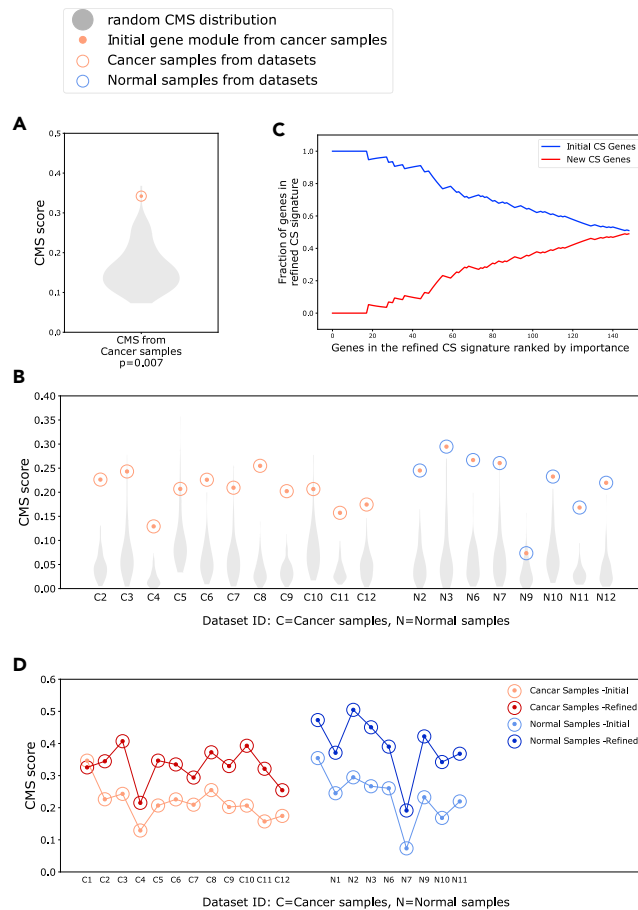


Figure 2. Integrity of initial and refined citrate – spermine (CS) gene signatures

(A) CMS (Correlation Module Scores) for the 150 genes most positively correlated with the CS metabolite profile for cancer samples in the Bertilsson dataset (Dataset ID 1). The CMS for was statistically significant when compared to CMSs using random metabolite profiles (lognorm test) (B) CMS for the initial cancer CS gene module from Bertilsson (red dots) evaluated in prostate cancer (red circles) and normal (blue circles) samples from 11 additional datasets. The CMS was statistically significant in all datasets (lognorm test, [Tables S1](#) and [S2](#)).

(C) Fraction of new genes in the refined CS gene signature that replaced genes in the initial CS gene signature. The genes is sorted by importance (rank) on the bottom axis. Of the 150 top ranked genes in the initial CS signature, 74 were replaced in the refined CS signature.

(D) Increase in CMS after refinement of the initial CS signature for prostate normal and cancer samples in 11 datasets. The CMS increased in the normal samples, even though the normal samples were not used to create or refine the signature. Cohort ID 4, 5 and 8 did not include normal samples. See also [Figure S1](#) and [Tables S1–S4](#).

signature (referred to as the CS signature from now) represents a more robust signature with improved integrity compared to the initial CS signature.

The citrate-spermine signature displays unique enrichment in prostate samples and negative correlation to stromal tissue compartments

Having established a CS signature of 150 genes associated with citrate and spermine secretion in the previous sections, we expanded our validation to additional datasets and sample types. For this analysis we used single samples Gene Set Enrichment Analysis (ssGSEA) ([Markert et al., 2011](#)), which is a method used to assess whether a specific gene signature is enriched or depleted in a single sample. Because citrate and spermine secretion is a highly specific function associated with prostate epithelium, higher ssGSEA scores in tissue samples from prostate compared to samples from other tissue types would confirm its functional relevance. Further, because prostate cancer samples are a heterogeneous tissue mixture of normal

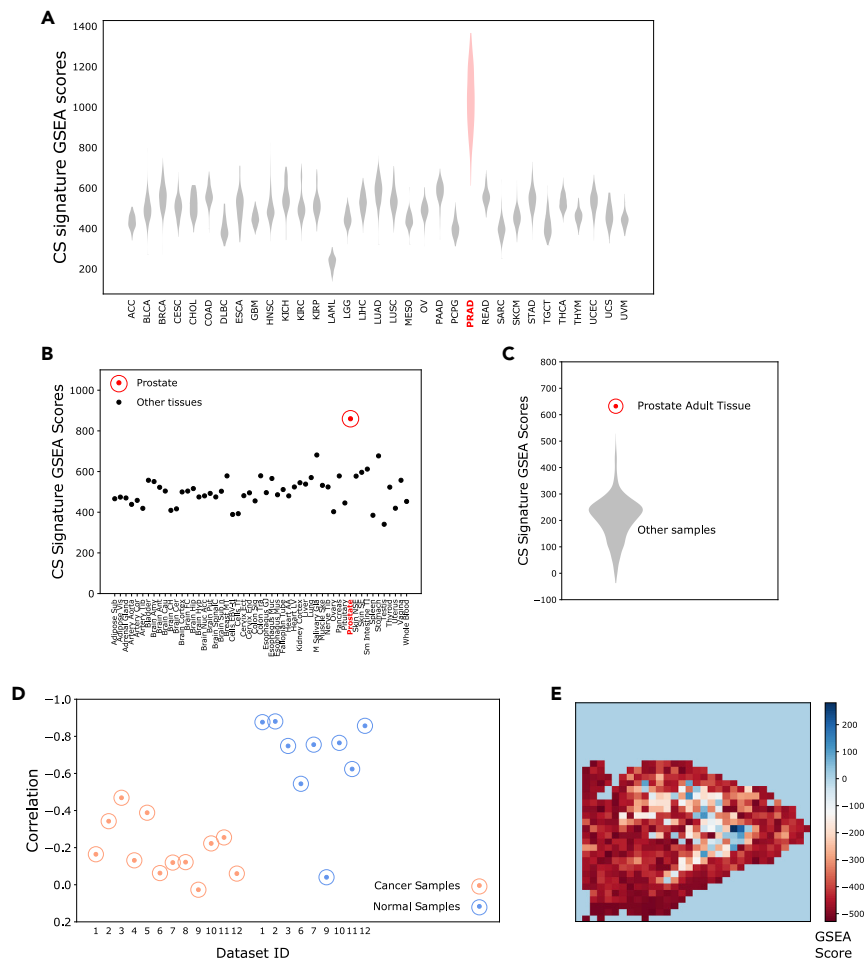


Figure 3. Validation of citrate - spermine (CS) gene signature

(A) CS signature ssGSEA scores for 11,093 cancer and normal samples in 33 tumor types from the TCGA-complete dataset. We have used standard TCGA cancer type abbreviations [<https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations>] where PRAD is the abbreviation for prostate cancer.

(B) CS signature ssGSEA scores for 53 averaged human normal tissue profiles from the GTex dataset.

(C) CS ssGSEA signature scores for 1829 human cell-line and tissue samples, including one prostate adult tissue sample, from the FANTOM dataset.

(D) Correlations between CS signature and stroma signature ssGSEA scores for cancer and normal samples in prostate cancer tissue datasets 1–12. The negative correlations in the normal samples are stronger due to less confounding from cancer tissue. Data from Mortensen (dataset ID 9) contains laser dissected prostate normal epithelium and cancer, which probably lack significant amounts of stroma.

(E) CS signature ssGSEA scores on spatial transcriptomics data from tissue slice 3.2 in Berglund (dataset ID 20). The colorbar indicates enrichment of the ssGSEA score in each pixel. The corresponding pathological tissue image can be found for image 3.2 in Supplementary Information –Figure S1B from Berglund et al. (Berglund et al., 2018), and comparison shows a strong overlap of high and low CS signature scores with epithelial/lumen and stroma tissue compartments, respectively. Results from all 12 images in Berglund are shown in Figure S2.

prostate epithelium, cancer and stroma, ssGSEA scores should be inversely proportional to the content of stroma tissue in the samples because stromal cells are unable to secrete citrate or spermine.

We first compared ssGSEA scores for cancer and normal samples in the prostate specific TCGA dataset (dataset ID 5) to all cancer and normal samples profiled in the TCGA-complete resource (11,093 samples from 33 cancer types, dataset ID 21). We observed strongly elevated ssGSEA scores for both prostate cancer and normal samples compared to cancer and normal tissues from other tissues (Figure 3A). We also tested the CS signature on average expression profiles for 53 normal tissues from the GTex portal (dataset

ID 22) (Figure 3B) and 1829 CAGE expression profiles from cell-lines and tissues in the FANTOM Consortium (dataset ID 23) (Figure 3C) where one tissue profile was from adult normal prostate. In both datasets, normal prostate tissue showed the highest ssGSEA scores.

We next correlated the CS signature ssGSEA scores to a previously identified gene signature for stroma content in prostate tissue (Rye et al., 2018) in datasets 1–12. We observed a strong reverse correlation between CS and stroma signatures for normal samples in all datasets (Figure 3D) (with an exception for Mortensen (dataset ID 9) which contained laser dissected epithelium/cancer with minimal amounts of stroma). The correlation was weaker in cancer samples. This is expected because prostate cancers vary in their ability to secrete citrate and spermine. The negative association between CS signature and stroma was also confirmed using spatial transcriptomics data from Berglund (dataset ID 20), where pixels with high and low ssGSEA scores overlaid regions with epithelium/lumen and stroma respectively (Figures 3E and S2).

In summary, the CS signature shows strong associations with expected properties of the prostate epithelium, which strengthens the biological validity of the genes in the signature.

