Supplementary Information

Table of contents

S1. Methods

S1.1. Plasma samples

Blood from localised PC patients was drawn in an EDTA tube by the anaesthetist prior to surgery, and centrifuged at 2,500g for 10 minutes at room temperature to separate the plasma which was then aliquoted and stored at -80°C. These localised PC plasma samples are referred to as Set 1 samples.

Blood from mHSPC and mCRPC (validation cohort) patients was collected in a tube with EDTA and 3.2% buffered sodium citrate as anti-coagulants, and centrifuged at 3,000 rpm for 10 minutes at 4°C. The supernatant was removed and centrifuged again, and the platelet-poor plasma was aliquoted and stored at -80°C. These mHSPC and mCRPC plasma samples, and 975 other plasma samples that are not part of this study are referred to as Set 2 samples.

Lipidomic analysis (lipid extraction and liquid chromatography-mass spectrometry (LC-MS)) of Set 1 and Set 2 samples are performed separately, approximately 2 years apart, using the same methodology (Huynh *et al* 2019)[1] but with different LC-MS instruments as described below.

S1.2. Lipid extraction

For both Set 1 and 2 samples, lipids were extracted from 10μl of plasma using a butanol/methanol extraction method as described by Alshehry *et al* (2015)[2]. Internal standards were added to the plasma prior to extraction (listed in Huynh *et al* 2019)[1]. These internal standards are used to calculate the concentration of the lipids from the LC-MS data, by relating the peak area of each species to the peak area of the corresponding internal standard.

S1.3. Quality controls

Replicates of two types of quality controls (QC) were extracted and run together with the study plasma samples:

- Pooled human plasma from healthy individuals (PQC)
- National Institute of Standards and Technology human plasma standard reference material 1950 (NIST1950). This was developed by NIST from a collaboration with the National Institute of Health (NIH), and the National Institute of Diabetes and Digestive and Kidney Disease (NIDDK), to allow comparisons between data sets run within the same laboratory and with other laboratories globally.

A total of 40 replicate PQC and 10 replicate NIST1950 samples were analysed with Set 1 samples; 63 replicate PQC and 24 replicate NIST1950 samples were analysed with the Set 2 samples. The coefficient of variation (%CV) of the lipid levels in these QC samples have passed the required threshold of mean %CV<15% and median %CV<10%.

S1.4. Liquid chromatography-mass spectrometry (LC-MS)

LC-MS analysis of Set 1 and Set 2 lipid extracts were performed as described by Huynh *et al* (2019)[1], except that an Applied Biosystems API 4000 Q/TRAP mass spectrometer with an Agilent 1200 liquid chromatography system was used for Set 1 samples, whereas the same LC-MS instrument described in Huynh *et al* was used for Set 2 lipid extracts which is an Agilent 6490 QQQ mass spectrometer with an Agilent 1290 series HPLC system.

Set 1 samples were all analysed in a single run on the LC-MS. Set 2 samples were analysed in 3 batch runs, where the batch differences were adjusted by median centering with PQC samples. The median concentration of each lipid species of the PQC samples in each batch was first calculated, and then used to derive a median center value for each lipid species (median center value = median of batch 1 / median of batch 1,2 or 3). The median center value of each batch was then multiplied to the concentration of lipids in the study samples, resulting in the alignment of all batches to the first batch.

The number of lipids detected and quantified in Set 1 and Set 2 samples were 609 and 772 respectively, with an overlap of 512 lipids. The measurements of the lipids in plasma samples of the study and QCs are provided in the supplementary Excel files "Lipid quantitation_localisedPC.xlsx" and "Lipid quantitation_metastaticPC.xlsx".

