

Glycosaminoglycan Profiling in Patients' Plasma and Urine Predicts the Occurrence of Metastatic Clear Cell Renal Cell Carcinoma

Francesco Gatto,¹ Nicola Volpi,² Helén Nilsson,³ Intawat Nookaew,^{1,7} Marco Maruzzo,⁴ Anna Roma,⁴ Martin E. Johansson,³ Ulrika Stiemer,⁵ Sven Lundstam,⁶ Umberto Basso,⁴ and Jens Nielsen^{1,*}

¹Department of Biology and Biological Engineering, Chalmers University of Technology, 41296 Göteborg, Sweden

²Department of Life Sciences, University of Modena and Reggio Emilia, 41125 Modena, Italy

³Department of Translational Medicine Malmö, Center for Molecular Pathology, Lund University, Skåne University Hospital, 20502 Malmö, Sweden

⁴Medical Oncology Unit 1, Istituto Oncologico Veneto IOV - IRCCS, 35128 Padova, Italy

⁵Department of Oncology, Institute of Clinical Sciences, Sahlgrenska Academy at the University of Gothenburg, Sahlgrenska University Hospital, 41345 Göteborg, Sweden

⁶Department of Urology, Sahlgrenska University Hospital and Sahlgrenska Academy, University of Gothenburg, 41345 Göteborg, Sweden

⁷Present address: Department Biomedical Informatics, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA

*Correspondence: nielsenj@chalmers.se

<http://dx.doi.org/10.1016/j.celrep.2016.04.056>

SUMMARY

Metabolic reprogramming is a hallmark of clear cell renal cell carcinoma (ccRCC) progression. Here, we used genome-scale metabolic modeling to elucidate metabolic reprogramming in 481 ccRCC samples and discovered strongly coordinated regulation of glycosaminoglycan (GAG) biosynthesis at the transcript and protein levels. Extracellular GAGs are implicated in metastasis, so we speculated that such regulation might translate into a non-invasive biomarker for metastatic ccRCC (mccRCC). We measured 18 GAG properties in 34 mccRCC samples versus 16 healthy plasma and/or urine samples. The GAG profiles were distinctively altered in mccRCC. We derived three GAG scores that distinguished mccRCC patients with 93.1%–100% accuracy. We validated the score accuracies in an independent cohort (up to 18 mccRCC versus nine healthy) and verified that the scores normalized in eight patients with no evidence of disease. In conclusion, coordinated regulation of GAG biosynthesis occurs in ccRCC, and non-invasive GAG profiling is suitable for mccRCC diagnosis.

INTRODUCTION

Clear cell renal cell carcinoma (ccRCC) is the most common form of kidney cancer (Rini et al., 2009), and it is responsible for 100,000 deaths worldwide (Ferlay et al., 2010). Approximately 50% of patients with ccRCC are expected to develop metastatic disease, which is usually incurable. In sharp contrast to early diagnosed ccRCC, the median survival of patients with metas-

tasis is significantly worse (Gupta et al., 2008), even with improved prognosis after the introduction of modern targeted therapies (Wahlgren et al., 2013). This fact constitutes a major and unmet clinical problem because, despite the need for both early prediction and frequent monitoring of metastatic ccRCC, no biomarkers are currently approved as part of the clinical management of the disease, resulting in late diagnosis or unknown responses to treatment (Jonasch et al., 2012; Moch et al., 2014).

The search for molecular biomarkers has focused on ccRCC genetics and angiogenesis, but none of these biomarkers has entered routine clinical practice, nor are they easily accessible or indicative of metastasis (Finley et al., 2011; Moch et al., 2014). In contrast, other molecular processes prominent in ccRCC might fill this gap. In this sense, accumulating evidence has suggested that the proliferation and survival of cancer cells rely upon a shift in their metabolism (Schulze and Harris, 2012; Vander Heiden et al., 2009; Ward and Thompson, 2012). In particular, ccRCC has recently been shown to feature strong regulation and dependence on distinctive metabolic reprogramming, which is pivotal to its progression (Creighton et al., 2013; Gatto et al., 2014, 2015; Hakimi et al., 2016; Li et al., 2014). These outstanding metabolic changes might be of clinical interest because they have the potential to be used as ccRCC biomarkers.

Under these premises, we followed up on our recent study, which revealed a deviating regulation of metabolism in ccRCC in contrast with seven common epithelial tumors (Gatto et al., 2014). Moreover, we further computationally characterized metabolic regulation of ccRCC, leveraging a larger number of samples using state-of-the-art genome-scale metabolic modeling (Bordbar et al., 2014; Jerby and Ruppin, 2012; Mardinoglu et al., 2013; Mardinoglu and Nielsen, 2015). As a result, we discovered previously unreported coordinated regulation of glycosaminoglycan (GAG) biosynthesis, which is exacerbated in metastasis. This discovery led us to speculate that this

regulation might be detectable in metastatic ccRCC. Hence, we designed an observational study to measure GAG profiles in the accessible fluids of metastatic ccRCC patients and sought to characterize the suitability of GAG profiles as diagnostic markers for the disease.

RESULTS

Metabolic Modeling Reveals Differential Regulation of Glycosaminoglycan Biosynthesis in Clear Cell Renal Cell Carcinoma versus Non-cancerous Kidneys

Our recent study suggested that metabolic reprogramming in ccRCC is unique and likely occurs due to genetic alterations in tumor progression (Gatto et al., 2014). The exceptional nature of metabolic regulation in ccRCC could have important clinical implications as a potential molecular biomarker. Thus, we sought to characterize fully metabolic regulation in ccRCC computationally. We retrieved a larger number of gene expression profiles from The Cancer Genome Atlas (TCGA) than in our previous study (481 tumor samples versus 71 tumor-adjacent normal samples, here simply referred to as “non-tumor,”; Table S1) and performed differential gene expression analysis (Law et al., 2014). We integrated the gene expression changes associated with ccRCC using genome-scale metabolic modeling to pinpoint deregulation at the levels of metabolic pathways and connected components in the metabolic network. In the first case, we used piano (Våremo et al., 2013) to perform consensus gene-set analysis of KEGG metabolic pathways and to determine the pathways that were significantly deregulated in ccRCC and in which direction (i.e., up- or downregulated). In the second case, we first calculated the metabolites that were mostly affected by gene expression changes and then clustered these metabolites in the human metabolic network using Kiwi (Våremo et al., 2014) to emphasize whether specific pathway components were regulated. At the pathway level, we observed widespread downregulation of central carbon and amino acid metabolism, steroid biosynthesis, heparan sulfate biosynthesis, and other catabolic processes (Figure 1A). Biosynthesis of chondroitin sulfate was the only significantly upregulated pathway when both considering only upregulated genes (mixed directionality) and considering all genes weighted by their statistical significance (distinct directionality). Consistently, at the metabolite level, we observed downregulation of genes associated with metabolites related to branched-chain amino acids (e.g., 3-hydroxyisobutyrate) and steroids (e.g., 3-hydroxy-3-methylglutaryl-coenzyme A, also known as HMG-CoA) (Figure S1). Interestingly, whereas most metabolites interact through central carbon metabolites, this analysis returned an unconnected sub-network of metabolites that comprised precursors of chondroitin and heparan sulfate, with opposite directions of regulation (Figure 1B). The repression of central carbon and amino acid metabolism was in agreement with previous computational analyses (Creighton et al., 2013; Gatto et al., 2014), which were recently experimentally validated (Cuperlovic-Culf et al., 2016; Nilsson et al., 2015). However, both analyses revealed distinct and opposite regulation of chondroitin and heparan biosynthesis, which closely interacts within the glycosaminoglycan biosynthesis pathway. Given that this finding was not previously

reported, we sought to explore glycosaminoglycan biosynthesis in further detail.