The citrate-spermine gene signature is gradually lost from normal samples through low grade, high grade and metastatic lesions

Having established the relevance of the CS signature to prostate citrate and spermine secretion, we next investigated how this signature associates with the various stages of prostate cancer aggressiveness. Low citrate concentrations have previously been associated with high-grade prostate cancer (Giskeodegard et al., 2013). We identified nine datasets (1779 prostate cancer samples) where cancer samples had been classified as either high-grade or low-grade, or assigned a Gleason score - the standard form of prostate cancer grading. We found significant changes in CS signature ssGSEA scores between high- and low-grade cancers in all eight datasets where samples where Gleason score 4 + 3 was classified as high-grade (613 high and 737 low-grade samples), and six out of eight datasets where samples with Gleason score 4 + 3 was classified as low-grade (578 high- and 1258 low-grade samples) (Figure 4A, Table 1). The reduction in CS signature scores was not due to increased tumor fraction in high-grade samples (shown in Table 1). These results support the previous findings and suggest that the changes previously observed for citrate at metabolite level are accompanied by changes in gene expression. A further significant reduction in CS signature scores were observed in metastatic compared to cancer and normal samples in 8 datasets (157 metastatic samples) (Figure 4B and Table 2). Note that the lower CS-signature scores in normal samples compared to cancer samples is due to the high level of stromal tissue in normal prostate samples (Franklin and Costello, 1984; Rye et al., 2018) (Figure S3). Overall, our results indicate that the genes associated with citrate and spermine secretion is upregulated in the normal epithelium, and then gradually downregulated with tumor progression from early low-grade lesions, through high-grade cancers and finally metastatic tumors. This trajectory is nicely illustrated by the results from Tomlins (dataset ID 18) (Figure 4C), where laser dissection has been used to purify stroma, normal epithelium, Prostate Intraepithelial Neoplasia (PIN, an early pre-cancerous prostate lesion) cancer and metastatic tissues. In line with this trajectory, the PIN lesions show an average CS signature score level between the levels of normal epithelium and cancer. Though the ability of cells to secrete citrate and spermine seems to be lost in metastatic samples, metastatic tissue from prostate cancer still contains traces of the CS gene profile compared to metastatic tissue originating from other organs (Hsu, dataset ID 19) (Figure 4D).

Citrate-spermine signature genes are enriched for genes with zinc-binding, mitochondrial and secretory functions

Gene Ontology analysis anchored the genes in the CS gene signature to the citrate secretory pathways from mitochondria towards the cell exterior via the ER and golgi. The analysis also highlighted several genes related to zinc-binding, in particular several metallothionines, and a role of branched chain fatty acid catabolism in upholding the normal prostate secretory function.

When examining gene functions on gene-by-gene basis using the NCBI-gene resource [<https://www.ncbi.nlm.nih.gov/gene>] (Table S5) we found that several genes in the CS signature were related to zinc-binding. Among these, six were Metallothioneins (MT1G, MT1M, MT1F, MT1X, MT1E and MT2A), one zinc-transporter (SLC39A10) and the two genes AZGP1 (alpha-2-glycoprotein 1, zinc-binding) and ANPEP (alanyl

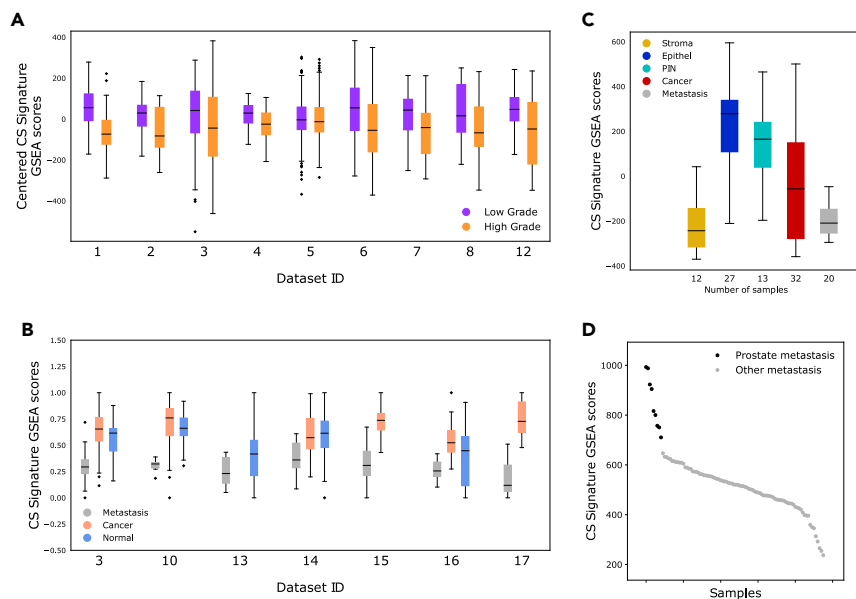


Figure 4. Citrate – spermine (CS) signature, tumor grade and metastasis

(A) CS signature ssGSEA scores for high-grade (Gleason higher or equal to 4 + 3) and low-grade (Gleason less than or equal to 3 + 4) cancers in 9 datasets. The samples compared correspond to the HG1 and LG1 groups in Table 1. The scores were centered in each cohort before plotting to visualize similarities between datasets better. See also Table S7B. (B) CS signature ssGSEA scores for metastatic, cancer and normal samples in 7 datasets. The scores were centered and normalized to range 0-1 before plotting to visualize similarities between datasets better. The low scores in normal samples is due to the higher content of stroma in normal samples (see Figure S3). (C) CS signature ssGSEA scores from five laser dissected prostate and prostate cancer tissue types from Tomlins (datasets Id 18). The GSEA scores illustrate the loss in CS secretion from normal epithelium through PIN, Cancer and finally Metastasis. The CS signature scores from metastatic samples are comparable to the non-secreting stroma tissue. (D) CS signature ssGSEA scores in 96 metastatic samples (9 from prostate and 87 from other organs) from the Hsu (dataset ID 19). Metastatic samples from prostate origin are more enriched for the CS signature than metastatic samples originating from other sites.

aminopeptidase, membrane), the latter containing a consensus sequence known from zinc-binding metalloproteinases.

We then used DAVID (Huang da et al., 2009a; b) and Enrichr (Chen et al., 2013; Kuleshov et al., 2016) to perform Gene Ontology (GO) analysis on the CS signature (Figures S4 and S5). For both GO tools the most significant terms were related to *zinc/metal-ion binding* and *branched chain amino acid (Leucine, Isoleucine and Valine) degradation*. When inspecting the CS signature genes at NCBI, we noticed that many genes did not have any associated functional description, while at the same time having high prostate specific expression. A potential challenge for our GO analysis is that the prostate specific function of many of the genes in our CS signature is not known. We therefore applied GAPGOM, an alternative GO analysis tool where each gene is associated with functions based on consensus ontology terms for genes it is correlated with (Ehsani and Drablos, 2018) (STAR Methods and Table S6). This analysis revealed strong gene ontology terms related to mitochondria, ER (endoplasmic reticulum), golgi, lysosome and exosome cellular components, which would fit well with a secretory path of citrate. When we performed Principal Component Analysis (PCA) to group genes according to their GO terms (see STAR Methods), the six metallothioneins formed a distinct cluster related strongly to zinc/metal-ion binding (Figure 5A and Table S6). In addition, the metallothioneins also showed strong associations with the mitochondrial respiratory complex and electron transport chain.

We also performed network analysis on the 150 genes in the CS signature (Figure S6), where the genes (*ALOX15B*, *RAB27A*, *ENDOD1*, *SLC45A3*, *NCAPD3*, *EHHADH*, *ACAD8*, *AFF3*, *NANS* and *YIPF1*) were identified as the top 10 hubs in the network. These hub-genes were all ranked among the top 20 in the CS-signature and had significant GO terms associated with branched chain fatty acid catabolism/degradation. Finally, we

Table 1. p-values from differential CS signature ssGSEA score analysis between high grade (HG) and low grade (LG) cancer samples in 9 datasets

Dataset ID	Dataset abbreviation	Number of HG1 samples	Number of LG1 samples	Number of HG2 samples	Number of LG2 samples	p-value HG1 vs LG1	p-value HG2 vs LG2	Correlation to tumor content
1	Bertilsson	56	60	36	80	4.7e-10	3.6e-5	-0.13
2	Chen	15	50	11	54	0.0040	0.0011	0.14
3	Taylor	36	94	15	115	0.015	0.010	NA
4	Sboner	119	162	81	200	3.7e-7	1.7e-4	-0.30
5	Erho	NA	NA	211	334	NA	0.060	NA
6	TCGA	296	200	200	296	2.5e-10	5.3e-13	-0.04
7	CMBR	30	82	9	103	0.014	0.09	0.11
8	STCK	34	57	15	76	0.021	0.007	NA
12	Kuner	27	32	NA	NA	0.021	NA	NA

Two comparisons are made: First HG1 vs LG1, where HG1 is classified as samples with Gleason score 4 + 3 or higher, and LG1 as samples with Gleason score 3 + 4 or lower. Second HG2 vs LG2, where HG2 is classified as samples with Gleason score 8 or higher, and LG2 as samples with Gleason score 7 or lower. The correlation of CS signature GSEA score to tumor content is also included when available, and show, in general, that decreased CS scores in high-grade samples is not due to increased tumor content in these samples.

compared CS signature ssGSEA scores with cell-types specific gene signatures from a single-cell RNA-Seq study on normal prostate tissue (Henry et al., 2018). The CS signature scores showed high correlation with prostate luminal cells (the cells mainly responsible for secretion), but not basal cells, and a negative correlation to stromal fibroblasts and smooth muscle cells (Figure 5B). Moreover, the luminal gene signature also included several of the metallothionein genes, corroborating the potential importance of these genes.

To investigate whether the target genes changed with cancer development, we performed a differential expression analysis between high- and low-grade cancer samples for all 9 datasets from Figure 4A. In particular, the zinc-binding genes (all six metallothioneins, *AZGP1* and *ANPEP*) showed consistent downregulation in high-grade cancer (Table S7). We also observed consistent downregulation in 9 of the 10 network HUB genes. This supports a role for these genes during cancer progression. We did not observe downregulation of the zinc-transporter *SLC39A10*.

The citrate-spermine signature is depleted in prostate model systems

To reveal detailed function of the genes highlighted in the previous section will need further wetlab experiments. We therefore wanted to identify suitable model-systems where the possible role of these genes and their mechanisms could be further studied. Unfortunately, we found the identification of model systems to be a challenge.

We tested the CS signature on publicly available gene expression datasets from various model systems for prostate cancer using ssGSEA. In the *Prensner* (dataset ID 10), 58 expression profiles from cell-lines were analyzed together with tissue and metastatic samples, including the four most common prostate cancer cell-lines PC3, DU145, LNCaP, VCaP, and the normal prostate cell lines RWPE and PrEC. This unique collection of sample types makes the *Prensner* dataset an excellent reference for comparisons.