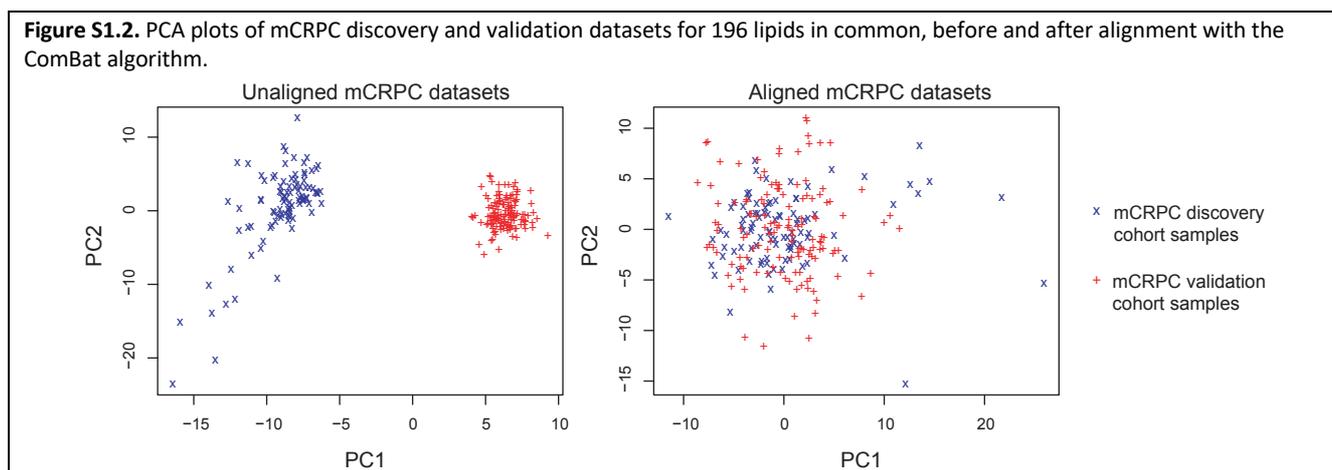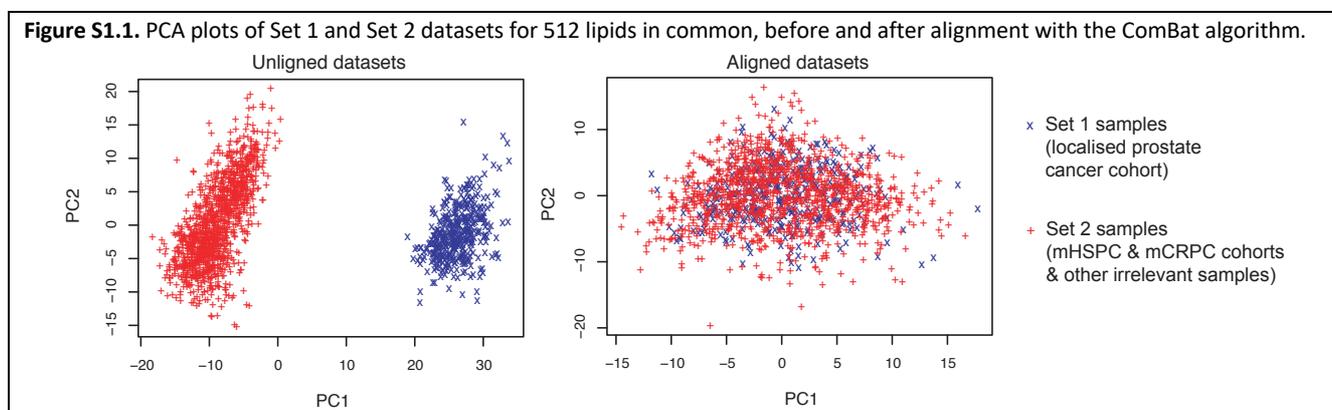### S1.5. Data normalization

The lipidomic datasets for Set 1 and Set 2 were normalized independently according to the Probabilistic Quotient (PBQ) normalization method as described in Lin *et al* (2017)[3] and adapted from Dietere *et al* (2006)[4]. Normalisation is a data pre-processing step that is essential for large scale analyses of multi-variable data (e.g.

genomic, proteomic). This normalisation step adjusts for biases that can arise from sample preparation (e.g. sample loss, evaporation, irregular extraction efficiency, pipetting errors), biological effects (e.g. differences in water content) or biological variation (e.g. differences in individuals unrelated to disease pathology). The reference sample used in PBQ normalization was created from the mean of the levels of each lipid species across all the plasma samples of each dataset respectively. Final values are logarithm-2 of pmol/ml.

S1.6. Dataset alignment

The normalized lipidomic datasets of Set 1 and Set 2 were aligned using the ComBat algorithm in the R package sva (v3.33.1)[5,6] for the 512 lipids in common, as PBQ normalisation was not sufficient to correct for LC-MS platform differences which was indicated by Principal Components Analysis (PCA)(Figure S1.1). Set 2 was aligned to Set 1 where Set 1 lipid levels remained unchanged. PCA and inspection of the total levels of each lipid subclass/class for each plasma sample confirmed that the algorithm was effective in removing batch differences (Figures S1.1 & S1.3). The normalised levels of the lipids that were not in common for each dataset were not altered, and were included in the statistical analyses of each respective datasets.

Similarly, the lipidomic datasets for the mCRPC validation cohort was aligned to the mCRPC discovery cohort using ComBat for 196 lipids in common prior to validation of the prognostic three-lipid signature, as the mCRPC discovery cohort was analysed with a different LC-MS platform described in Lin *et al* (2017)[3] (Figure S1.2).

Figure S1.1. PCA plots of Set 1 and Set 2 datasets for 512 lipids in common, before and after alignment with the ComBat algorithm.

Figure S1.2. PCA plots of mCRPC discovery and validation datasets for 196 lipids in common, before and after alignment with the ComBat algorithm.

S1.7. Latent class analysis

Unique lipid profiles were identified by latent class analysis (LCA) of the levels of prognostic lipids categorised into quartiles (R package poLCA v1.4.1). LCA is a non-supervised method that identifies class membership (lipid profiles) using the observed variables. The most parsimonious number of lipid profiles was determined with the minimum Akaike Information or Bayesian Information criterion.

Figure S1.3. Sum of normalised lipid levels by lipid subclass/class for each sample before and after alignment of Set 1 and Set 2 datasets with the ComBat algorithm.