Glycosaminoglycan Biosynthesis Displays Coordinated Regulation Specific to Clear Cell Renal Cell Carcinoma at the Transcript and Protein Levels

Chondroitin (CS) and heparan (HS) sulfates are glycosaminoglycans (GAGs) that share a common biosynthetic route in the linkage to the core protein, but thereafter, they differ in polymerization: CS, repeating disaccharides, is constituted by *N*-acetylgalactosamine and glucuronic acid residues, while HS, repeating disaccharides, is constituted by *N*-acetylglucosamine and glucuronic acid residues (Kreuger and Kjellén, 2012; Mikami and Kitagawa, 2013). In ccRCC, we observed coordinated regulation of GAG biosynthesis, defined by substantial upregulation of most genes specific to CS biosynthesis (11/13) and concurrent downregulation of genes specific to HS biosynthesis (8/13), indicating a potential change in GAG disaccharide composition, sulfation, and chain length in ccRCC (Figure 2A; Table S2). We confirmed this coordinated regulation of GAG biosynthesis in two independent datasets that compared gene expression in ccRCC versus non-tumor samples (Peña-Llopis et al., 2012; Wang et al., 2009), with strong and significant correlations between expression fold changes in these studies and the TCGA samples (Pearson's correlation coefficient $\rho = 0.87$ – 0.89 ; Figure 2B).

To verify the extent to which this regulatory pattern is ccRCC specific, we repeated an analogous analysis of six other epithelial cancer types for which at least 20 tumor-adjacent normal samples were found in TCGA (i.e., breast invasive carcinoma, colon adenocarcinoma, head and neck squamous cell carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, and uterine corpus endometrial carcinoma). None of these cancers displayed the same coordinated pattern as in ccRCC, which was a clear outlier according to unsupervised hierarchical clustering, although we found cancer-type-dependent regulation of individual enzymes involved in GAG biosynthesis (Figure 2C; Table S2). In addition, we never observed both the CS and the HS biosynthesis pathways among the top ranked regulated pathways in any of these cancer types (Figure S2).

To evaluate whether the coordinated regulation of GAG biosynthesis is also represented at the level of protein expression, we used immunohistochemistry on a ccRCC tissue microarray to detect the presence of three representative proteins characteristic of the pathway (CHPF2 in CS biosynthesis and HS6ST2 and EXTL1 in HS biosynthesis) in ccRCC versus normal kidney samples (Figure 2D). In accordance with gene expression changes, CHPF2 displayed strong staining in all of the tested tumor samples (positive in 21 of 21 samples) and only weak and likely unspecific staining in the kidney proximal tubule cells (0/2); HS6ST2 showed weak or no staining in all of the tested tumor samples (positive in zero of 32), while it was detected in both the podocytes in the kidney glomeruli and the endothelial cells of larger vessels (2/2); and EXTL1 was undetected in 96% of the tested tumor samples (positive in one of 27), but it was stained strongly in the kidney-collecting duct cells (two of two) (representative samples in Figure 2E). Taken together, these results suggested that coordinated regulation of GAG biosynthesis is a prominent metabolic event occurring exquisitely in the kidney during ccRCC transformation.

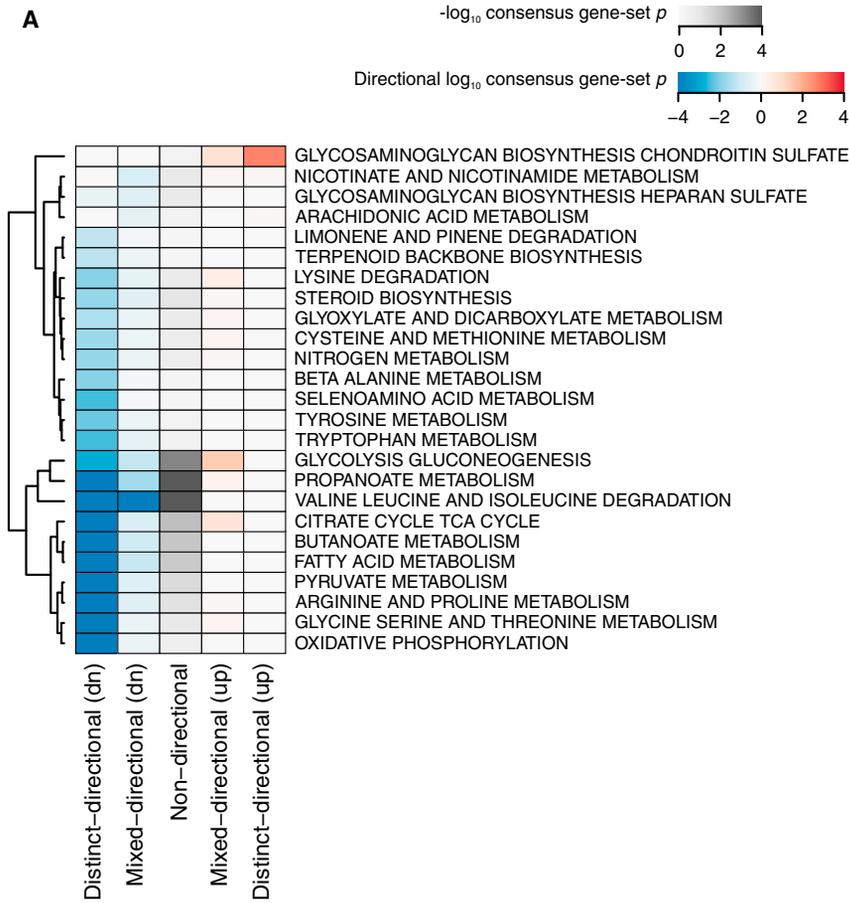
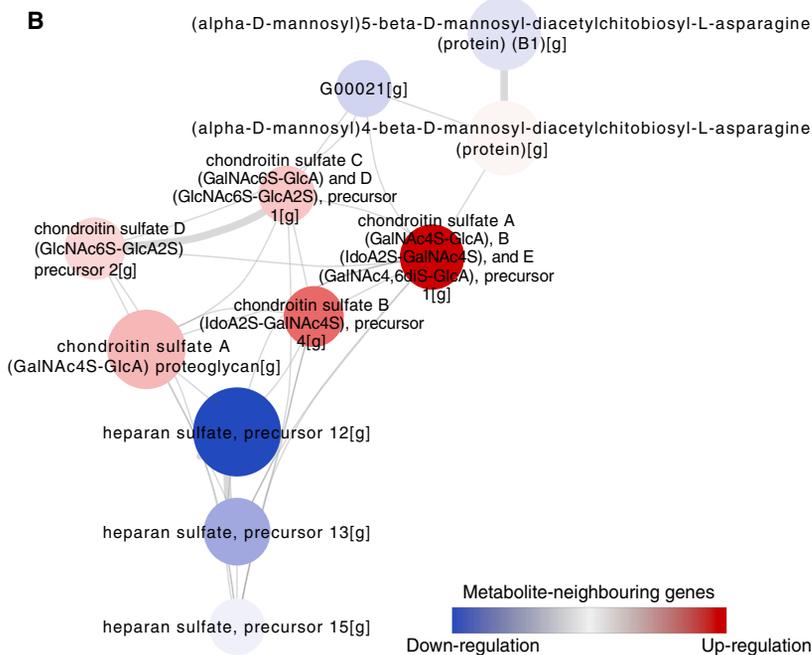


Figure 1. Integrated Metabolic Modeling and Differential Gene Expression Analysis in ccRCC versus Tumor-Adjacent Normal Kidney Tissue

(A) Consensus gene-set analyses for KEGG metabolic pathways in ccRCC versus tumor adjacent normal samples. Each entry shows the median gene-set *p* value for a pathway among six different gene-set analysis methods. Every column represents a different regulatory direction: Dist(dn), distinct directional down; Mix(dn), mixed directional down; Nondir, non-directional; Mix(up), mixed directional up; Dist(up), distinct directional up.

(B) An extract from an unbiased analysis of the metabolic network aimed at detecting connected metabolites, in which the neighboring genes were prevalently up- (red) or downregulated (blue) (see also Figure S1 for the complete network results).