CS signature ssGSEA scores from the *Prensner* cell-lines were consistently at the low end, compared to both tissue and metastatic samples (Figure 6A), indicating that cell-lines derived from prostate do not secrete citrate and spermine. This observation was confirmed in *Taylor* (dataset ID 3), which also contained four cell-line samples (Figure S7). When comparing different cell-lines, androgen responsive cell-lines (LNCaP and VCaP) generally have higher CS signature ssGSEA scores than androgen resistant (DU145 and PC3) cell-lines (Figure 6B). The overall highest CS scoring cell-line is MDA-Pca-2b (an androgen responsive cancer cell-line). Interestingly, normal prostate cell-lines (RWPE and PrEC) do not score higher than the cancer cell-lines. The relative differences in CS ssGSEA scores between the different cell-lines were highly stable, which was confirmed when assessing additional cell-line samples in *CCLE* (Cancer Cell Line Encyclopedia, dataset ID 24) (Figure 6C), *Søgaard* (Dataset ID 26) and E-MTAB-2706 (dataset ID 25) (Figures S8 and S9, respectively). In addition, results from cohort *CCLE* and E-MTAB-2706 reveal that

Table 2. p-values from comparing CS signature ssGSEA scores from metastatic prostate cancer samples to cancer and normal samples and in 8 datasets

Dataset ID	Dataset abbreviation	Number of metastatic samples	Number of cancer samples	Number of normal samples	p-value metastatic vs cancer	p-value metastatic vs normal
3	Taylor	19	131	29	7.4e-15	2.8e-7
10	Prensner	12	78	38	3.4e-11	2.9e-10
13	Aryee	18	NA	21	NA	0.024
14	Chandran	25	65	81	8.8e-6	6.9e-6
15	Cai	29	22	NA	5.8e-13	NA
16	Poisson	13	12	16	1.5e-4	0.09
17	Monzon	21	10	NA	7.4e-11	NA
18	Tomlins	20	32	27	0.007	1.2e-10

androgen resistant prostate cancer cell-lines (DU145 and PC3) do not separate from cell-lines of other cancer origin in terms of CS signature GSEA scores (Figures 6C and S8).

To enable the comparison of ssGSEA scores between datasets, we implemented a method to adapt ssGSEA scores from different datasets to the *Prensner* dataset (see STAR Methods). To demonstrate the utility of this implementation, we first adapted the CS ssGSEA scores from *Taylor* to *Prensner* because both datasets included tissue samples (normal and cancer), metastatic samples as well as cell-lines. The CS ssGSEA scores were highly concordant down to the level of different cell-types between the two cohorts after adaptation (Figure S10). We then used the adaptation strategy to compare datasets from organoids and mouse models (dataset IDs 27–31) to the *Prensner* dataset to investigate whether they would resemble tissue or cell-lines. None of the model-systems analyzed seemed to produce CS ssGSEA scores at levels similar to prostate epithelial tissue or low-grade primary cancer tissue. Instead their CS scoring range typically compare to the cell-lines from *Prensner* (Figure 6D). Overall, these results call attention to the research challenge for current prostate model systems to recapitulate the function of normal prostate epithelium in terms of citrate and spermine secretion.

DISCUSSION

Expanding the view of zinc-binding proteins and zinc-transporters in the prostate

Six metallothioneins were identified as part of the CS gene signature in this study and constituted a functionally distinct cluster particularly associated with zinc-binding. Metallothioneins are low molecular weight, metal binding proteins localized to the Golgi, and have different expression in normal prostate and prostate cancer (Garrett et al., 2000; Si and Lang, 2018). In addition, expression levels can change in response to zinc-stimuli in prostate cell-lines (Lin et al., 2009). It has also been speculated that metallothioneins can affect mitochondrial function because they are small enough to enter the mitochondrial membrane bilayer carrying zinc (Franklin et al., 2005). This suggestion fits with the GO associations of metallothioneins to mitochondria, respiratory complex and the electron transport chain. Otherwise, the other potential zinc-binding genes discovered in this study have not been studied previously in the context of citrate accumulation and secretion in the prostate.

The zinc-transporter *SLC39A1* (ZIP1) was not a part of the CS-signature, though this gene has previously been shown to be important for zinc-homeostasis in prostate cells. There could be several reasons for this discrepancy. First, this study only measures genes at the transcript level, while it is possible that *SLC39A1* is regulated at the protein level (Costello and Franklin, 2016). However, expression levels from genes displayed in NCBI-gene resource demonstrate that *SLC39A1* is not prostate specific, but expressed in most tissues. Though *SLC39A1* has the ability to transport zinc within prostate tissue, it may not be the main determinant of prostate specific zinc regulation *in vivo*. In this context it should also be noted that none of the metallothioneins displayed prostate specific expression in the NCBI-gene database. Of the zinc transporters, most prostate specific expression is observed for *SLC39A2*, *SLC39A6*, *SLC39A7* and *SLC39A10*, where *SLC39A10* was part of the CS signature. *SLC39A6* and *SLC39A7* (though not part of the signature) were also positively correlated with citrate in our analysis whereas *SCL39A1* and *SLC39A2*

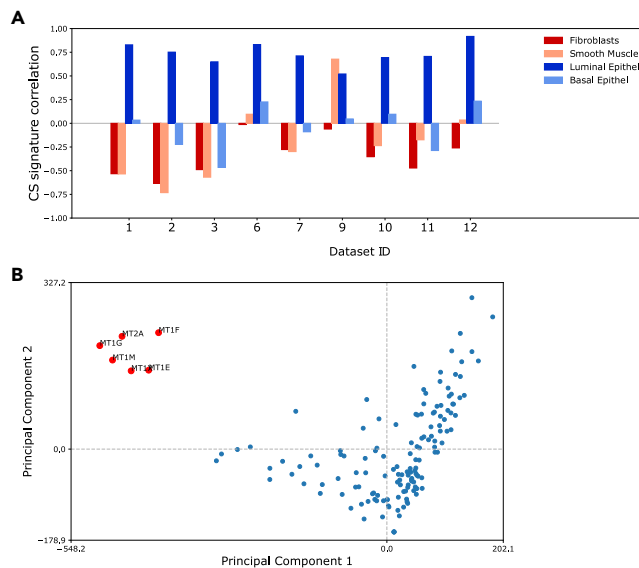


Figure 5. Genes and ontologies

(A) Principal Component Analysis (PCA) for Gene Ontology terms associated with CS signature genes. The GO terms are based on a consensus analysis over datasets 1–12 (see [STAR Methods](#)). The only observable cluster is formed by the six Metallothioneins from the CS signature.

(B) Correlation of ssGSEA signature with four gene signatures from distinct cell types in a single cell sequencing dataset (basal and luminal from the prostate epithelium, fibroblasts and smooth muscle cells from stroma) ([Henry et al., 2018](#)) in 9 datasets with normal prostate samples. The CS gene signature correlates with luminal epithelial cells (the cell-type mainly responsible for secreting citrate and spermine) whereas the correlation to stroma tissue types is negative. Dataset 9 is from laser dissected epithelium, which lack stroma tissue. See also [Figure S4–S6](#) and [Tables S5–S7](#).

showed no correlation. Of the potential zinc-binders, *AZGP1* demonstrate the most prostate-biased expression, but the function of this gene is unknown. In contrast to *ANPEP*, *AZGP1*, the six metallothioneins and the network HUB genes, only one of the zinc-transporters (*SLC39A1* - upregulated) showed any consistent differential expression between high- and low-grade cancer samples. Overall, more targeted research experiments are needed to identify the genes that control zinc-levels in the prostate, and how they perform their function.

Previous studies have shown that prostate epithelial cells can utilize glucose and aspartate as the carbon sources for citrate production. The gene *GOT2* (*glutamic-oxaloacetic transaminase 2*, also named *mAAT-mitochondrial aspartate aminotransferase* in previous reports) was shown to be responsible for the conversion of aspartate to oxaloacetate and citrate in the mitochondria. *GOT2* was a part of the CS signature, in addition to *SMS* (*spermine synthase*), and both *GOT2* and *SMS* showed high tissue-expression in prostate according to NCBI-Gene. For these two genes our results fit with previous mechanistic knowledge ([Costello et al., 1988](#); [Juang et al., 1995](#)).

Very little is known about the role of branched chain amino acid catabolism for normal prostate function. All of the three amino acids leucine, isoleucine and valine are upregulated in prostate cancer ([Tessem et al., 2016](#)), indicating their relevance for cancer transformation, and this fits with our finding that the degradation of these amino acids is important to maintain normal prostate function. Branched chain amino acids can possibly work as precursors for citrate production ([Zhang et al., 2017b](#)), and leucine can act as a sensor for mTOR-pathway activation ([Ananieva and Wilkinson, 2018](#)), which is generally regarded as an important pathway in prostate cancer development ([Edlind and Hsieh, 2014](#)).

Expression of citrate-spermine signature genes are with prostate cancer progression

There has been a discussion on whether the reduced citrate levels observed in prostate cancer are merely a result of a reduction in luminal glands leading to reduced amounts of prostatic fluid ([Tayari et al., 2017](#)). However, our results clearly indicate that changes in citrate and spermine levels are accompanied by changes at the

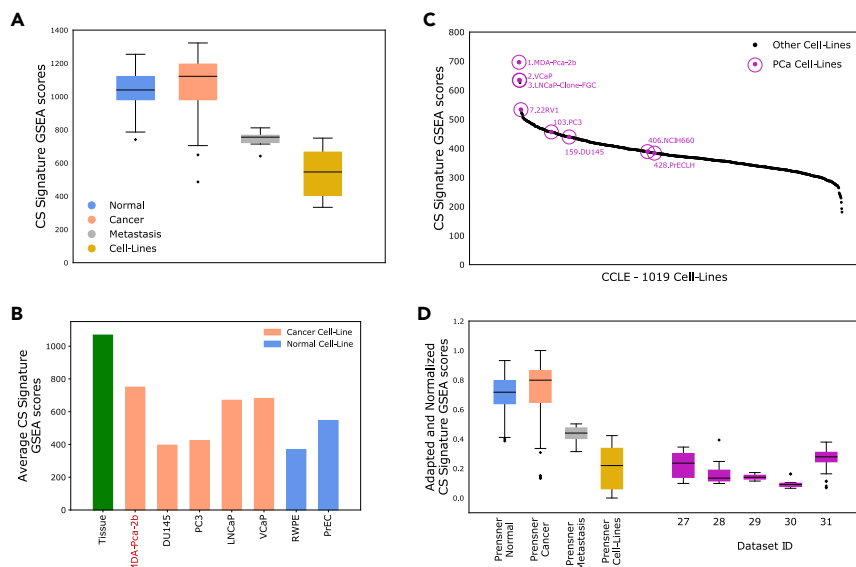


Figure 6. Model systems

(A) CS signature ssGSEA scores for all sample types in Prensner (dataset ID 10), including 58 prostate normal and cancer cell-line samples. Cell-line samples, in general, have very low CS signatures scores.