Set 1 = Localised prostate cancer cohort; Set 2 = mHSPC, CRPC & other irrelevant samples
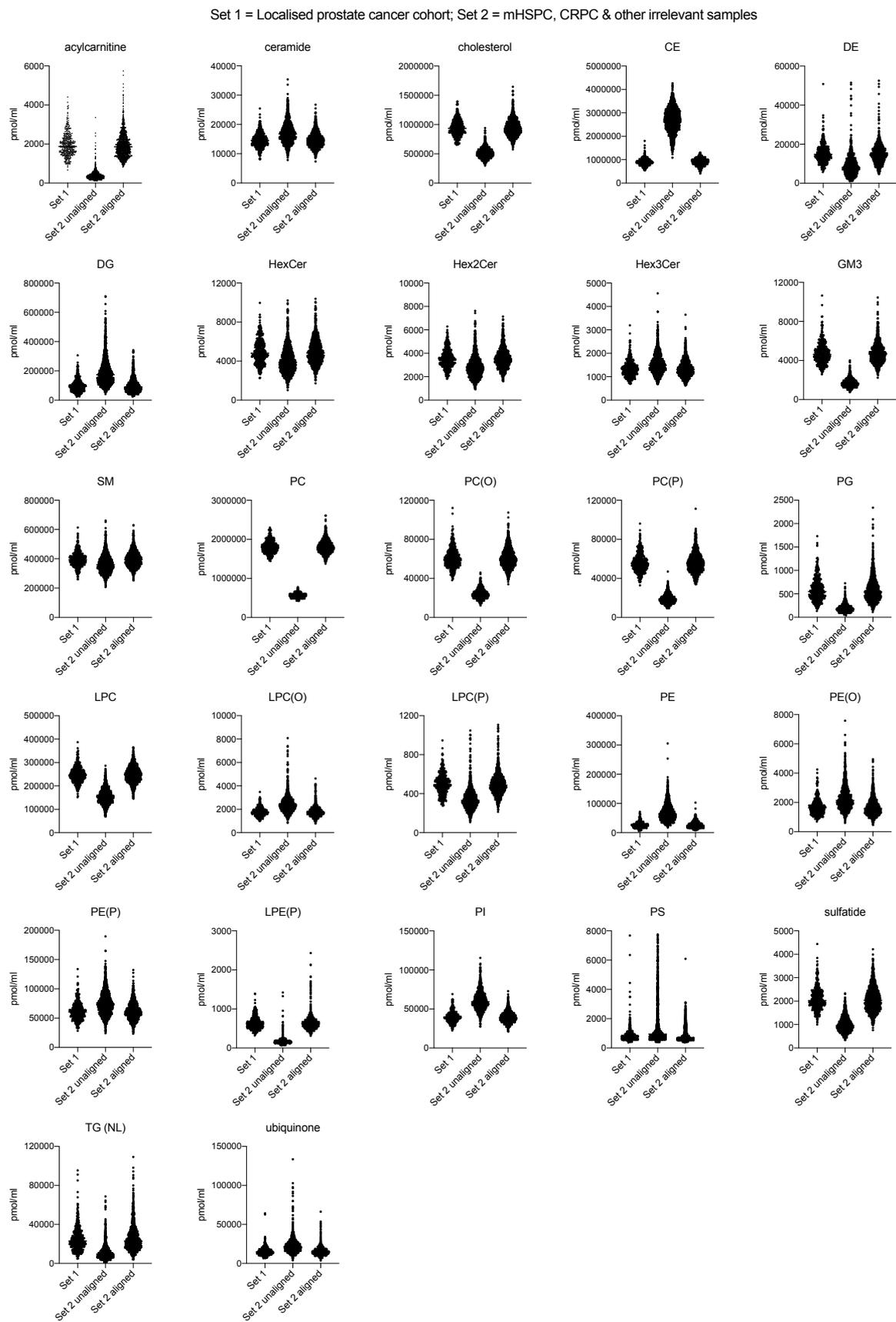
S1.8 Age of plasma sample age versus ceramide levels

The age of the plasma samples do not correlate with the total levels of ceramide or Cer(d18:1/24:1) indicating that age is not a confounding factor in the results (Figure S1.4).

Figure S1.4. Graphs of the levels of total ceramides or Cer(d18:1/24:1) in each individual plasma sample plotted against the age of the plasma sample.

(Lipid levels are not normalised; age = time from plasma collection to lipidomic analysis; R = Pearson coefficient).
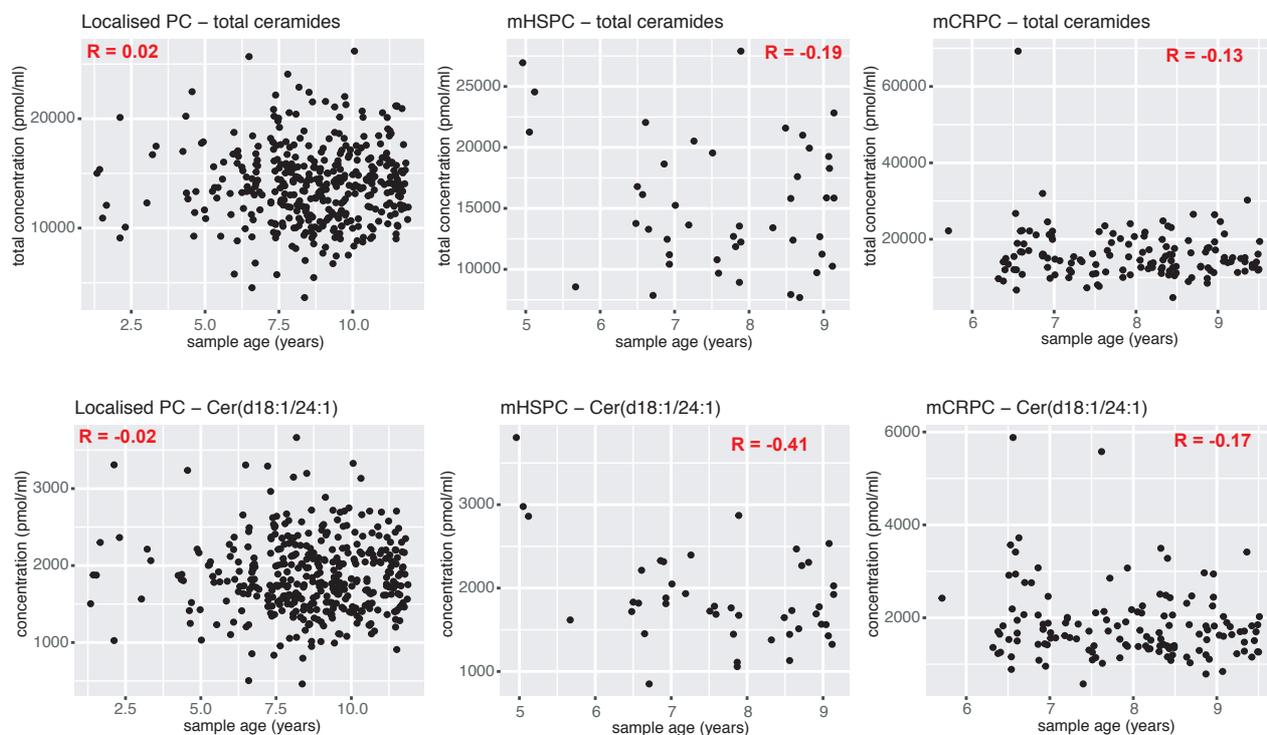
S2. Localised prostate cancer – Cox regression analyses