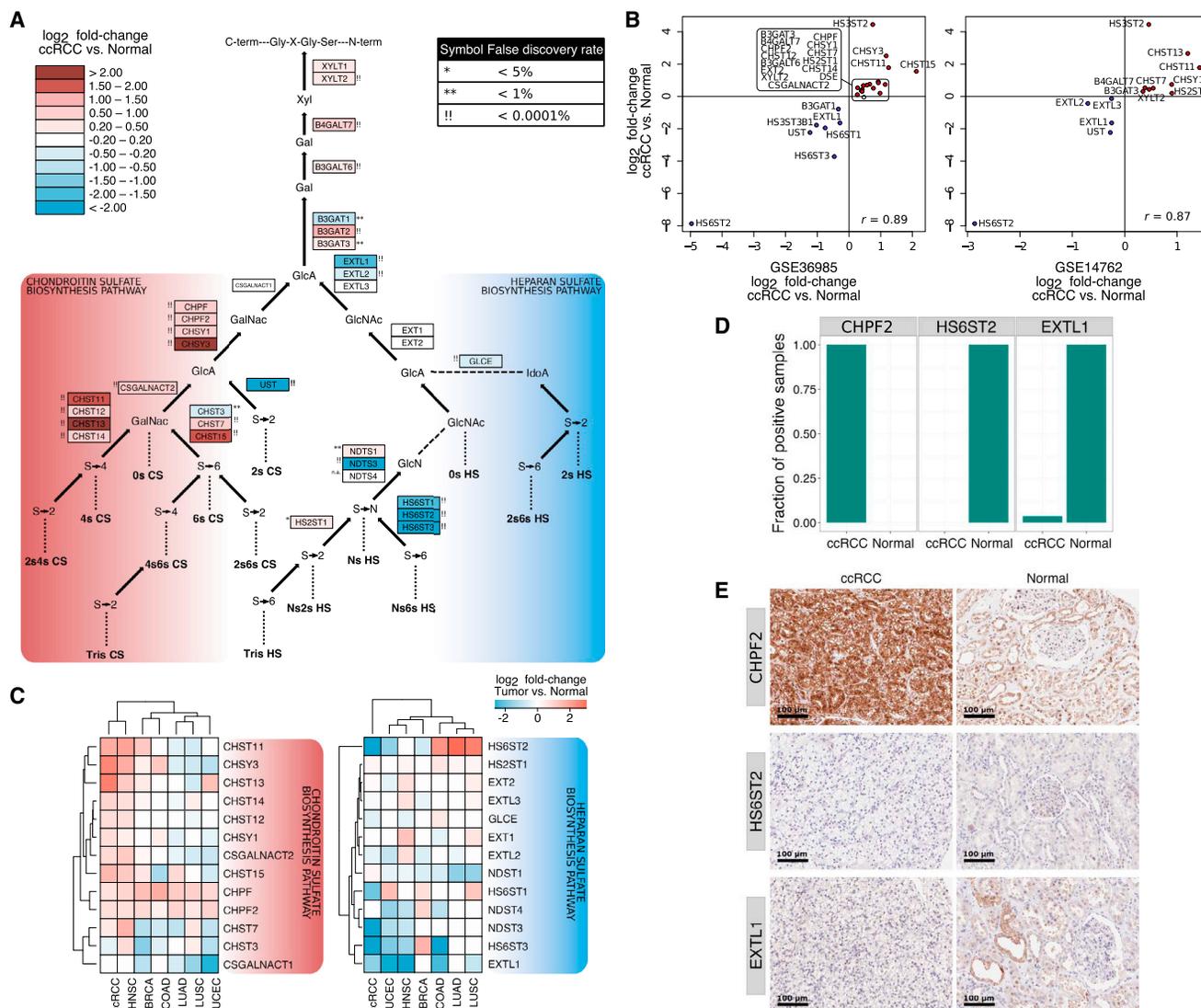


Figure 2. Coordinated Regulation of Glycosaminoglycan Biosynthesis in ccRCC versus Tumor-Adjacent Normal Kidney Tissue at the Transcript and Protein Levels

(A) Pathway view of glycosaminoglycan biosynthesis in ccRCC. Each box shows the enzyme(s) carrying out a given reaction in the pathway. The color represents the \log_{10} fold change in ccRCC versus non-tumor tissue for the enzyme-coding gene, while the symbol next to each box indicates the significance of the corresponding gene regulation (in terms of false discovery rate). The pathway is drawn according to KEGG gene associations (note that genes related to dermatan sulfate biosynthesis or sulfation at C3 in heparan sulfate are not shown, the latter event being rarely observed [Thacker et al., 2014]). Solid arrows indicate the addition of a molecule, dashed lines indicate the conversion of a molecule, and dotted lines indicate the final disaccharide composition up to that point.

(B) Correlation of gene expression \log_2 fold changes in the glycosaminoglycan biosynthesis pathway between TCGA samples (y axis) and two independent studies (GSE36986 and GSE14762 [Peña-Llopis et al., 2012; Wang et al., 2009]).

(C) Gene expression \log_2 fold changes in the glycosaminoglycan biosynthesis pathway in ccRCC, compared to other cancers versus matched non-tumor tissue. HNSC, head and neck squamous cell carcinoma; BRCA, breast invasive carcinoma; COAD, colon adenocarcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; UCEC, uterine corpus endometrial carcinoma. See also Figure S3 for gene expression analysis in other cancers at the pathway level.

(D) Fraction of the samples positive for immunohistochemical staining of CHPF2, HS6ST2, and EXTL1 in ccRCC (21–27 tissue samples) versus normal kidney (two samples). The results are presented as the consensus of staining performed in duplicate.

(E) Immunohistochemical staining for CHPF2, HS6ST2, and EXTL1 in representative ccRCC and normal samples.

Altered Regulation of Glycosaminoglycan Biosynthesis Is Exacerbated in Metastasis and Is Detectable in Patients' Urine and Plasma

CS and HS have been long implicated in the regulation of angiogenesis, adhesion, invasion, and migration, which are key steps

in the metastatic cascade (Afratis et al., 2012; Jackson et al., 1991). We extended our differential gene expression analysis to verify whether the genes involved in GAG biosynthesis showed further regulation in ccRCC patients with metastasis. We found that 11 genes involved in GAG biosynthesis were differentially

Table 1. Clinical Data for the Discovery and Validation Cohorts

	mccRCC (n = 34)	Healthy (n = 16)	mccRCC (n = 18)	Healthy (n = 9)
Cohort Characteristics				
Class	Discovery	Discovery	Validation	Validation
Both plasma and urine samples	13	16	7	9
Only plasma samples	21	0	11	0
Baseline Characteristics				
Age (range, years)	64.6 (59.2–70.5)	62.7 (57.6–65)	56.1 (50.7–64.6)	55.2 (42.2–64.3)
Female	23.1%	62.5%	14.3%	44.4%
Caucasian	100%	100%	100%	100%
BMI (kg/m ²)	26.1 (23.7–26.7)	25.7 (22.8–29.1)	23.1 (22.3–28.8)	23.5 (21.7–25.8)
Current Oncological Therapy				
None	64.7%	100%	61.1%	100%
Sunitinib	20.6%	0%	27.8%	0%
Other antineoplastic agents	14.7%	0%	11.1%	0%
Lifestyle Characteristics				
Physical exercise (hr/week)	2 (2–3.5)	3 (2–6)	3 (3–4.75)	4 (3–7)
Bread consumption (servings/week)	10 (7–14)	10 (6.25–14.5)	14 (12–14)	7 (4–7)
Pizza consumption (servings/week)	1 (0.5–1)	0 (0–0)	1 (0.25–1)	0 (0–1)
Pasta consumption (servings/week)	7 (5–7)	1 (0.5–1)	7 (5–7)	2 (0–3)
Rice consumption (servings/week)	3 (2–3)	1 (1–2)	2 (1.5–3)	1 (1–1)
Alcohol consumers	23.1%	87.5%	71.4%	100%
Fiber consumers	23.1%	50%	57.1%	66.7%
Smoking Habits				
Never smokers	61.5%	56.25%	71.4%	66.7%
Ex smokers	38.5%	37.5%	28.6%	33.3%
Smokers	0%	6.25%	0%	0%

All results are presented as medians (25th, 75th percentile) or percentages. Missing values were omitted. Detailed clinical data are reported in [Table S3](#).