(B) Average CS signature ssGSEA scores for different cell-types in Prensner. The ssGSEA scores were highly reproducible within samples from the same cell-type. Androgen responsive cell-types (MDA-Pca-2b, LNCaP and VCaP) have somewhat higher ssGSEA scores than androgen resistant cell-types (DU145 and PC3). Prostate normal cell-types are similar to cancer cell-types. Number of samples of each cell-type are: MDA-Pca-2b (1), DU145(10), PC3(2), LNCaP(7), VCaP(10), RWPE(9), PrEC(4).

(C) CS signature ssGSEA scores in 1019 cancer cell-types from CCLF (dataset ID 24). Androgen responsive prostate cancer cell-types score higher than cancer cell-types from other cancers whereas androgen resistant cell-lines score similar to cell-types from other cancers.

(D) CS Signature ssGSEA scores in four datasets with prostate and prostate cancer derived organoids and model systems (dataset IDs 27–30) and one dataset with prostate tumors from mouse models (Aytes, dataset ID 31). The ssGSEA scores were adapted and normalized to CS signature GSEA scores from Prensner. All model systems have consistently low GSEA scores, similar to prostate cancer cell-lines. In total 64 model system samples and 384 mouse tumor samples were analyzed. Dataset descriptions: **27**: 2D and 3D cultures derived from prostate cancer cell-line LNCaP (9 samples) and normal cell-line RWPE (9 samples), in total 18 samples. **28**: Organoids from primary prostate epithelial cells in mono and co-culture with stromal cells (7 samples), stromal cells in co-culture (4 samples) and macrodissected tumor tissue (2 samples), in total 13 samples. **29**: Cells from benign prostatic bulk (3 + 3 samples), basal (3 + 3 samples) and luminal (3 + 3 samples) tissue cultured in two different media, in total 18 samples. **30**: Mouse prostate organoids with WT (6 samples) and mutated (6 samples) gene SPOP, in total 12 samples. **31**: Prostate tumor tissue from mouse models with different genetic modifications, in total 384 samples. See also [Figures S7–S10](#).

gene expression level, and this study supports the hypothesis that changes in genes responsible for citrate and spermine secretion are important for the cell transformations leading to cancer. Further research into these mechanisms can lead to new discoveries and treatment targets in the management of prostate cancer.

The challenge of finding a model system for normal prostate

We could not find CS signature enrichment comparable to tissue samples in any of the cell-lines and model systems we investigated. Androgen responsive cell-lines seem to maintain some traces of prostate specific functions compared to other model systems, but their CS signature enrichments are far below that of *in vivo* tissue and indicate that most parts of the prostate-lineage specificity is lost. Thus, these observations question the relevance of these model systems for studying normal prostate function and transitions from normal cells to prostate cancer cells. The secretory function can possibly be triggered by proper stimuli like dihydrotestosterone (DHT), testosterone or prolactin ([Costello et al., 1995](#); [Franklin et al., 1995](#); [Liu et al., 1997](#)). For example, LNCaP (but not PC3) cells were able to secrete citrate when stimulated by DHT ([Franklin et al., 1995](#)). However, it was also observed that the rate of citrate consumption by the TCA cycle increased proportionally in the same experiment ([Franklin et al., 1995](#)). Because these

experiments were performed before the era of transcriptomics, it would be interesting to repeat these experiments with accompanying transcriptomics analysis to identify the genes that change during such stimuli.

Signature refinement produces more robust gene signatures with improved integrity

In this manuscript, we have argued that improved gene signature integrity (assessed by CMS) can be achieved by integrating data from multiple prostate cancer datasets. The aim was to remove noise and produce better and more robust gene signatures in terms of biological interpretation. We found several evidences that more robust signatures were achieved by integration of datasets. First, we observed that the refined signature improved CMS scores compared to the initial signature in independent normal prostate samples (Figure 2D). Second, when we used the initial gene signature for ssGSEA in the *Hsu* dataset, we observed no clear separation between prostate and other metastatic samples (Figures 4D and S11). Third, in the *FANTOM* dataset, the one Prostate Adult Tissue sample were only ranked as the highest scoring sample after signature refinement, but not when the initial signature was used (Figures 3D and S12).

The refinement procedure generated an improved CS signature ssGSEA scores which was particularly pronounced for prostate samples. We also tested the effect of refinement on gene expression data from other tissue types in the *TCGA-complete* dataset (33 cancer and 22 normal tissue types). Other cancer types showed both elevated and decreased GSEA scores after refinement, and elevations were consistently lower than in prostate samples (Figure S13).

The best CS signature ssGSEA score improvement after refinement were also observed particularly for the metabolites citrate and spermine. For this test, we redid the procedure to create initial and refined signatures for all metabolites and lipids in the metabolite data from *Bertilsson* (23 metabolites and 17 lipid signals). We then calculated ssGSEA scores in all cancer and normal tissue types in the *TCGA-complete* dataset for each metabolite/lipid signature (Figures S14–S17). The most elevated ssGSEA scores were for citrate and spermine in prostate compared to other tissue types, whereas other metabolites showed both elevated and decreased ssGSEA score for other tissue types compared to prostate. This observation was similar for both cancer and normal tissue samples. Thus, the effect of refinement was highest for citrate and spermine specifically in the prostate. Interestingly, we noted that other metabolites/lipid signatures were elevated in prostate, particularly putrescine (a precursor for spermine in the polyamine pathway) a few lipid signals and glucose, indicating prostate specific regulation of other metabolites than just citrate and spermine. These observations were confirmed in the *FANTOM* dataset (dataset ID 23) (Figure S18).

In summary, we conclude that the identification followed by refinement procedure used in this study created a robust and unbiased gene signature biologically relevant for *in vivo* spermine and citrate secretion in the prostate. In a wider perspective, the module approach may represent a general strategy to find robust genesets related to other –omics data (for example metabolomics, proteomics or lipidomics), as long as these-omics data are accompanied with gene expression data.

Conclusion

The genes and mechanisms governing normal prostate function, in particular the ability to accumulate and secrete large amounts of citrate and spermine, are mostly unknown. We have used bioinformatics on an extensive collection on prostate datasets to discover genes relevant for secretion of citrate and spermine in the prostate. Based on this we have created a 150 gene signature which can be used to assess the degree of citrate and spermine secretion in a single sample. The bioinformatics approach enabled us to validate and explore the gene signature in a large number of contexts based on public data, including normal versus cancer enrichment, spatial colocalization on tissue, tumor grade, metastasis and different model systems. The employed approach is generalizable to other types of data when measured together with gene expression. The signature showed an inverse association to cancer progression from normal through low and high grade to metastasis. The genes in the signature were enriched for zinc-binding, mitochondrial function, respiratory complex, secretion through the ER –golgi – lysosome – exosome pathway and branched chain amino acid catabolism, pointing to both known and unexplored mechanisms for normal prostate function. The lack of suitable model systems is a challenge and need to be established to further study these findings.

Limitations of the study

Because this is a purely computational study, the conclusions are based on correlative and associative methods, which do not directly give information on causality. To verify that these association represent relevant prostate cancer biology we have 1) Ensured that the correlations and association are highly robust, and can be verified in 31 datasets with samples related to prostate and prostate cancer. 2) That the results are functionally meaningful according to what is known about the prostate biology and citrate/spermine secretion. When this is established, we further conclude that new targets and mechanisms highlighted in the same framework are also likely robust and functionally meaningful. The actual causality is difficult to pinpoint at this stage, and would need further laboratory experiments with relevant model systems.

In this study we have compared datasets from different sources in the same analysis. Such comparisons may not be able to remove all individual biases in the datasets. However, similar trends observed repeatedly across several datasets will minimize the effect of individual biases. We also demonstrate that ssGSEA results are highly comparable across tissue types and model system when comparing the datasets from Taylor (dataset ID 3) and Prensner (dataset ID 10) (Figure S10).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Module refinement procedure
 - Correlation Module Score (CMS)
 - Refinement procedure - Stage 1
 - Refinement procedure - Stage 2
 - Refinement procedure - Stage 3
 - Single sample Gene Set Enrichment Analysis (ssGSEA)
 - Adapted and normalized ssGSEA
 - GO analysis with GAPGOM
 - Network analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104451>.

ACKNOWLEDGMENTS

This work is supported by the Liaison Committee between the Central Norway Regional Health Authority (RHA) and the Norwegian University of Science and Technology (NTNU) to [MBR]; the Norwegian Cancer Society to [TFB] and [MBT] and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 758306) to [MBT]. The technique for fresh frozen tissue biobanking and cylinder extraction for reference E-MTAB-1041 was developed by Biobank1, St. Olavs Hospital, Trondheim, Norway. Fundingsupport for MPC Transcriptome sequencing to identify non-coding RNAs in prostate cancer was provided through the NIH Prostate SPORE P50CA69568, R01 R01CA132874, the Early Detection Research Network (U01 CA111275), the Department of Defense grant W81XWH-11-1-0331 and the National Center for Functional Genomics (W81XWH-11-1-0520). The results shown here are in part based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>. The Geno-type-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 11/23/2017.