A significantly higher proportion of Profile L2 men had diabetes and body mass index (BMI) above the median or were on statin medication compared to the others (Chi-square P≤0.02, Table S2.1). However, none of these metabolic indicators except for diabetes (univariable Cox regression P=0.002), were significantly associated with metastatic relapse (Table S2.2). Multivariable Cox regression of these metabolic indicators with lipid profile showed that only diabetic status and lipid profile were independently associated with metastatic relapse (P≤0.008, Table S2.2 & S2.3). However, diabetes was not independently associated with metastatic relapse when modelled with lipid profile, Gleason score and P-stage (Model 2 of Table 2 in main article).

Table S2.1. Proportion of men with metabolic disorders among the lipid profile groups.

Metabolic factors	% of Profile 2 men	% of other profiles	Chi-Square P-value
Body mass index > median	72% (8% unknown)	44% (47% unknown)	0.01
Diabetic	11% (21% unknown)	3.4% (38% unknown)	0.02
Statin usage	66% (21% unknown)	25% (38% unknown)	1×10^{-7}
Hypertension	60% (23% unknown)	44% (38% unknown)	0.05

Table S2.2. Cox regression analyses of the association of metastatic relapse with lipid profile and metabolic factors in the localised prostate cancer cohort.

Metabolic factors	cases	events	Univariable Cox regression		Multivariable Cox regression	
			HR (95% CI)	P-value	HR (95% CI)	P-value
Lipid profile (Profile L2 vs others)	389	40	5.80 (3.04-11.1)	4×10^{-7}	5.52 (2.47-12.3)	3×10^{-5}
Diabetes	251	33	5.25 (1.81-15.2)	0.002	4.65 (1.51-14.4)	0.008
Statin usage	251	33	1.70 (0.85-3.37)	0.1	0.76 (0.34-1.68)	0.5
Hypertension	249	32	1.49 (0.74-3.00)	0.3	1.26 (0.61-2.62)	0.5

Table S2.3. Cox regression analyses of the association of metastatic relapse with lipid profile and body mass index (BMI) in the localised prostate cancer cohort.
BMI was not included in the above analysis (Tabel S2.2) as BMI was only known for 45% of the cohort.

Metabolic factors	cases	events	Univariable Cox regression		Multivariable Cox regression	
			HR (95% CI)	P-value	HR (95% CI)	P-value
Lipid profile (Profile L2 vs others)	389	40	5.80 (3.04-11.1)	4×10^{-7}	7.18 (2.94-17.5)	1×10^{-5}
BMI (continuous variable)	174	20	1.03 (0.93-1.15)	0.6	1.01 (0.89-1.14)	0.9

S3. Localised prostate cancer – AJCC TNM staging

Figure S3.1. Kaplan-Meier curves of localised PC grouped by the TNM staging system.

The TNM staging system by the American Joint Committee on Cancer (AJCC) classified the localised prostate cancer cohort into four groups – Stage I, IIA, IIB and III, where their Kaplan-Meier curves of metastatic relapse are shown in (A). Patients with the highest TNM staging in the cohort (Stage III) had a faster rate of metastatic relapse if they had lipid Profile L2 (B).

A TNM staging

TNM staging	N	events
I	9	0
IIA	18	0
IIB	178	5
III	184	35

B TNM staging Group IIB & III with stratification by lipid profile

TNM staging	Lipid profile	N	events
IIB	other	158	4
IIB	L2	20	1
III	other	151	20
III	L2	33	15

HR 6.30 (95% CI 3.12-12.72) $p\ 2\times10^{-6}$

Number at risk (A)

TNM staging							
IIB	178	178	176	142	76	15	0
III	184	169	154	114	50	11	0

Number at risk (B)

TNM staging	Lipid profile							
IIB	other	158	158	156	125	67	14	0
IIB	L2	20	20	20	17	9	1	0
III	other	151	142	134	101	46	10	0
III	L2	33	27	20	13	4	1	0

7

S4. CRPC validation cohort – Cox regression analyses & three-lipid signature

Table S4.1. Cox regression analyses of the association of overall survival with lipid profile, age and body mass index (BMI) in mCRPC.

Metabolic factors	cases	events	Univariable Cox regression		Multivariable Cox regression	
			HR (95% CI)	P-value	HR (95% CI)	P-value
Lipid profile (Profile C2 vs others)	137	122	2.54 (1.73-3.72)	2×10^{-6}	2.48 (1.65-3.72)	1×10^{-5}
Age (continuous variable)	137	122	1.03 (1.01-1.05)	0.01	1.01 (0.99-1.04)	0.3
BMI (continuous variable)	133	118	0.96 (0.92-1.00)	0.04	0.96 (0.93-1.00)	0.05

Table S4.2. Cox regression analyses of the association of overall survival with lipid profile and clinicopathological variables in mCRPC.