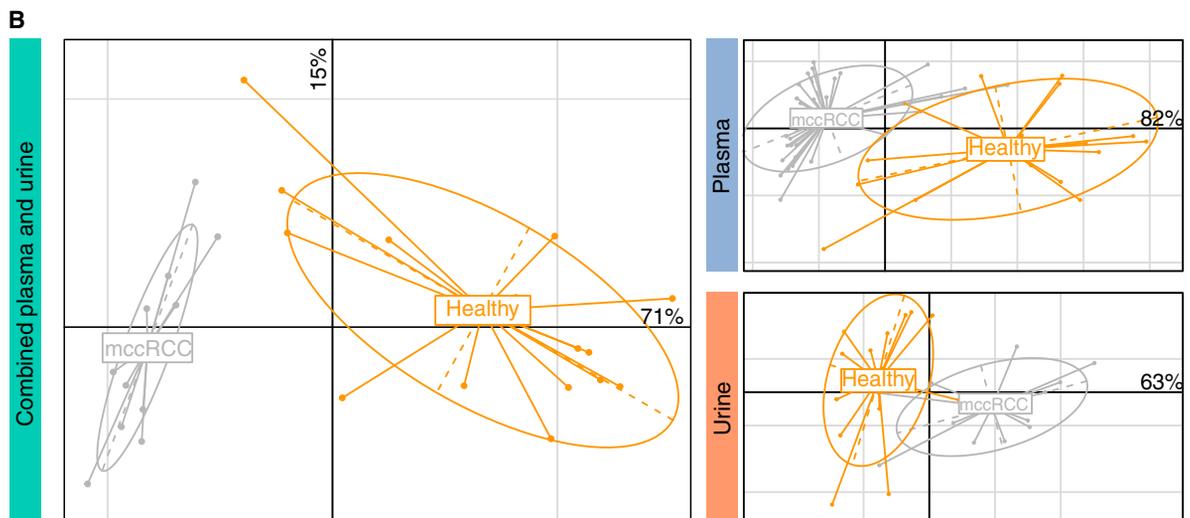
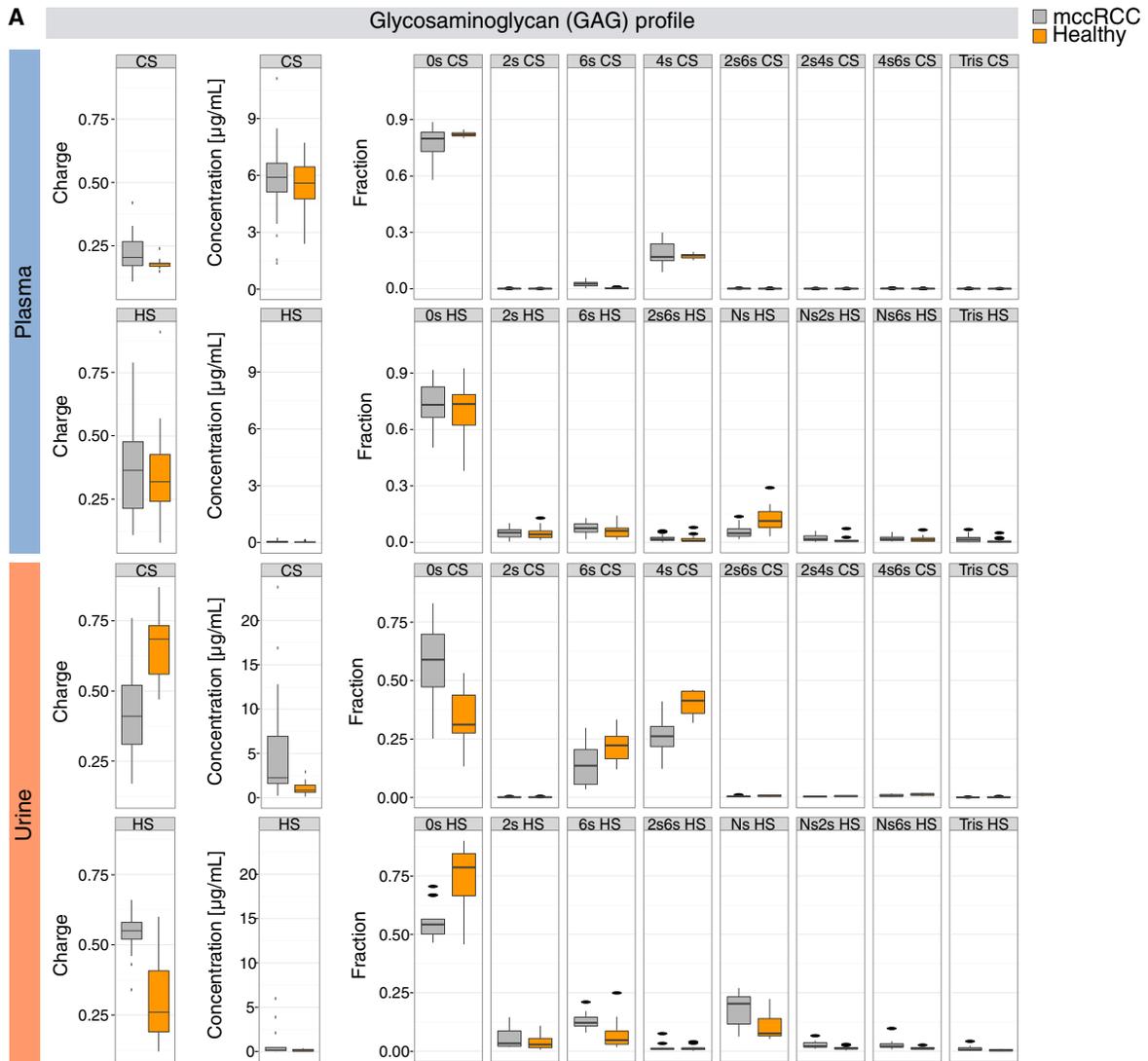
regulated in metastasis, exacerbating the overexpression of CS-associated genes and the suppression of HS-associated genes ([Figure S3](#)). This finding suggested that coordinated regulation of GAG biosynthesis is an event exacerbated by metastasis. While the assembly of GAG chains occurs intracellularly, the completed proteoglycan is secreted in the extracellular matrix ([Silbert and Sugumaran, 2002](#)). Hence, considered together, we speculated that not only might eventual changes in GAGs due to ccRCC progression be reflected in kidney-proximal fluids, but also these changes should be easier to detect in metastatic ccRCC (mccRCC) patients. This speculation leverages on variations in GAG concentration and composition having been observed in the proximal fluids of other diseases in which GAGs were implicated ([Anower-E-Khuda et al., 2013](#); [Mannello et al., 2014, 2015](#); [Schmidt et al., 2014](#); [Schmidt et al., 2016](#); [Volpi et al., 2015](#)).

To verify whether changes in the GAG profile occur in mccRCC and could be measured in accessible body fluids, we recruited a discovery cohort of 50 subjects, consisting of 34 patients with mccRCC and 16 healthy individuals ([Table 1](#); [Table S3](#)). Plasma and urine samples were obtained from all of the subjects, except for 21 mccRCC patients from whom only plasma samples were available. CS and HS concentrations and their disaccharide compositions were quantified in the samples using liquid chromatography with online electrospray ionization mass spectrometry (ESI-MS). In total, 18 independent GAG properties were

measured in every fluid sample (note that the GAG charge is the sum of all of the sulfated disaccharide fractions). The collection of all of these data points defined a GAG profile. We observed remarkable differences between the GAG profiles of mccRCC patients and those of healthy individuals, both in the plasma and urine samples ([Figure 3A](#)). Principal component analysis (PCA) of GAG profiles that combined plasma and urine measurements revealed that mccRCC patients were clearly separate from healthy individuals (71% of the variance was explained by the first component; [Figure 3B](#)). Similar separations were achieved using only plasma measurements (81% variance) or urine measurements (63% variance). These results indicated that mccRCC entails alterations in systemic GAG composition that are markedly distinct from those of healthy individuals.