AUTHOR CONTRIBUTIONS

MBR, MBT, and HB conceived the idea and developed the concept. MBR, MBT, and HB performed data curation. MBR developed methods and performed overall analysis. FD and CM developed and performed the GAPGOM analysis. MH performed network analysis. MBR, MBT, and TFB acquired funding. MBR, MBT, SB, FD and TFB performed supervision and provided intellectual input. MBR, MBT, SB, HB, FD, and TFB prepared the original draft. MBR prepared figures. All authors wrote and reviewed the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 4, 2022

Revised: April 12, 2022

Accepted: May 17, 2022

Published: June 17, 2022

REFERENCES

- Ananieva, E.A., and Wilkinson, A.C. (2018). Branched-chain amino acid metabolism in cancer. *Curr. Opin. Clin. Nutr. Metab. Care* 21, 64–70. <https://doi.org/10.1097/MCO.0000000000000430>.
- Aryee, M.J., Liu, W.N., Engelmann, J.C., Nuhn, P., Gurel, M., Haffner, M.C., Esopi, D., Irizarry, R.A., Getzenberg, R.H., Nelson, W.G., et al. (2013). DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Sci. Transl. Med.* 5, 169ra10. <https://doi.org/10.1126/scitranslmed.3005211>.
- Aytes, A., Mitrofanova, A., Lefebvre, C., Alvarez, M.J., Castillo-Martin, M., Zheng, T., Eastham, J.A., Gopalan, A., Pienta, K.J., Shen, M.M., et al. (2014). Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* 25, 638–651. <https://doi.org/10.1016/j.ccr.2014.03.017>.
- Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. <https://doi.org/10.1038/nrg1272>.
- Berglund, E., Maaskola, J., Schultz, N., Friedrich, S., Marklund, M., Bergenstrahle, J., Tarish, F., Tanoglidis, A., Vickovic, S., Larsson, L., et al. (2018). Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun.* 9, 2419. <https://doi.org/10.1038/s41467-018-04724-5>.
- Bertilsson, H., Tessem, M.B., Flatberg, A., Viset, T., Gribbestad, I., Angelsen, A., and Halgunset, J. (2012). Changes in gene transcription underlying the aberrant citrate and choline metabolism in human prostate cancer samples. *Clin. Cancer Res.* 18, 3261–3269. <https://doi.org/10.1158/1078-0432.Ccr-11-2929>.
- Bettuzzi, S., Davalli, P., Astancolle, S., Carani, C., Madeo, B., Tampieri, A., and Corti, A. (2000). Tumor progression is accompanied by significant changes in the levels of expression of polyamine metabolism regulatory genes and clusterin (sulfated glycoprotein 2) in human prostate cancer specimens. *Cancer Res.* 60, 28–34.
- Blattner, M., Liu, D.L., Robinson, B.D., Huang, D., Poliakov, A., Gao, D., Nataraj, S., Deonaraine, L.D., Augello, M.A., Sailer, V., et al. (2017). SPOP mutation drives prostate tumorigenesis *in vivo* through coordinate regulation of PI3K/mTOR and AR signaling. *Cancer Cell* 31, 436–451. <https://doi.org/10.1016/j.ccell.2017.02.004>.
- Braadland, P.R., Giskeodegard, G., Sandmark, E., Bertilsson, H., Euceda, L.R., Hansen, A.F., Guldvik, I.J., Selnaes, K.M., Grytli, H.H., Katz, B., et al. (2018). Erratum: *ex vivo* metabolic fingerprinting identifies biomarkers predictive of prostate cancer recurrence following radical prostatectomy. *Br. J. Cancer* 118, e11. <https://doi.org/10.1038/bjc.2017.470>.
- Cai, C.M., Wang, H.Y., He, H.H., Chen, S., He, L.F., Ma, F., Mucci, L., Wang, Q.B., Fiore, C., et al. (2013). ERG induces androgen receptor-mediated regulation of SOX9 in prostate cancer. *J. Clin. Invest.* 123, 1109–1122. <https://doi.org/10.1172/Jci66666>.
- Chandran, U.R., Ma, C.Q., Dhir, R., Bisceglia, M., Lyons-Weiler, M., Liang, W.J., Michalopoulos, G., Becich, M., and Monzon, F.A. (2007). Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. *BMC Cancer* 7, 64. <https://doi.org/10.1186/1471-2407-7-64>.
- Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinf.* 14, 128. <https://doi.org/10.1186/1471-2105-14-128>.
- Cheng, L.L., Wu, C., Smith, M.R., and Gonzalez, R.G. (2001). Non-destructive quantitation of spermine in human prostate tissue samples using HRMAS 1H NMR spectroscopy at 9.4 T. *FEBS Lett.* 494, 112–116. [https://doi.org/10.1016/S0014-5793\(01\)02329-8](https://doi.org/10.1016/S0014-5793(01)02329-8).
- Costello, L.C., Akuffo, V., and Franklin, R.B. (1988). Testosterone stimulates net citrate production from aspartate by prostate epithelial cells. *Horm. Metab. Res.* 20, 252–253. <https://doi.org/10.1055/s-2007-1010807>.
- Costello, L.C., and Franklin, R.B. (2000). The intermediary metabolism of the prostate: a key to understanding the pathogenesis and progression of prostate malignancy. *Oncology* 59, 269–282. <https://doi.org/10.1159/000012183>.
- Costello, L.C., and Franklin, R.B. (2016). A comprehensive review of the role of zinc in normal prostate function and metabolism; and its implications in prostate cancer. *Arch. Biochem. Biophys.* 611, 100–112. <https://doi.org/10.1016/j.abb.2016.04.014>.
- Costello, L.C., Franklin, R.B., and Narayan, P. (1999). Citrate in the diagnosis of prostate cancer. *Prostate* 38, 237–245. [https://doi.org/10.1002/\(sici\)1097-0045\(19990215\)38:3<237::aid-pros8>3.0.co;2-o](https://doi.org/10.1002/(sici)1097-0045(19990215)38:3<237::aid-pros8>3.0.co;2-o).
- Costello, L.C., Liu, Y., and Franklin, R.B. (1995). Testosterone stimulates the biosynthesis of m-aconitase and citrate oxidation in prostate epithelial cells. *Mol. Cell. Endocrinol.* 112, 45–51. [https://doi.org/10.1016/0303-7207\(95\)03582-r](https://doi.org/10.1016/0303-7207(95)03582-r).
- Costello, L.C., Liu, Y., Zou, J., and Franklin, R.B. (2000). Mitochondrial aconitase gene expression is regulated by testosterone and prolactin in prostate epithelial cells. *Prostate* 42, 196–202.
- Erlind, M.P., and Hsieh, A.C. (2014). PI3K-AKT-mTOR signaling in prostate cancer progression and androgen deprivation therapy resistance. *Asian J. Androl.* 16, 378–386. <https://doi.org/10.4103/1008-682X.122876>.
- Ehsani, R., and Drablos, F. (2018). Measures of co-expression for improved function prediction of long non-coding RNAs. *BMC Bioinf.* 19, 533. <https://doi.org/10.1186/s12859-018-2546-y>.
- Erho, N., Crisan, A., Vergara, I.A., Mitra, A.P., Ghadessi, M., Buerki, C., Bergstralh, E.J., Kollmeyer, T., Fink, S., Haddad, Z., et al. (2013). Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS One* 8, e66855. <https://doi.org/10.1371/journal.pone.0066855>.
- Franklin, R.B., and Costello, L.C. (1984). Glutamate dehydrogenase and a proposed glutamate-aspartate pathway for citrate synthesis in rat ventral prostate. *J. Urol.* 132, 1239–1243. [https://doi.org/10.1016/s0022-5347\(17\)50113-5](https://doi.org/10.1016/s0022-5347(17)50113-5).