Metabolic factors	cases	events	Univariable Cox regression		Bivariable Cox regression			
					Model 1		Model 2	
			HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
Lipid profile (Profile C2 vs others)	137	122	2.54 (1.73-3.72)	2×10^{-6}	2.45 (1.65-3.64)	8×10^{-6}	2.22 (1.49-3.31)	1×10^{-4}
Alkaline phosphatase	132	118	1.01 (1.00-1.00)	0.03	1.00 (1.00-1.00)	0.1		
PSA	130	117	1.00 (1.00-1.00)	0.08			1.00 (1.00-1.00)	0.09
Lactase dehydrogenase	64	58	1.00 (1.00-1.01)	0.5				
Visceral metastases	45	44	1.15 (0.56-2.38)	0.7				
Haemoglobin	43	42	0.80 (0.66-0.97)	0.03				
Albumin	38	37	1.01 (1.00-1.03)	0.04				

Figure S4.1. Heatmap of sphingolipid levels in mCRPC men of discovery and validation cohort.

Men with the 3-lipid signature have higher levels of sphingolipids (except for 1-deoxydihydroceramide).
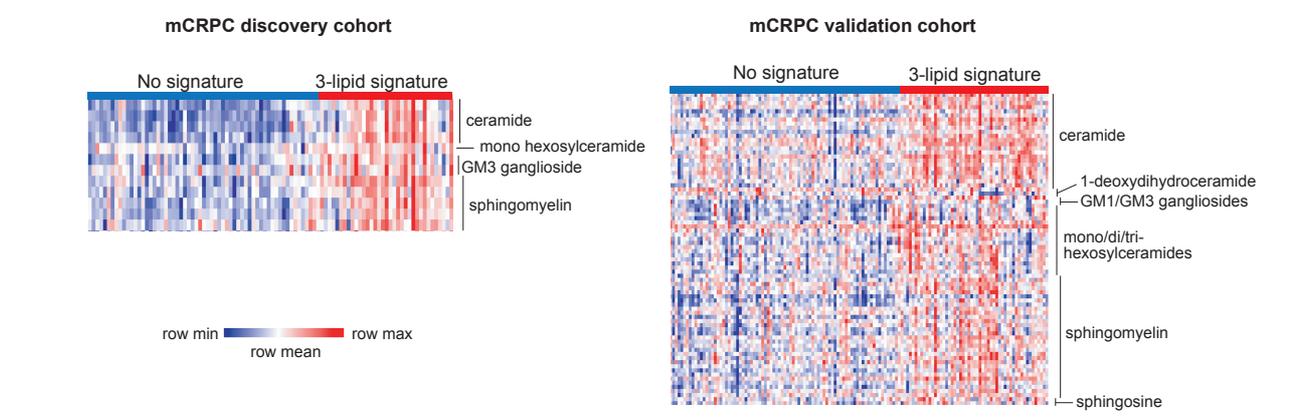

Table S4.3. Cox regression analyses of the 3-lipid signature and its individual lipids in the validation mCRPC cohort.

Ceramide(d18:1/24:1) alone was as prognostic as the three-lipid signature.

variables (137 cases, 122 events)	Variable type	Univariable Cox regression	
		Validation cohort	
		HR (95% CI)	P-value
3-lipid signature	Categorised: High risk (53 men) vs low risk (84 men)	2.39 (1.63-3.51)	1×10^{-5}
3-lipid signature	Continuous	4.44 (2.31-8.54)	1×10^{-5}
Ceramide(d18:1/24:1)	Continuous	3.19 (1.88-5.40)	4×10^{-5}
Sphingomyelin(d18:2/16:0)	Continuous	3.49 (1.59-7.65)	2×10^{-3}
Phosphatidylcholine(16:0/16:0)	Continuous	1.26 (0.56-2.86)	0.6

Figure S4.2. ROC analyses of 3-lipid signature and Cer(d18:1/24:1) in predicting 1 year in the validation mCRPC cohort.
The area under the curve (AUC) show that the 3-lipid signature perform better than Cer(d18:1/24:1) alone.

ROC analysis of 1 year survival - 137 men (113 alive)

AUC 0.664 (95% CI 0.536-0.753)   AUC 0.666 (95%CI 0.547-0.784)   AUC 0.625 (95%CI 0.502-0.747)
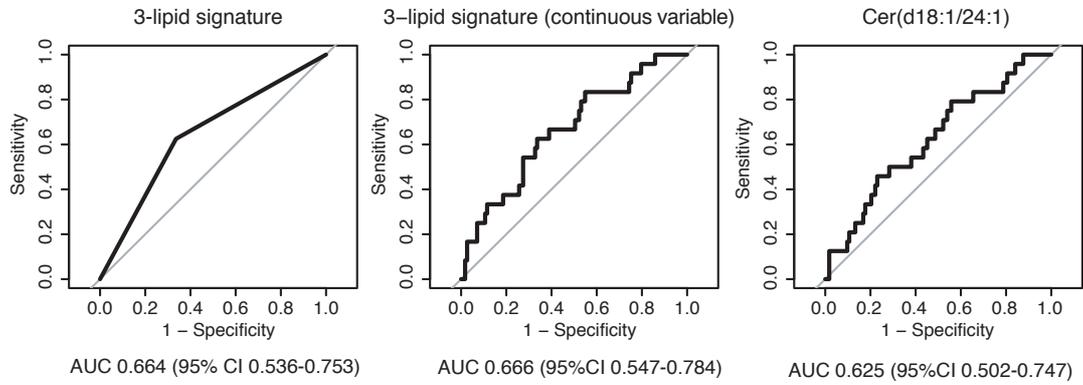
S5. Sphingolipids and their biological relevance to prostate cancer progression

Figure S5.1. Metabolism of ceramide and sphingomyelin by cancer cells into sphingosine-1-phosphate[7].