Design of mccRCC Biomarkers Based on Plasma and Urine GAG Profiles

The changes in the plasma and urine GAG profiles, which were largely attributable to the occurrence of mccRCC, opened the opportunity to design accessible biomarkers based on the GAG properties that best distinguished the disease from a healthy state. We utilized Lasso penalized logistic regression ([Tibshirani, 1996](#)) with leave-one-out cross-validation to select robust GAG properties that are most predictive of clinical outcomes (i.e., mccRCC versus healthy). A biomarker score was subsequently



(legend on next page)

designed as a ratio, where the numerator is the sum of the properties associated with mcrRCC, and the denominator is the sum of the properties associated with the healthy state. Each term was normalized using the regression coefficients. We derived three potential disease biomarker scores, based on either plasma or urine or on combined measurements:

$$\text{Plasma score} = \frac{[6s \text{ CS}] + CS_{tot}}{\frac{3}{10} \frac{[4s \text{ CS}]}{[6s \text{ CS}]} + [Ns \text{ HS}]}$$

$$\text{Urine score} = \frac{[Ns6s \text{ HS}] + 60 \cdot \text{Charge HS}}{[4s \text{ CS}]}$$

$$\text{Combined score} = \text{mean}(\text{Plasma score}, \text{Urine score})$$

where terms in brackets represent the fraction of the disaccharide for the corresponding GAG (the abbreviations describe different sulfation patterns for CS and HS as per Figure 1B), CS_{tot} is the total concentration of CS (in $\mu\text{g/mL}$), and Charge HS is the total fraction of sulfated disaccharides of HS. We then calculated the three scores for each sample and observed that the mcrRCC samples had recurrently elevated scores, compared to healthy samples (Figure 4A). We computed significant non-null mean differences in all three scores between the two groups using robust Bayesian estimation. The mean difference was equal to 2.15 for the combined score (95% high-density interval [HDI] 1.72–2.60), 2.49 for the plasma score (95% HDI 1.94–3.05), and 0.79 for the urine score (95% HDI 0.52–1.06). The performance of the three biomarkers was evaluated using receiver operating characteristic (ROC) curves, and the area under the curve (AUC) was found to be 1 (perfect classifier) in the case of the combined and plasma scores and 0.966 for the urine score (Figure 4B; Table 2). A straightforward clinical implementation of these biomarkers would be to predict occurrence of the disease, for example, to monitor mcrRCC patients after surgery or to diagnose response to treatment using a simple non-invasive test, in addition to or as a substitute for standard radiological tests. Thus, from each ROC curve, we computed a score cutoff that maximized the positive predictive value (PPV) of the biomarker (Lopez-Raton et al., 2014) (Table 2). Taken together, these findings demonstrated that alterations in plasma and urine GAG composition occurring in mcrRCC could be summarized as scores. In turn, these scores accurately distinguished patients with from healthy individuals, emphasizing their potential as disease biomarkers.

Validation of the mcrRCC Biomarkers in an Independent Cohort and in Patients with No Evidence of Disease

To validate whether these scores had reproducible accuracy in an independent cohort, we recruited 27 subjects, consisting of

18 patients with mcrRCC and nine healthy individuals (Table 1; Table S3). Plasma and urine samples were obtained from all of the subjects, except for 11 mcrRCC patients from whom only plasma samples were available. We analyzed the three biomarkers for each individual and computed the corresponding scores. The scores were also remarkably higher in mcrRCC patients compared with healthy controls in this validation cohort (Figure 4C). We computed an AUC value equal to 1 for all three biomarkers (Figure 4D). Additionally, the specificity at the previously determined cutoff score was 100% for all of the biomarkers, consistent with their potential to predict positive diagnoses (Table 2). This evidence strongly suggested that the three biomarkers could indicate the occurrence of mcrRCC by means of a non-invasive analytical test.

We sought to verify whether the scores would normalize in subjects previously diagnosed with mcrRCC but with no evidence of disease, which would strongly suggest that the biomarkers follow mcrRCC and are suitable for monitoring its progression. In addition, we could not exclude that previous exposure to the disease might have prolonged effects on systemic GAG composition. Therefore, we analyzed GAGs in the plasma and urine of a prospective cohort of eight individuals diagnosed with mcrRCC but with no evidence of disease at the time of sampling. The GAG profiles in the urine and the plasma were remarkably distinct from those of patients with mcrRCC and shifted toward the profiles of healthy individuals (Figure 5A). When we computed the biomarker scores using the same formulas designed above, we observed a significant decrease compared to the expected value in mcrRCC (mean difference in the combined score, 0.94, 95% HDI 0.72–1.17; in the plasma score, 0.86, 95% HDI 0.69–1.03; in the urine score, 0.72, 95% HDI 0.40–1.07). The accuracy of the test based on the previously identified cutoffs was high for both the plasma and urine scores, with seven of eight cases less than the cutoff and hence 87.5% of the subjects were correctly identified as not having mcrRCC (Figure 5B). The accuracy was lower for the combined score, with six of eight subjects (75%) correctly classified according to the cutoff. Although only a longitudinal study could corroborate a positive correlation between the tumor burden and these scores, these results argued that plasma and urine GAG composition could be used as a robust and accurate diagnostic biomarker for the occurrence of mcrRCC.

Analysis of the Predictive Value of mcrRCC Biomarkers, Accounting for Confounding Factors

We sought to identify the extent to which the measured systemic GAG alterations were purely attributable to mcrRCC progression, as suggested by the underlying transcriptional regulation, or whether they were also dependent on other confounding factors. Therefore, we gathered clinical and dietary information that could confound the associations of the scores with clinical outcomes for 33 individuals (17 mcrRCC patients and 16 healthy

Figure 3. The Glycosaminoglycan Plasma and Urine Profiles of mcrRCC Patients Are Markedly Distinct from Those of Healthy Individuals (A) The glycosaminoglycan profiles of mcrRCC patients (gray boxplots) versus healthy individuals (orange boxplots) in plasma (top, 34 versus 16 samples) and urine (bottom, 13 versus 16 samples). Each profile consists of 18 independent measurements of GAGs (nine related to CS and nine related to HS), which refer to the total concentration and the disaccharide composition. (B) Principal component analysis of sample GAG profiles, using measurements from plasma, urine, or both.