- Franklin, R.B., Juang, H.H., Zou, J., and Costello, L.C. (1995). Regulation of citrate metabolism by androgen in the LNCaP human prostate carcinoma cell line. *Endocrine* 3, 603–607. <https://doi.org/10.1007/BF02953026>.
- Franklin, R.B., Milon, B., Feng, P., and Costello, L.C. (2005). Zinc and zinc transporters in normal prostate function and the pathogenesis of prostate cancer. *Front. Biosci. Landmark* 10, 2230–2239. <https://doi.org/10.2741/1692>.
- Garrett, S.H., Sens, M.A., Shukla, D., Flores, L., Somji, S., Todd, J.H., and Sens, D.A. (2000). Metallothionein isoform 1 and 2 gene expression in the human prostate: downregulation of MT-1X in advanced prostate cancer. *Prostate* 43, 125–135. [https://doi.org/10.1002/\(sici\)1097-0045\(20000501\)43:2<125::aid-pros7>3.0.co;2-s](https://doi.org/10.1002/(sici)1097-0045(20000501)43:2<125::aid-pros7>3.0.co;2-s).
- Ghandi, M., Huang, F.W., Jane-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the cancer cell line Encyclopedia. *Nature* 569, 503–508. <https://doi.org/10.1038/s41586-019-1186-3>.
- Giskeodegard, G.F., Bertilsson, H., Selnæs, K.M., Wright, A.J., Bathen, T.F., Viset, T., Halgunset, J., Angelsen, A., Gribbestad, I.S., and Tessem, M.B. (2013). Spermine and citrate as metabolic biomarkers for assessing prostate cancer aggressiveness. *PLoS One* 8, e62375. <https://doi.org/10.1371/journal.pone.0062375>.
- Henry, G.H., Malewska, A., Joseph, D.B., Malladi, V.S., Lee, J., Torrealba, J., Mauck, R.J., Gahan, J.C., Raj, G.V., Roehrborn, C.G., et al. (2018). A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell Rep.* 25, 3530–3542.e5. <https://doi.org/10.1016/j.celrep.2018.11.086>.
- Hsu, S.D., Kim, M.K., Foye, A., Silvestri, A., Lyerly, H.K., Morse, M., Petricoin, E., and Febbo, P.G. (2010). Use of gene expression signatures to identify origin of primary and therapeutic strategies for patients with advanced solid tumors. *J. Clin. Oncol.* 28, 10504. https://doi.org/10.1200/jco.2010.28.15_suppl.10504.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. <https://doi.org/10.1093/nar/gkn923>.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. <https://doi.org/10.1038/nprot.2008.211>.
- Jia, Z., Wang, Y., Sawyers, A., Yao, H., Rahmatpanah, F., Xia, X.Q., Xu, Q., Pio, R., Turan, T., Koziol, J.A., et al. (2011). Diagnosis of prostate cancer using differentially expressed genes in stroma. *Cancer Res.* 71, 2476–2487. <https://doi.org/10.1158/0008-5472.CAN-10-2585>.
- Juang, H.H., Costello, L.C., and Franklin, R.B. (1995). Androgen modulation of multiple transcription start sites of the mitochondrial aspartate aminotransferase gene in rat prostate. *J. Biol. Chem.* 270, 12629–12634. <https://doi.org/10.1074/jbc.270.21.12629>.
- Klijn, C., Durinck, S., Stawiski, E.W., Haverly, P.M., Jiang, Z.S., Liu, H.B., Degenhardt, J., Mayba, O., Gnad, F., Liu, J.F., et al. (2015). A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* 33, 306–312. <https://doi.org/10.1038/nbt.3080>.
- Kolenko, V., Teper, E., Kutikov, A., and Uzzo, R. (2013). Zinc and zinc transporters in prostate carcinogenesis. *Nat. Rev. Urol.* 10, 219–226. <https://doi.org/10.1038/nrurol.2013.43>.
- Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97. <https://doi.org/10.1093/nar/gkw377>.
- Kuner, R., Falth, M., Pressinotti, N.C., Brase, J.C., Puig, S.B., Metzger, J., Gade, S., Schafer, G., Bartsch, G., Steiner, E., et al. (2013). The maternal embryonic leucine zipper kinase (MELK) is upregulated in high-grade prostate cancer. *J. Mol. Med.* 91, 237–248. <https://doi.org/10.1007/s00109-012-0949-1>.
- Lin, S.F., Wei, H., Maeder, D., Franklin, R.B., and Feng, P. (2009). Profiling of zinc-altered gene expression in human prostate normal vs. cancer cells: a time course study. *J. Nutr. Biochem.* 20, 1000–1012. <https://doi.org/10.1016/j.jnutbio.2008.09.004>.
- Liu, Y., Franklin, R.B., and Costello, L.C. (1997). Prolactin and testosterone regulation of mitochondrial zinc in prostate epithelial cells. *Prostate* 30, 26–32. [https://doi.org/10.1002/\(sici\)1097-0045\(19970101\)30:1<26::aid-pros4>3.0.co;2-j](https://doi.org/10.1002/(sici)1097-0045(19970101)30:1<26::aid-pros4>3.0.co;2-j).
- Lizio, M., Abugessaisa, I., Noguchi, S., Kondo, A., Hasegawa, A., Hon, C.C., de Hoon, M., Severin, J., Oki, S., Hayashizaki, Y., et al. (2019). Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res.* 47, D752–D758. <https://doi.org/10.1093/nar/gky1099>.
- Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Sahin, S., Abugessaisa, I., Hori, F., Ishikawa-Kato, S., Arner, E., Baillie, J.K., et al. (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 16, 22. <https://doi.org/10.1186/s13059-014-0560-6>.
- Lynch, M.J., and Nicholson, J.K. (1997). Proton MRS of human prostatic fluid: correlations between citrate, spermine, and myo-inositol levels and changes with disease. *Prostate* 30, 248–255. [https://doi.org/10.1002/\(sici\)1097-0045\(19970301\)30:4<248::aid-pros4>3.0.co;2-h](https://doi.org/10.1002/(sici)1097-0045(19970301)30:4<248::aid-pros4>3.0.co;2-h).
- Markert, E.K., Mizuno, H., Vazquez, A., and Levine, A.J. (2011). Molecular classification of prostate cancer using curated expression signatures. *Proc. Natl. Acad. Sci. U S A* 108, 21276–21281. <https://doi.org/10.1073/pnas.1117029108>.
- Mortensen, M.M., Hoyer, S., Lynnerup, A.S., Orntoft, T.F., Sorensen, K.D., Borre, M., and Dyrskjot, L. (2015). Expression profiling of prostate cancer tissue delineates genes associated with recurrence after prostatectomy. *Sci. Rep.* 5, 16018. <https://doi.org/10.1038/srep16018>.
- Penney, K.L., Sinnott, J.A., Tyekucheva, S., Gerke, T., Shui, I.M., Kraft, P., Sesso, H.D., Freedman, M.L., Loda, M., Mucci, L.A., and Stampfer, M.J. (2015). Association of prostate cancer risk variants with gene expression in normal and tumor tissue. *Cancer Epidemiol. Biomark. Prev.* 24, 255–260. <https://doi.org/10.1158/1055-9965.Epi-14-0694-T>.
- Poisson, L.M., Sreekumar, A., Chinnaiyan, A.M., and Ghosh, D. (2012). Pathway-directed weighted testing procedures for the integrative analysis of gene expression and metabolomic data. *Genomics* 99, 265–274. <https://doi.org/10.1016/j.ygeno.2012.03.004>.
- Prensner, J.R., Iyer, M.K., Balbin, O.A., Dhanasekaran, S.M., Cao, Q., Brenner, J.C., Laxman, B., Asangani, I.A., Grasso, C.S., Kominsky, H.D., et al. (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* 29, 742–749. <https://doi.org/10.1038/nbt.1914>.
- Rhodes, D.R., Barrette, T.R., Rubin, M.A., Ghosh, D., and Chinnaiyan, A.M. (2002). Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res.* 62, 4427–4433.
- Richards, Z., McCray, T., Marsili, J., Zenner, M.L., Manlucu, J.T., Garcia, J., Kajdacsy-Balla, A., Murray, M., Voisine, C., Murphy, A.B., et al. (2019). Prostate stroma increases the viability and maintains the branching phenotype of human prostate organoids. *iScience* 12, 304–317. <https://doi.org/10.1016/j.isci.2019.01.028>.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43, e47. <https://doi.org/10.1093/nar/gkv007>.
- Ross-Adams, H., Lamb, A.D., Dunning, M.J., Halim, S., Lindberg, J., Massie, C.M., Egevad, L.A., Russell, R., Ramos-Montoya, A., Vowler, S.L., et al. (2015). Integration of copy number and transcriptomics provides risk stratification in prostate cancer: adiscovery and validation cohort study. *EBioMedicine* 2, 1133–1144. <https://doi.org/10.1016/j.ebiom.2015.07.017>.
- Rye, M.B., Bertilsson, H., Andersen, M.K., Rise, K., Bathen, T.F., Drablos, F., and Tessem, M.B. (2018). Cholesterol synthesis pathway genes in prostate cancer are transcriptionally downregulated when tissue confounding is minimized. *BMC Cancer* 18, 478. <https://doi.org/10.1186/s12885-018-4373-y>.
- Sboner, A., Demichelis, F., Calza, S., Pawitan, Y., Setlur, S.R., Hoshida, Y., Perner, S., Adami, H.O., Fall, K., Mucci, L.A., et al. (2010). Molecular sampling of prostate cancer: a dilemma for predicting disease progression. *BMC Med. Genom.* 3, 8. <https://doi.org/10.1186/1755-8794-3-8>.
- Si, M., and Lang, J. (2018). The roles of metallothioneins in carcinogenesis. *J. Hematol. Oncol.* 11, 107. <https://doi.org/10.1186/s13045-018-0645-x>.
- Sogaard, C.K., Moestue, S.A., Rye, M.B., Kim, J., Nepal, A., Liabakk, N.B., Bachke, S., Bathen, T.F.,

- Otterlei, M., and Hill, D.K. (2018). APIM-peptide targeting PCNA improves the efficacy of docetaxel treatment in the TRAMP mouse model of prostate cancer. *Oncotarget* **9**, 11752–11766. <https://doi.org/10.18632/oncotarget.24357>.
- Tayari, N., Heerschap, A., Scheenen, T.W.J., and Kobus, T. (2017). In vivo MR spectroscopic imaging of the prostate, from application to interpretation. *Anal. Biochem.* **529**, 158–170. <https://doi.org/10.1016/j.ab.2017.02.001>.
- Taylor, B.S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B.S., Arora, V.K., Kaushik, P., Cerami, E., Reva, B., et al. (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22. <https://doi.org/10.1016/j.ccr.2010.05.026>.
- Tessem, M.B., Bertilsson, H., Angelsen, A., Bathen, T.F., Drablos, F., and Rye, M.B. (2016). A balanced tissue composition reveals new metabolic and gene expression markers in prostate cancer. *PLoS One* **11**, e0153727. <https://doi.org/10.1371/journal.pone.0153727>.
- Tomlins, S.A., Mehra, R., Rhodes, D.R., Cao, X.H., Wang, L., Dhanasekaran, S.M., Kalyana-Sundaram, S., Wei, J.T., Rubin, M.A., Pienta, K.J., et al. (2007). Integrative molecular concept modeling of prostate cancer progression. *Nat. Genet.* **39**, 41–51. <https://doi.org/10.1038/ng1935>.
- Wang, Y., Xia, X.Q., Jia, Z., Sawyers, A., Yao, H., Wang-Rodriguez, J., Mercola, D., and McClelland, M. (2010). In silico estimates of tissue components in surgical samples based on expression profiling data. *Cancer Res.* **70**, 6448–6455. <https://doi.org/10.1158/0008-5472.CAN-10-0021>.
- Yu, Y.P., Landsittel, D., Jing, L., Nelson, J., Ren, B.G., Liu, L.J., McDonald, C., Thomas, R., Dhir, R., Finkelstein, S., et al. (2004). Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J. Clin. Oncol.* **22**, 2790–2799. <https://doi.org/10.1200/Jco.2004.05.158>.
- Zhang, D.X., Lin, K., Lu, Y., Rycaj, K., Zhong, Y., Chao, H.P., Calhoun-Davis, T., Shen, J.J., and Tang, D.G. (2017a). Developing a novel two-dimensional culture system to enrich human prostate luminal progenitors that can function as a cell of origin for prostate cancer. *Stem Cell Transl. Med.* **6**, 748–760. <https://doi.org/10.5966/sctm.2016-0243>.
- Zhang, S., Zeng, X., Ren, M., Mao, X., and Qiao, S. (2017b). Novel metabolic and physiological functions of branched chain amino acids: a review. *J. Anim. Sci. Biotechnol.* **8**, 10. <https://doi.org/10.1186/s40104-016-0139-z>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Dataset ID 1 – <i>Bertilsson</i> : 156 prostate tissue samples (116 cancer and 40 normal) 129 of these samples (95 cancer and 34 normal) with concordant metabolite and gene expression measurement	Bertilsson et al. (2012)	Array Express: E-MTAB-1041
Dataset ID 2 – <i>Wang</i> : 136 prostate tissue samples 65 cancer and 71 normal	Wang et al. (2010) , Jia et al. (2011)	Gene Expression Omnibus: GEO: GSE8218
Dataset ID 3 – <i>Taylor</i> : 160 prostate tissue samples (131 cancer and 29 normal) 19 prostate metastatic samples 4 cell-line samples (DU145, PC3, LNCaP, VCaP)	Taylor et al. (2010)	Gene Expression Omnibus: GEO: GSE21034
Dataset ID 4 – <i>Sboner</i> : 281 prostate cancer samples	Sboner et al. (2010)	Gene Expression Omnibus: GEO: GSE16560
Dataset ID 5 – <i>Erho</i> : 545 prostate cancer samples	Erho et al. (2013)	Gene Expression Omnibus: GEO: GSE46691
Dataset ID 6 – <i>TCGA-PRAD</i> : 549 prostate tissue samples (497 cancer and 52 normal)	https://www.cancer.gov/tcga	https://portal.gdc.cancer.gov/repository
Dataset ID 7 – <i>CMBR (Cambridge)</i> : 186 prostate tissue samples (112 cancer and 74 normal)	Ross-Adams et al. (2015)	Gene Expression Omnibus: GEO: GSE70768
Dataset ID 8 – <i>STCK (Stockholm)</i> : 94 prostate cancer samples	Ross-Adams et al. (2015)	Gene Expression Omnibus: GEO: GSE70769
Dataset ID 9 – <i>Mortensen</i> : 50 prostate tissue samples (36 cancer and 14 normal) Laser dissected tissue	Mortensen et al. (2015)	Gene Expression Omnibus: GEO: GSE46602
Dataset ID 10 – <i>Prensner</i> : 116 prostate tissue samples (78 cancer and 38 normal) 12 prostate metastatic samples 58 cell-line samples	Prensner et al. (2011)	Gene Expression Omnibus: GEO: GSE31728 (Cell-lines) dbGap: phs000443.v1.p1 (tissue and metastatic samples)
Dataset ID 11 – <i>Penney</i> : 424 prostate tissue samples (264 cancer and 160 normal)	Penney et al. (2015)	Gene Expression Omnibus: GEO: GSE62872
Dataset ID 12 – <i>Kuner</i> : 98 prostate tissue samples (59 cancer and 39 normal)	Kuner et al. (2013)	Gene Expression Omnibus: GEO: GSE32571
Dataset ID 13 – <i>Aryee</i> : 21 normal prostate samples 18 prostate metastatic samples	Aryee et al. (2013)	Gene Expression Omnibus: GEO: GSE38241
Dataset ID 14 – <i>Chandran</i> : 145 prostate tissue samples (65 cancer and 81 normal) 25 prostate metastatic samples	Chandran et al. (2007) , Yu et al. (2004)	Gene Expression Omnibus: GEO: GSE6919 Data from GPL8300 Plat-form
Dataset ID 15 – <i>Cai</i> : 29 prostate cancer samples 22 prostate metastatic samples	Cai et al. (2013)	Gene Expression Omnibus: GEO: GSE32269
Dataset ID 16 – <i>Poisson</i> : 28 prostate tissue samples (12 cancer and 16 normal) 13 prostate metastatic samples	Poisson et al. (2012)	Gene Expression Omnibus: GEO: GSE8511
Dataset ID 17 – <i>Monzon</i> : 10 prostate cancer samples 21 prostate metastatic samples	Chandran et al. (2007)	Gene Expression Omnibus: GEO: GSE6752