Hydrolysis of ceramides by ceramidases releases the fatty acyl chains and produces sphingosine, which is phosphorylated by sphingosine kinases (SPHK) into sphingosine-1-phosphate (S1P). Sphingomyelins can be converted into ceramides by sphingomyelinase. S1P mediates its effect by acting on G-protein coupled S1PR receptors in an autocrine or paracrine manner.
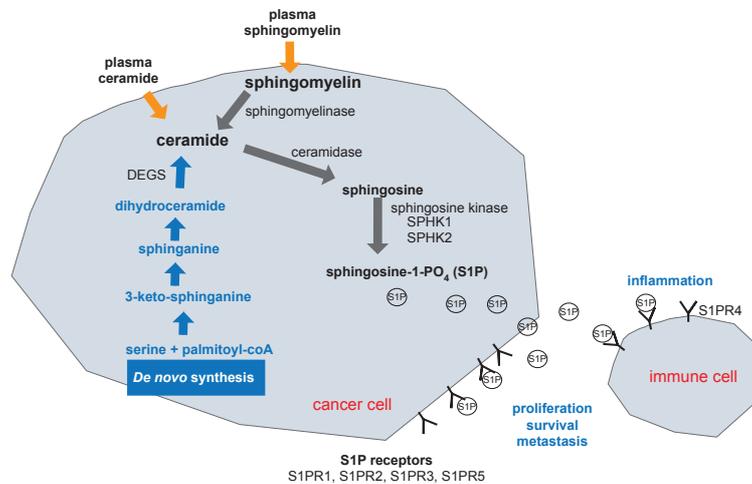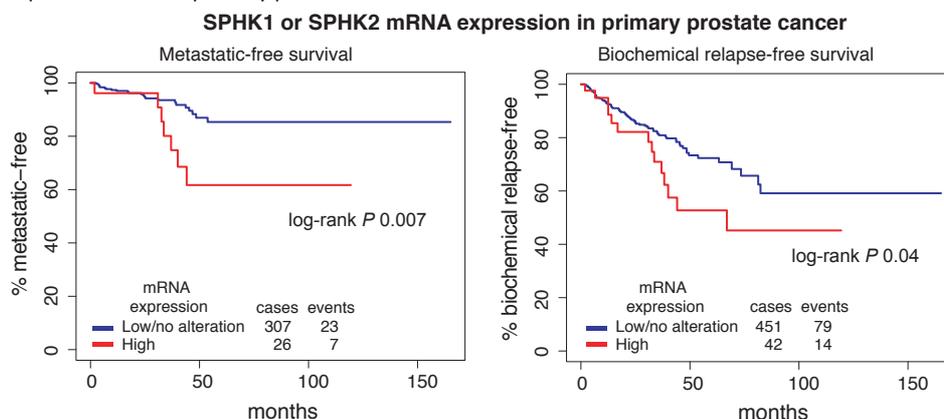

Figure S5.2 Higher expression of DEGS1, or sphingosine kinases SPHK1 or SPHK2, in localised prostate cancer is associated with higher rates of biochemical and metastatic relapse.

DEGS1 (delta 4-desaturase, sphingolipid 1) catalyses the insertion of the double bond in the sphingoid base of dihydroceramide to form ceramide in *de novo* ceramide synthesis, and is an androgen regulated gene[8]. SPHK1 and SPHK2 phosphorylates sphingosine into S1P (Figure S5.1). (Genomic data source: TCGA PanCan Atlas, cBioPortal accessed on 1 July 2020, clinical data last updated on 2018).

(A) mRNA expression of SPHK1, SPHK2 and DEGS1 in primary prostate tumours.
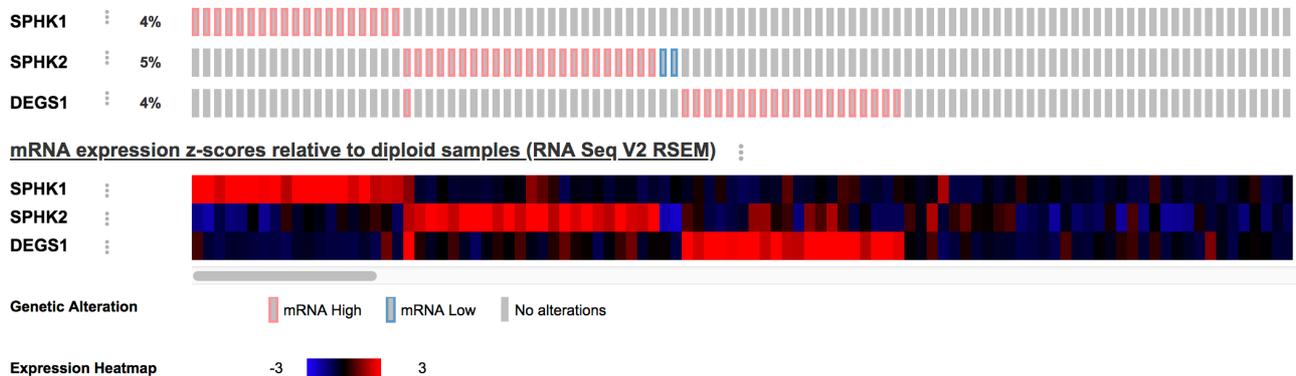
(B) Kaplan-Meir curves of biochemical and metastatic relapse for patients with high mRNA expression of SPHK1 or SPHK2 compared to those with low expression in their primary prostate tumours