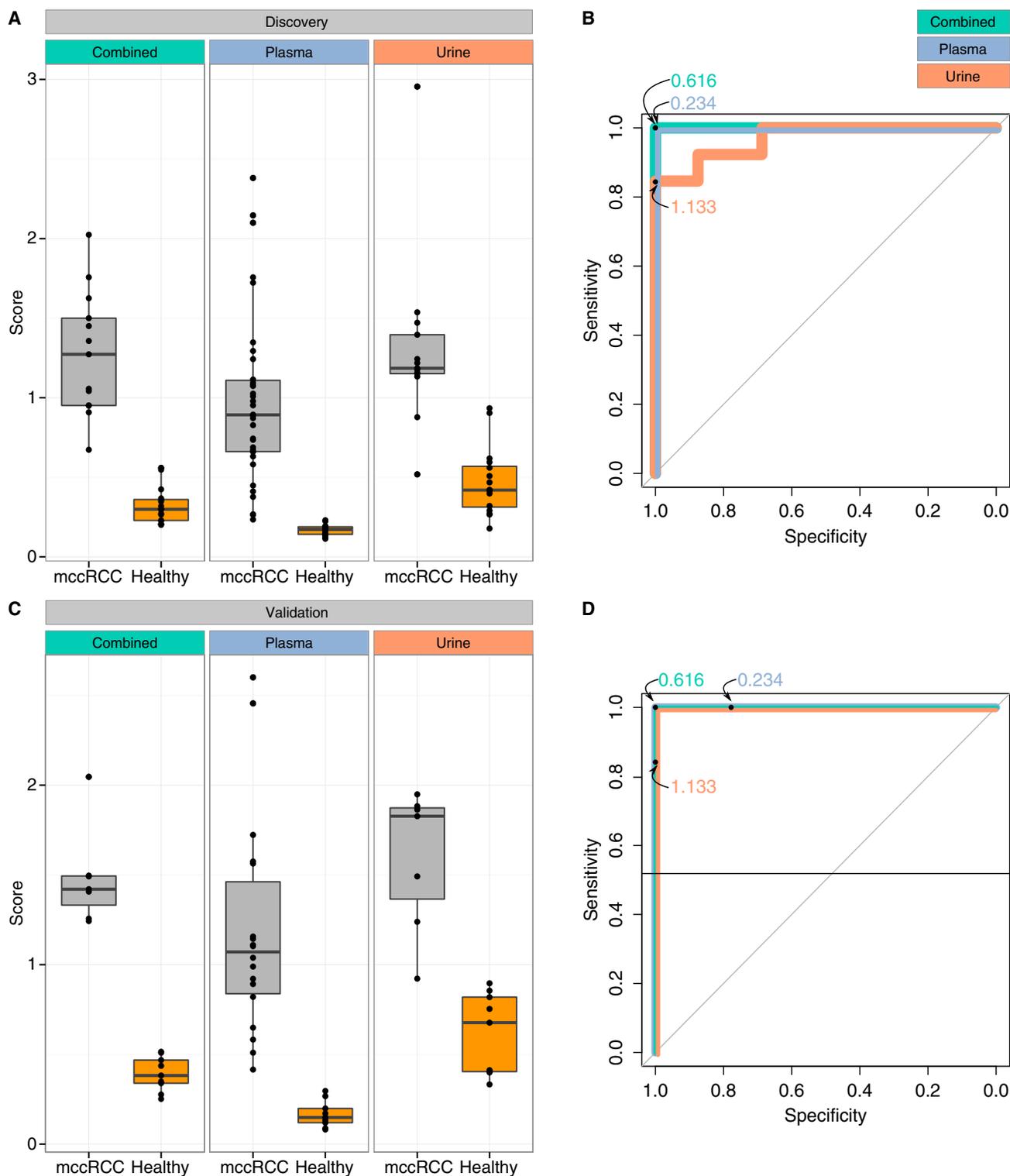


Figure 4. The Glycosaminoglycan Profile Can Be Summarized in Three Scores Based on Measurements in the Plasma, Urine, or Both that Can Accurately Predict the Occurrence of mCCRCC

(A) Plasma, urine, and combined scores in mCCRCC patients (gray boxplots) versus healthy individuals (orange boxplots) belonging to the discovery cohort (34 versus 16 plasma samples and 13 versus 16 urine samples).

(B) Receiver operating characteristic (ROC) curves in the classification of samples of the discovery cohort as either mCCRCC or healthy, based on the combined, plasma, and urine scores. For each biomarker, an optimal cutoff scores that maximized the positive predictive value is indicated.

(legend continued on next page)

Table 2. Measures of Accuracy for GAG Scores in the Prediction of mcrRCC for the Discovery and Validation Cohorts at the Optimal Cutoff Score

	AUC	Optimal Cutoff Score	Accuracy	Specificity	Sensitivity
Discovery Cohort					
Combined marker	1	0.616	100%	100%	100%
Plasma marker	1	0.234	100%	100%	100%
Urine marker	0.966	1.133	93.1%	100%	84.6%
Validation Cohort					
Combined marker	1	—	100%	100%	100%
Plasma marker	1	—	92.6%	100%	77.8%
Urine marker	1	—	93.7%	100%	85.7%

The optimal cutoff score was calculated in the discovery cohort and verified in the validation cohort.

patients; Table S3). As reported in Table 1, we observed an uneven distribution of some baseline characteristics, for example, gender, pasta consumption, and alcohol consumption. Therefore, we tested whether the clinical outcomes could be purely inferred by some of the confounding factors, rather than by the biomarker scores. First, we determined the most biased factors between the mcrRCC versus healthy groups. To this end, we regressed the clinical outcome based on the confounding factors and the combined score using Lasso penalized logistic regression. This analysis selected four potentially relevant confounding factors: age; weekly consumption of pasta and rice; and use of alcohol. Then, we performed analysis of covariance using logistic regression to test the strength of the association between clinical outcome and the combined score, using the four confounding factors as covariates. Notably, none of the covariates made a significant contribution to the regression of the clinical outcome ($p = 0.27$ and 0.44 ; Figure S4). In addition, we calculated that the logistic regression model based solely on the combined score was the most likely model ($p = 99.2\%$), according to the minimum Kullback-Leibler divergence criterion: the Akaike information criterion for the regression based on the combined score only was significantly lower than for the regression based also on the four covariates (7.8 versus 17.5, respectively). A similar conclusion was reached for the plasma score (6.0 versus 17.9) but not for the urine score (23.0 versus 17.5), for which pasta consumption displayed a significant effect in the regression of the clinical outcome ($p = 0.03$). Taken together, these results indicated that the combined and plasma scores alone (but not the urine score) had strong associations with the clinical outcome regardless of any here-considered confounding factors, prompting the use of GAG measurements as unbiased predictors of the occurrence of mcrRCC.

Finally, we explored whether systemic therapy had an effect on the biomarker scores, given that these scores were calculated by profiling body fluids. We limited our analysis to the patients for whom only plasma samples were collected (and hence, we checked solely the effect on the plasma scores) because, for

this group, we noted a comparable number of treated ($n = 19$) and untreated ($n = 33$) patients. We did not observe any significant correlation between the plasma score and the use of systemic therapy, based on a linear regression of the score on the treatment status of the sample ($p = 0.518$) and the type of treatment (sunitinib versus other regimens, $p = 0.508$). Overall, these analyses of covariance showed that GAG measurements, in the form of the proposed scores, could robustly predict the occurrence of mcrRCC despite baseline and treatment differences across patients. This robustness was likely due to coordinated regulation of GAG biosynthesis intrinsic to ccRCC progression, mirrored at the level of kidney-adjacent fluids.

DISCUSSION

This study revealed that coordinated regulation of GAG biosynthesis, which features concurrent upregulation of the branch leading to CS formation and downregulation of the branch leading to HS formation, is a prominent event in ccRCC. Additionally, many pathway-associated genes are further up- or downregulated in metastasis. This discovery can be attributed to an increased number of samples and recent advances in metabolic network analysis: indeed, traditional gene-set enrichment analysis likely misses the distinctive regulation of the two branches within the gene set because the opposite fold changes would cancel each other out. However, even considering the large sample size and the independent validation with tissue microarrays, we cannot completely exclude that other systematic confounding factors might partially explain the observed differential regulation, most importantly local inflammation in the kidney, which can occur concomitantly with other renal diseases unrelated to cancer. At the same time, altered expression of CS and HS, particularly in glycan composition and sulfation, has been indicated in the promotion of migration, metastasis, and angiogenesis in a number of tumor models, including skin (Smetsers et al., 2004), lung (Mizumoto et al., 2012), brain (Wade et al., 2013), and breast (Fernández-Vega et al.,

(C) Plasma, urine, and combined scores in mcrRCC patients (gray boxplots) versus healthy individuals (orange boxplots) belonging to the validation cohort (18 versus nine plasma samples and seven versus nine urine samples).

(D) ROC curves in the classification of samples of the validation cohort as either mcrRCC or healthy, based on the combined, plasma, and urine scores. See also Figure S5 for score correlations with confounding factors.