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dataset ID 18 –Tomlins: 84 prostate tissue samples (32 cancer, 13 PIN lesions, 27 normal epithelium and 12 normal stroma) 20 prostate metastatic samples Laser dissected tissue	Tomlins et al. (2007)	Gene Expression Omnibus: GEO: GSE6099
Dataset ID 19 –Hsu: 96 metastatic samples from various cancers (9 from prostate metastasis)	Hsu et al. (2010)	Gene Expression Omnibus: GEO: GSE18549
Dataset ID 20 –Berglund: Spatial transcriptomics on 12 slices from prostate cancer and normal tissue	Berglund et al. (2018)	http://www.spatialtranscriptomicsresearch.org/
Dataset ID 21 –TCGA-complete: 11,093 samples from 33 cancer types	https://www.cancer.gov/tcga	https://portal.gdc.cancer.gov/repository
Dataset ID 22 –GTEx Median profiles for 53 normal tissue types	https://gtexportal.org/home/	https://gtexportal.org/home/datasets
Dataset ID 23 –FANTOM 1829 samples from various human cell-lines and tissues	Lizio et al. (2019) , Lizio et al. (2015)	fantom.gsc.riken.jp/5/data/
Dataset ID 24 –CCLE:1019 cell-line samples from various cancer types	Ghandi et al. (2019)	https://portals.broadinstitute.org/ccle/data
Dataset ID 25 –E-MTAB-2706: 622 cell-line samples from various cancer types	Klijn et al. (2015)	RNA-Seq data Array Express: E-MTAB-2706
Dataset ID 26 –Sogaard: 36 samples from 3 prostate cancer cell-lines: DU145, PC3 and LNCaP	Sogaard et al. (2018)	Array Express: E-MTAB-4858
Dataset ID 27 –Su: 18 samples: 2D and 3D cultures (spheroids) derived from prostate cancer cell-line LNCaP (9 samples), and prostate normal cell-line RWPE (9 samples)	No article reference found	Gene Expression Omnibus: GEO: GSE30304
Dataset ID 28 –Zachary: 13 samples: Organoids from primary prostate epithelial cells in mono and co-culture with stromal cells (7 samples), stromal cells in co-culture (4 samples) and macrodissected tumor tissue (2 samples)	Richards et al. (2019)	Gene Expression Omnibus: GEO: GSE115052
Dataset ID 29 –Zhang: 18 samples: Cells from benign prostatic bulk (3 + 3 samples), basal (3 + 3 samples) and luminal (3 + 3 samples) tissue cultured in two different media	Zhang et al. (2017a)	Gene Expression Omnibus: GEO: GSE74698
Dataset ID 30 –Blattner: 12 samples: Mouse prostate organoids with WT (6 samples) and mutated (6 samples) gene SPOP	Blattner et al. (2017)	Gene Expression Omnibus: GEO: GSE94839
Dataset ID 31 –Aytes: 384 samples: Prostate tumor tissue from mouse models with different genetic modifications	Aytes et al. (2014)	Gene Expression Omnibus: GEO: GSE53202
Software and algorithms		
Python version 2.7	Python Software Foundation	https://www.python.org
Limma 3.26.9 with edgeR 3.12.1	Bioconductor 3.1 R 3.2.2	https://bioconductor.org/packages/release/bioc/html/limma.html
Code for gene signature refinement procedure	This paper	https://zenodo.org/record/6303619#.YnuoQqam19A

RESOURCE AVAILABILITY

Lead contact

The lead contact for this study is Morten Beck Rye: morten.rye@ntnu.no.

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data: This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the [key resources table/supplemental information](#).

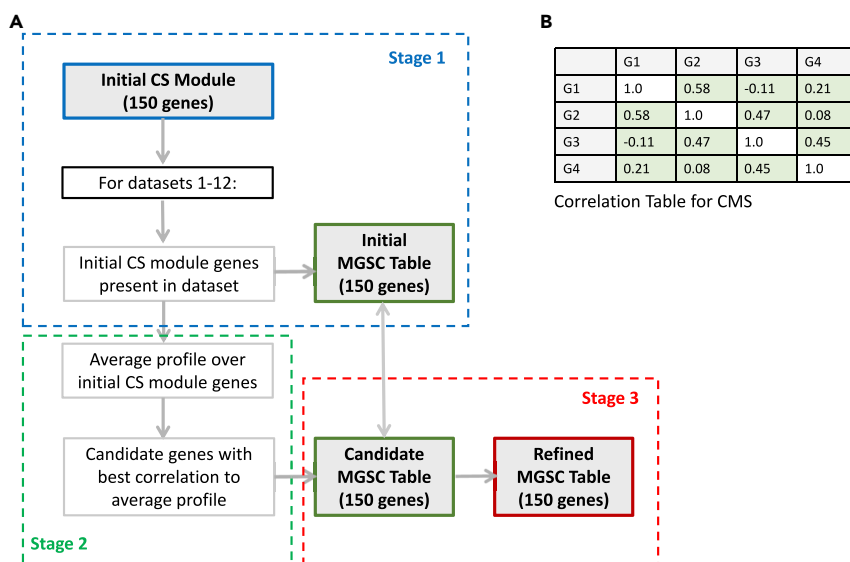
Code: Example code is available at Zenodo with accession <https://zenodo.org/record/6303619#.YnuoQqam19A>.

Any additional information required to reanalyze the data reported in this paper is available from Morten Beck Rye (morten.rye@ntnu.no— lead contact) upon request.

METHOD DETAILS

Module refinement procedure

A schematic representation of the procedure with stages 1-3 is shown in below figure. The procedure makes use of the following abbreviations: **M** = Module, **G** = several genes, **g** = single gene, **D** = Dataset, **P** = gene expression Profile, **TB** = Table, **av** = average, **in** = initial, **rf** = refined, **c** = candidate, **m** = missing, **cr** = correlation, **ic** = intra-correlation, **CMS** = Correlation Module Score (CMS), **MGCR** = Module Gene Contribution Score (MGCR).



Module refinement procedure and Correlation Module Score (CMS)

(A) Flowchart of the module refinement procedure. Each stage 1-3 are explained in the main text.

(B) Example of a correlation table used to calculate the Correlation Module Score (CMS). This gene module has four genes (G1-G4) and a CMS of 0.32. See also [Figures S11–S18](#) for performance of the module refinement procedure.

Correlation Module Score (CMS)

To assess how well the correlations between genes in a module (or geneset) is preserved for all (or a subset of) samples in a particular dataset, we created the Correlation Module Score (abbreviated CMS). The CMS for a module in a dataset is calculated by first creating a table of intra-correlations between expression profiles for all genes in the module for that dataset (Figure). Here the expression profile for a gene is the vector consisting of gene expression values overall samples in that dataset. For a 150-gene module, this will create

a table with $150 \times 150 = 22,500$ correlation values (a table for a hypothetical 4-gene module is shown in Figure). The CMS is then calculated as the average overall correlations included the table (but excluding self-correlations). A high CMS represents a high integrity of the gene module in the dataset, indicating the preservation of strong correlations between the module-genes in the datasets. Likewise, a low CMS would indicate weak relations between the genes in the module, and the loss of gene module integrity. We further assume if a gene module has a high CMS in a dataset, it would indicate that the biological process the module represents is important or active for samples in this dataset. In this study the datasets are patient cohorts with normal and cancer tissue samples from the prostate.