(C) Kaplan-Meir curves of biochemical and metastatic relapse for patients with high mRNA expression of DEGS1 compared to those with low expression in their primary prostate tumours
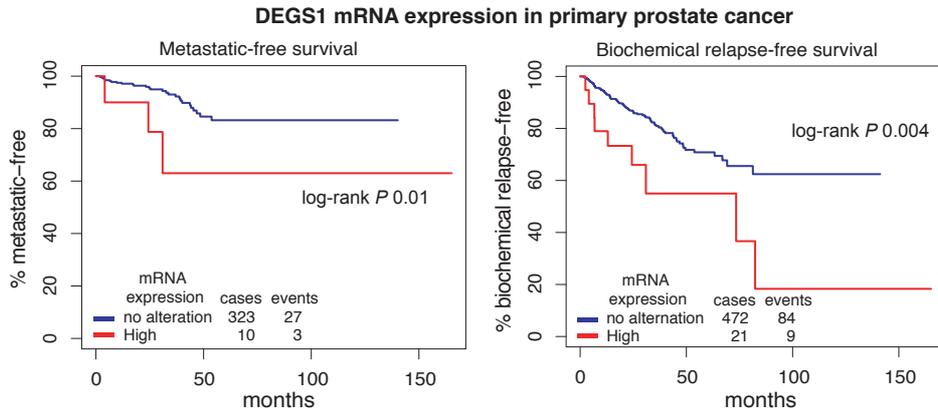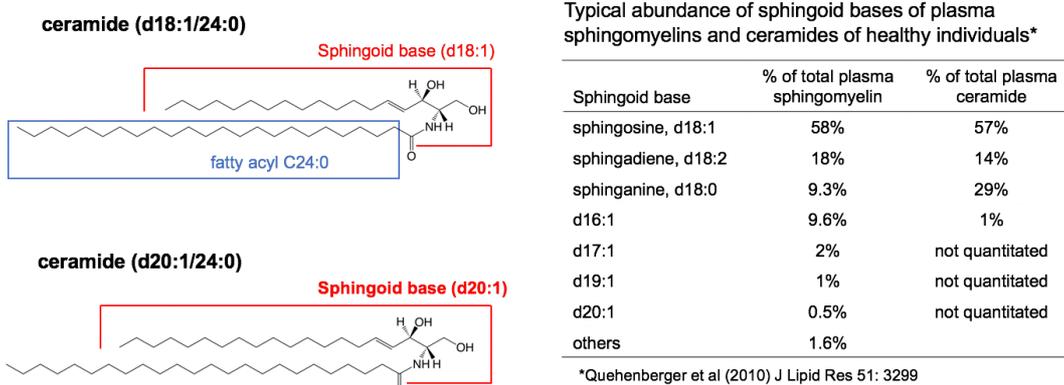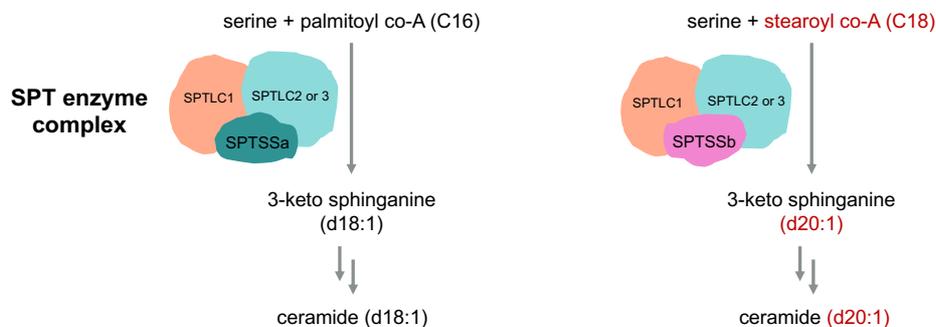
DEGS1 mRNA expression in primary prostate cancer