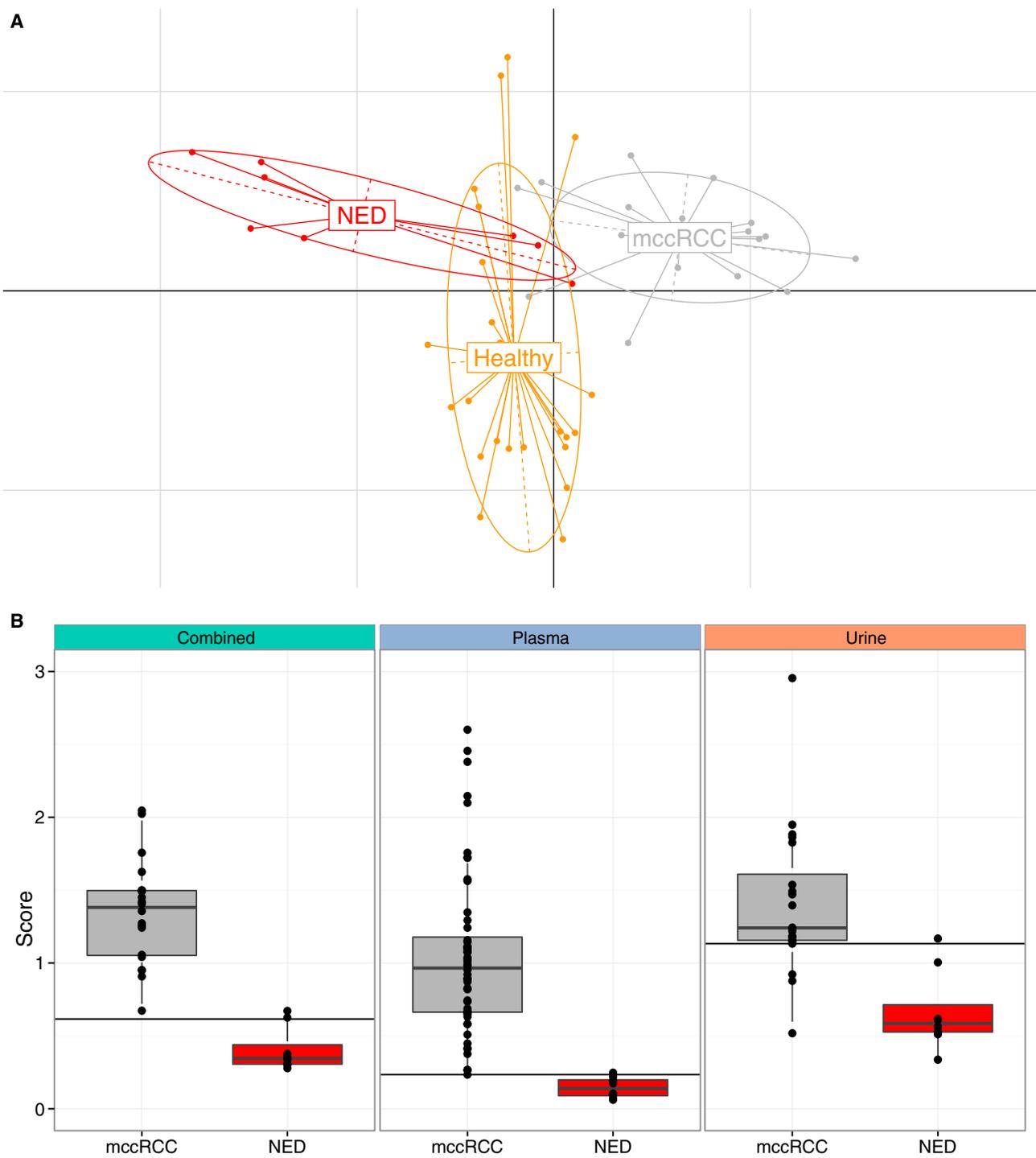


Figure 5. The Glycosaminoglycan Profile and Combined, Plasma, and Urine Scores in Subjects Previously Diagnosed with mccRCC but with No Evidence of Disease at the Time of Sampling

(A) Principal component analysis of GAG profiles using measurements from both plasma and urine in healthy subjects, mccRCC patients, and subjects with no evidence of disease (NED, eight samples).

(B) Biomarker scores in NED and all mccRCC subjects analyzed in this study. The horizontal lines represent the optimal cutoff scores at which a subject was classified as either with mccRCC or as healthy.

2013); however, contrary to these studies that focused on individual GAG types, our work reports an extensive, consistent, and coordinated regulation of the whole biological process of GAG biosynthesis in a cancer type. The relevance of such precise regulation of GAGs in ccRCC could be attributed to the roles of GAGs in the remodeling of the extracellular matrix, which strongly depends on their composition and abundance (Afratis et al., 2012). For example, a chondroitin sulfate-rich matrix was linked to the development of self-contained and defined lesions in lower grade glioma (compared to the microscopic infiltrations typical of glioblastomas) (Silver et al., 2013), which is a tumor growth model closely resembling ccRCC (Rini et al., 2009). However, it remains to be explored how this regulatory program is mechanistically linked to metastasis, rather than representing a coordinated metabolic event attributable to the remodeling of the kidney caused by the disease.

Because GAGs localize and act in the extracellular matrix, we assumed that changes in their regulation would reflect changes in their profiles in body fluids proximal to the kidney, e.g., blood and urine, as observed in other pathologies (Anower-E-Khuda et al., 2013; Mannello et al., 2014, 2015; Schmidt et al., 2014, 2016; Volpi et al., 2015). In particular, this behavior should be exacerbated in metastasis. Currently, there is no diagnostic biomarker that has entered routine practice for metastatic ccRCC (Jonasch et al., 2012; Moch et al., 2014). At the same time, metastatic disease is invariably incurable, although rare complete responses have been reported in association with oncological targeted therapies with or without metastasectomy (Albiges et al., 2012). Therefore, it would undoubtedly represent an important clinical advancement if changes in the GAG profile could constitute an indicator of the occurrence of the disease. The availability of such a test would be valuable for a number of medical decisions: to monitor ccRCC before and after surgery or systemic treatment; to exclude relapse of the disease also over longer periods of time, after which a patient is typically declared cured; to assess the occurrence of ccRCC in a population at risk, such as genetically predisposed individuals; to ascertain whether metastasis is due to ccRCC or other neoplasms; and to follow treatment response in mcrRCC. Our validation findings provided proof of concept that the here-designed biomarker scores might effectively aid in undertaking some of these clinical decisions because they are calculated based on non-invasive measurements that are predictive of the clinical outcome and are independent of the here-considered confounding factors; most importantly, they are accurate and robust predictors of the disease.

The plasma and urine GAG profiles loosely resembled the expected patterns from the underlying transcriptional regulation, i.e., increased concentration and sulfation of CS relative to HS in ccRCC. Notably, a recent characterization of the GAG profile in early-stage ccRCC tissues was strongly correlated with the here-uncovered regulatory program (Ucakturk et al., 2016). In addition, a previous study examined CS/HS concentrations in early-stage ccRCC tissue samples (Batista et al., 2012), and re-elaboration of these data to emphasize the CS/HS ratio delineated a consistent trend (Figure S5). Overall, these data and our study were suggestive of an active and early role for ccRCC in de novo GAG production, likely stemming from gene expression

regulation. At the same time, the GAG profiles revealed some novel biological insights attributable to the occurrence of this cancer type. The GAG composition in the plasma of healthy individuals is typically not affected by any tissue. Here, we observed systemic alteration of GAG composition concomitant with metastatic ccRCC. The enrichment of chondroitin-4-sulfate and chondroitin-6-sulfate and 6-O-sulfated HS in mcrRCC samples was strikingly reminiscent of the GAG composition of lymphocytes (Shao et al., 2013). It is therefore tempting to speculate that infiltration of the immune system in mcrRCC could underlie the observed transcriptional regulation in the tumor. In the urine, the GAG composition in healthy individuals has not been as well characterized. The alterations reported here in the GAG profiles of mcrRCC samples might reflect progressive damage to cells in the kidney glomeruli (McCarthy and Wassenhove-McCarthy, 2012; Miner et al., 2011). Collectively, this evidence seemed to emphasize the importance of alterations in GAGs in the progression of ccRCC. Intriguingly, the uniqueness of these GAG alterations could be exploited to deliver drugs specifically to ccRCC, as recently shown by a study in which cancer was targeted using a GAG-binding malaria protein (Salanti et al., 2015).