Refinement procedure - Stage 1

The purpose of Stage 1 in the module refinement procedure is to find the contribution of each gene in the initial CS module from Bertilsson to the overall CMS for the module. To represent this contribution we introduce the Module Gene Contribution Score (MGCS), which is the average overall correlations between that gene and each of the other genes in the module (or, the overall contribution of that gene to the overall CMS for the module). In figure this will be the average of the rows (excluding the diagonal), where each average value represent the MGCS score for a gene. A high MGCS score will indicate that the expression profile of the gene fits better with expression profile of other genes in the module, while a low MGCS score indicates a gene expression profile that deviates from other genes in the module.

Procedure:

For each dataset **D** (dataset ID 1–12 with normal and cancer tissue samples from the prostate):

Find the genes from the initial CS module, **in-M-G**, present in the dataset **D**, **in-M-G-D**.

Calculate the intra-correlation table, **ic-TB-D**, for the initial CS module gene profiles, **in-M-G-P-D**, in dataset **D**.

Calculate the corresponding CMS score, **CMS-D**, from the intra-correlation table, **ic-TB-D**, in dataset **D**.

Calculate the MGCS score for each initial CS gene **in-g** in **CMS-D**, **in-g-MGSC-D**.

For missing initial CS module genes, **m-in-M-G-D**, in dataset **D**: Set the **in-g-MGSC-D** to a *dataset specific missing gene constant* equal to the **CMS-D** multiplied by 0.5.

Store the **in-g-MGSC-D** for each initial CS module gene in each dataset **D** in an *Initial MGSC table*, **in-MGSC-TB** for comparison in Stage 3.

Store MGSC for missing genes in each cohort, **m-MGSC**, for Stage 2.

Refinement procedure - Stage 2

The purpose of Stage 2 is to calculate, for all candidate genes **c-g** present in any one of the 12 datasets **D**, a score which describes how well each candidate gene expression profile fit with the expression profiles of genes in the initial CS module. This is done by comparing each candidate gene expression profile **c-g-P-D** to the average expression profile, **in-M-G-P-D-av**, of the genes in the initial CS module for each dataset **D**. We regard any gene present in any of the 12 datasets **D** as candidate genes.

Procedure:

Define candidate genes, **c-G**, by the union of all genes present in any dataset **D**.

For each dataset **D**:

Find the genes from the initial CS module, **in-M-G**, present in dataset **D**.

Calculate the average expression profile, **in-M-G-P-D-av**, over expression profiles for all initial module genes **in-M-G-P-D** in dataset **D**.

For each gene g present in dataset D :

Calculate the correlation between the expression profile of gene g and the average profile in-M-G-P-D-av for dataset D .

For genes not present in this dataset, the correlation is set to the *cohort specific missing gene constant* in m-MGSC from **Stage 1**.

Store all correlations in a table, cr-TB , for evaluation in **Stage 3**.

Refinement procedure - Stage 3

The purpose of Stage 3 is to identify the genes that correlates best with genes in the initial CS module over all 12 datasets, but that were not a part of the initial CS module to begin with. We combine the 150 best new candidate genes with the initial 150 CS module genes from **Stage 1**, and sort the resulting 300 genes according to their MGSC score. The top 150 genes are selected for the refined CS module. Thus, new candidate genes will replace initial genes if they have higher MGSC scores. The final output is a refined CS module, where the robustness of gene correlations across all 12 prostate cancer datasets are taken into consideration in the selection of genes for the module. This refined module is defined as the CS gene signature.

Procedure:

Exclude initial CS module genes in-M-G from the correlation table cr-TB from **Stage 2**.

Calculate the average correlation, c-g-cr-av , for each of the remaining candidate genes overall 12 datasets in cr-TB .

Sort genes in correlation table cr-TB by average correlation c-gr-cr-av .

Select 150 candidate genes, c-g-150 , with the highest average correlation, c-gr-cr-av .

For each dataset D :

For each of the top 150 candidate genes, c-g-150 :

Calculate the average correlation of gene c-g-150 to all genes from the initial CS module, in-M-G-D , and use the average correlation as the candidate gene MGSC, c-g-MGSC-D . This is a measure of the fit for the candidate gene, c-g-150 , to the initial CS module in dataset D .

For candidate genes not present in dataset D , the MGSC is set to the *cohort specific missing gene constant* in m-MGSC from **Stage 1**.

Combine the candidate MGSC table c-g-MGSC-D with the initial MGSC table in-g-MGSC-D from **Stage 1**. The combined table, comb-TB , will contain 300 genes, 150 from the initial CS module and 150 new candidates.

Calculate the average MGSC score, g-MGSC-av , for each gene overall 12 datasets for the combined table comb-TB .

Sort genes by average MGSC score, g-MGSC-av .

Select the 150 genes with the highest average MGSC score, g-MGSC-av , as genes for the refined CS module, rf-G-M .

Example Code for the refinement procedure is available at Zenodo with accession <https://zenodo.org/record/6303619#.YnuoQqam19A>.

Single sample Gene Set Enrichment Analysis (ssGSEA)

Single sample Gene Set Enrichment Analysis (ssGSEA) was performed by sorting the genes in each sample according to their expression level in descending order. Then a ssGSEA score was calculated for each sample implementing a previously published algorithm (Markert et al., 2011). We did not perform centering and normalization of the expression values for ssGSEA in this study. This is because citrate and spermine secretion, as well as the content of stroma tissue in the sample, are dominating biological features of these samples where differences should be visible at absolute expression level. However, we observed good correlations (0.86–0.97) for ssGSEA scores based on normalized and absolute expression values. A few of the datasets were only presented with centered expression values in the public resource, and for these we used centered values for ssGSEA. For the spatial transcriptomics data from Berglund (dataset ID 20), we did use centered and normalized expression for ssGSEA in each pixel. This was necessary due to many pixels with a high number of unexpressed genes, and/or very few genes with abundant expression.

Adapted and normalized ssGSEA

The magnitude of GSEA scores depend on both the size of the gene set used, and the number of total genes in the ranked list input to the GSEA calculations. To be able to compare ssGSEA scores from *Premsner* to other datasets, we adapted the ssGSEA by only using the genes shared between the datasets in each comparison. In this way, both the number and identity of genes, as well as the size of the geneset will be identical for both datasets. For the comparison of several datasets in Figure 4D, we also normalized the GSEA scores in *Premsner* to 0–1 scale for each comparison. The adapted and normalized score calculated in each comparison were highly correlated (average 0.99), and had a small standard deviation of the mean (0.004), and we conclude that they are comparable between the datasets.

GO analysis with GAPGOM

The GO analysis was based on gene annotation generated with the GAPGOM tool (Ehsani and Drablos, 2018) [PMID: 30567492, GAPGOM in Bioconductor]. GAPGOM will predict gene annotation for a target gene by identification of well-annotated genes showing correlated expression pattern with the target gene across several experiments, and then estimate a predicted annotation for the target gene as a consensus over the co-expressed genes, based on the hypothesis that co-expressed genes may be involved in similar or related processes. This prediction of annotation terms may facilitate a richer functional annotation of genes, in particular for genes where there is a lack of experimental annotation data.

Since GAPGOM only allows us to do annotation prediction one gene at a time, GAPGOM was used in conjunction with Snakemake [https://snakemake.readthedocs.io/en/stable/project_info/citations.html] to create a software pipeline. This pipeline was used to predict GO term annotation for all 150 genes in the CS-signature on each of dataset 1–12. Since GO annotation consists of three different ontologies (MF, BP, CC), each prediction is done separately for each. This then outputs a list for each gene in each dataset for each GO ontology, in total 1800 lists. Each list contains one or multiple GO-term predictions, with the following information for each prediction; GO-ID, Ontology, p-value, FDR/q-value (Bonferoni normalized p-value), The description of the GO term, and the used correlation method for the prediction. For each list, only GO terms with a q-value > 0.05 were selected. This generated a total of 21,080 GO terms for all lists, where 5354 GO terms were unique. A table was created with the number of genes as rows and the number of unique GO-terms as columns. For each gene and GO-term, we summed $-\log_{10}$ q-values from all datasets, producing an overall score for each GO-term and gene. We made three tables, one for all 150 genes, one for the six Metallothioneins, and one for the 10 network Hub-genes.

Network analysis

Correlation-based networks of the CS signature genes for each dataset are created as follows: The 150 genes are represented as nodes in the network and the pairwise Pearson correlation between genes represents the interactions between nodes. To reduce the complexity and highlight the most powerful interactions, only the 20 strongest outgoing links (in absolute Pearson correlation) from each node are kept. In this way, the most central nodes will have 20 outgoing links and up to 130 ingoing links from other nodes, while some of the non-central nodes will only have 20 links. By calculating the node-degree (number of links to other nodes) for each node, genes with high node-degree reflects central genes for driving the biological processes (Barabasi and Oltvai, 2004). This network construction is performed on cancer samples for all

12 datasets, and the 10 genes with the highest mean degree across the datasets are considered as the top 10 hub genes.

QUANTIFICATION AND STATISTICAL ANALYSIS

Figures 1A and 1B: Correlation coefficients was calculated using Pearson correlation. Figure 1C: The citrate and spermine concentration profiles was both normalized to sum = 1 before averaging. Figure 2: Definition and calculation of CMS is described in Method details. p-values for Statistical significance of CMS was calculated using the lognormal distribution in python 2.7 – scipy.stats.lognorm. Correlation between metabolite profile and gene profiles were calculated using Pearson correlation. Ranking of genes by importance is described in Method details. Figures 3A, 3B, 3C, and 3E: Calculation of ssGSEA scores is described in Method details. Figure 3D: Correlation coefficients are calculated using Pearson correlation. Figure 4: Calculation of ssGSEA scores is described in Method details. Figure 4A: ssGSEA scores were centered around zero for each dataset for improved visual comparison. Figure 4B: ssGSEA scores were scaled to values between 0 (lowest) and 1 (highest) in each dataset for improved visual comparison. The boxplots were created using matplotlib.pyplot.boxplot with default settings and layout. Figure 5A: Correlation coefficients was calculated using Pearson correlation. Calculation of ssGSEA scores is described in Method details. Figure 6: Calculation of ssGSEA scores and Adapted/Normalized ssGSEA scores is described in Method details. The boxplots were created using matplotlib.pyplot.boxplot with default settings and layout. Tables 1 and 2: p-values was calculated using the limma package in R. Correlation to tumor content in Table 1 was calculated using Pearson correlation. Differential expression analysis between high- and low-grade cancer samples (dataset ID 1–8 and 12, Table S7) was performed in R using limma for microarray data, and limma-voom for RNA-Seq data (Ritchie et al., 2015).