Metastatic-free survival

log-rank *P* 0.01

mRNA expression	cases	events
no alteration	323	27
High	10	3

Biochemical relapse-free survival

log-rank *P* 0.004

mRNA expression	cases	events
no alternation	472	84
High	21	9

Figure S5.3 Circulating ceramides with the d20:1 sphingoid base are likely to originate from prostate cancer cells instead of the liver.
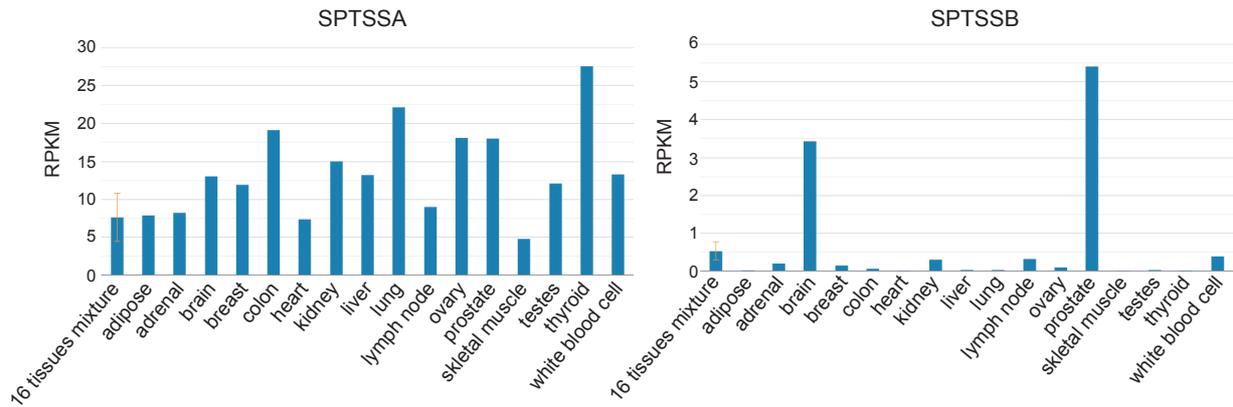
(A) The most common sphingoid base is sphingosine (d18:1), where 18 = carbon-chain length of 18, ':1' = one double bond, 'd' = two hydroxyl group. Other variations are present in plasma at lower levels. The ceramide with the d20:1 sphingoid base has a carbon-chain length of 20, and one double bond and two hydroxyl group. Dihydroceramides (sphingoid base d18:0), lack a double bond in the sphingoid base and are precursors of ceramides where the double bond is inserted by the enzyme DEGS1.

ceramide (d18:1/24:0)
Sphingoid base (d18:1)
fatty acyl C24:0

ceramide (d20:1/24:0)
Sphingoid base (d20:1)

Typical abundance of sphingoid bases of plasma sphingomyelins and ceramides of healthy individuals*

Sphingoid base	% of total plasma sphingomyelin	% of total plasma ceramide
sphingosine, d18:1	58%	57%
sphingadiene, d18:2	18%	14%
sphinganine, d18:0	9.3%	29%
d16:1	9.6%	1%
d17:1	2%	not quantitated
d19:1	1%	not quantitated
d20:1	0.5%	not quantitated
others	1.6%	

*Quehenberger et al (2010) J Lipid Res 51: 3299

(B) Ceramides with the d20:1 sphingoid base are formed from the condensation of serine with stearoyl-coA instead of palmitoyl-coA during *de novo* ceramide synthesis by serine palmitoyl transferase enzyme (SPT) that has the small subunit SPTSSB instead of SPTSSA[9].


serine + palmitoyl co-A (C16)

SPT enzyme complex
SPTLC1
SPTLC2 or 3
SPTSSa

3-keto sphinganine (d18:1)

ceramide (d18:1)

serine + stearoyl co-A (C18)

SPTLC1
SPTLC2 or 3
SPTSSb

3-keto sphinganine (d20:1)

ceramide (d20:1)

11

(C) SPTSSA is expressed ubiquitiously, whereas SPTSSB is only expressed by certain tissues such as prostate but not liver[10] (RPKM = reads per kilobase of transcript per million mapped reads; source: Illumina bodyMap2 transcriptome displayed in NCBI Gene: https://www.ncbi.nlm.nih.gov/gene/165679?report=expression; https://www.ncbi.nlm.nih.gov/gene/171546?report=expression)

(D) Primary prostate tumours and CRPC metastases have either increased mRNA expression or high level amplification of SPTSSB compared to SPTSSA (Source: TCGA, PanCancer Atlas; SU2C/PCF Dream Team study published in PNAS 2019[11]. Both datasets accessed on cBioPortal on 21 Jan 2020)
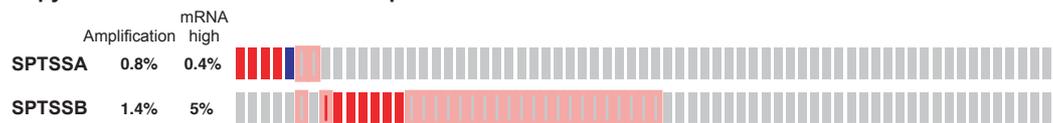
*mRNA expression measured in samples from 493 patients, of which 489 were analysed for copy number alterations. Amplification % is for these 489 patients.
**Copy number alterations measured in samples from 429 patients, of which 205 were analysed for mRNA expression. mRNA high % is for these 205 patients.

12

References

1   Huynh, K. *et al.* High-Throughput Plasma Lipidomics: Detailed Mapping of the Associations with Cardiometabolic Risk Factors. *Cell Chem Biol* **26**, 71-84.e74, (2019).

2   Alshehry, Z. H. *et al.* An Efficient Single Phase Method for the Extraction of Plasma Lipids. *Metabolites* **5**, 389-403, (2015).

3   Lin, H. M. *et al.* A distinct plasma lipid signature associated with poor prognosis in castration-resistant prostate cancer. *Int J Cancer* **141**, 2112-2120, (2017).

4   Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem* **78**, 4281-4290, (2006).

5   Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127, (2007).

6   Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883, (2012).

7   Ogretmen, B. Sphingolipid metabolism in cancer signalling and therapy. *Nat Rev Cancer* **18**, 33-50, (2018).

8   McNair, C. *et al.* Cell cycle-coupled expansion of AR activity promotes cancer progression. *Oncogene* **36**, 1655-1668, (2017).

9   Han, G. *et al.* Identification of small subunits of mammalian serine palmitoyltransferase that confer distinct acyl-CoA substrate specificities. *Proc Natl Acad Sci U S A* **106**, 8186-8191, (2009).

10   Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* **13**, 397-406, (2014).

11   Abida, W. *et al.* Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci U S A* **116**, 11428-11436, (2019).