Thus far, among the major difficulties that have impaired biomarker discovery and its translation into clinical practice have been the detection of targets in accessible samples and the reproducibility of results (Sawyers, 2008). Here, we provided evidence for a plasmatic and/or urinary biomarker of metastatic ccRCC that was supported by an intensely and consistently regulated biological process in ccRCC samples. We envision that future longitudinal studies that monitor the trend between the tumor load and scores might establish these biomarkers for a diverse range of diagnostic tools in the clinical management of ccRCC.

EXPERIMENTAL PROCEDURES

Gene Expression Analysis

RNA sequencing (RNA-seq) gene expression profiles for 481 ccRCC primary tumor and 71 tumor-adjacent normal-like samples were retrieved at The Cancer Genome Atlas (TCGA) (Table S1). Differential expression analysis for ccRCC versus non-tumor was performed using voom (Law et al., 2014). 2,090 genes with no annotation (3%) or no more than ten counts in less than 10% of the samples (7%) were discarded. The effect of metastasis was accounted by adding the metastatic status of each sample as a covariate in the linear model used in voom. Two independent microarray-generated datasets were retrieved in GEO (GEO: GSE36895 [Peña-Llopis et al., 2012] and GEO: GSE14762 [Wang et al., 2009]), and the differential expression analysis for ccRCC versus non-tumor was performed using limma (Smyth, 2004). The significance for changes in gene expression using either RNA-seq or microarray data was tested using empirical Bayes estimation on a linear model for a given comparison (in the case of RNA-seq the count variance was moderated as proposed in voom [Law et al., 2014]). Consensus gene-set enrichment analysis (GSA) using piano (Väremo et al., 2013) was performed using as gene sets either KEGG metabolic pathways or metabolites (i.e., a gene set is the list of reaction-encoding genes that involve a given metabolite [Patil and Nielsen, 2005]), where the gene-set p value is defined as the median p value among the following GSA methods: Fisher's test, Stouffer's test, reporter test, tail-strength test, mean, and median. The significance of a gene set for each GSA method was tested using a permutation test by shuffling gene labels 10,000 times. The gene sets ranked among the top 30 by most GSA methods are shown in a heatmap that is hierarchically clustered. The differential gene expression analysis and multiple gene-set analysis (limited to KEGG metabolic pathways) was then repeated for six other cancer types (breast invasive

carcinoma, colon adenocarcinoma, head and neck squamous cell carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, and uterine corpus endometrial carcinoma) compared to matched tumor-adjacent normal samples (which were also retrieved at TCGA; [Table S1](#)). All analyzed cancer types were subsequently hierarchically clustered upon \log_2 fold change in the expression of genes belonging to the KEGG glycosaminoglycan biosynthesis pathways (excluding genes belonging to dermatan sulfate biosynthesis and sulfotransferases on the C3 of heparan sulfate) compared to matched non-tumor samples. Gene-set relatedness between gene sets was computed in terms of the underlying network using Kiwi ([Våremo et al., 2014](#)), where gene sets were considered related if the mutual shortest path length is lesser than 2 in the network (to increase interpretability, gene sets with more than ten genes were neglected). In the case of metabolites, the gene-set network was extracted from the genome-scale metabolic model HMR2 ([Agren et al., 2014](#)). The methods outlined above were implemented using the respective R-packages, except Kiwi that is a Python module.

Immunohistochemical Staining

A tissue microarray containing 32 ccRCC samples and two normal kidney samples in duplicates was prepared and used for immunohistochemistry. An experienced urological pathologist selected all cases. The ethical approval was granted by the ethical committee at Lund University (LU289-07). Tissue sections of 4 μm were deparaffinized and rehydrated according to standard protocols. Antigen retrieval was performed using pressure cooking of the samples for 20 min in 10 mmol/l citrate buffer (pH 6.0). Immunohistochemical staining was performed using a Dako Techmate 500 unit, according to the manufacturer's instructions (Dako). Antibodies and dilutions used were HPA020992 (CHP2 1:35), HPA034625 (HS6ST2 1:125), and HPA037749 (EXTL1 1:35), all from Atlas Antibodies AB. Only tumor samples where both duplicates could be scored were included in the analysis (21 for CHP2, 32 for HS6ST2, and 27 for EXTL1).

Sample Collection

In the discovery cohort, plasma and urine samples were obtained from 34 patients with metastatic clear cell renal carcinoma in two sites, IOV-IRCCS, Padova, Italy and Sahlgrenska University Hospital, Göteborg, Sweden. For 21 patients, only plasma samples were obtained. A control group was formed using 16 healthy individuals without any renal or liver malignancy, nor inflammatory pathologies. In the validation cohort, plasma and urine samples were obtained from 18 patients with metastatic clear cell renal carcinoma in two sites, IOV-IRCCS, Padova, Italy and Sahlgrenska University Hospital, Göteborg, Sweden. For 11 patients, only plasma samples were obtained. A control group was formed using nine healthy individuals without any renal or liver malignancy, nor inflammatory pathologies. Samples from IOV-IRCCS, Padova, Italy were collected from a consecutive series of patients scheduled for mCCRCC follow-up, prospectively. Samples from Sahlgrenska University Hospital, Göteborg, Sweden were retrieved from the bio-bank in the Department of Urology and Oncology, retrospectively. All subjects provided written informed consent. The present observational study was notified to the Ethics Committee at IOV-IRCCS, Padova, Italy in January 2013. The approval to collect and analyze blood samples at the Sahlgrenska University Hospital, Göteborg, Sweden was obtained from the Regional Ethics Board of Västra Götaland, Sweden. Whole-blood samples were collected in EDTA-coated tubes. The tubes were centrifuged (2,500 $\times g$ for 15 min at 4°C), and the plasma was extracted and collected in a separate tube. Urine were collected in polypropylene tubes. The samples were stored at -80°C until they were shipped for analysis in dry ice. Clinical and dietary information is available in [Table S3](#).

Glycosaminoglycan Analysis

Sample preparation including extraction and purification steps were performed as previously described by [Coppa et al. \(2011\)](#) and [Volpi and Maccari \(2005a, 2005b\)](#), while sample GAGs separation and quantification were performed as described in [Volpi et al. \(2014\)](#) and [Volpi and Linhardt \(2010\)](#).

Briefly, to extract the GAGs, 500 μl of sample was lyophilized, reconstituted with 1 ml of a 20-mM TRIS-Cl buffer (pH 7.4), and treated with protease (proteinase K from Tritirachium album [Enzyme Commission 3.4.21.64], >500 U/ml from Sigma-Aldrich) at 60°C for 12 hr. After boiling for 10 min, centrifugation

and filtration was done on 0.45- μm filters, and the filtrate was lyophilized. The powder was dissolved in 1 ml of distilled water by prolonged mixing. After centrifugation at 5,000 $\times g$ for 15 min, 0.2 ml of 20% trichloro-acetic acid was added to the supernatant and stored for 2 hr at 4°C. The mixture was centrifuged at 5,000 $\times g$ for 15 min, and the supernatant was recovered and lyophilized for further purification on anion-exchange resin (QAE Sephadex A-25). After reconstitution with 500 μl 10 mM NaCl and centrifugation at 10,000 $\times g$ for 5 min, the supernatant was applied to a column (0.5 \times 2 cm) packed with about 0.4 ml of resin previously equilibrated with 10 mM NaCl. After washing the resin with 2 ml of 10 mM NaCl, 1 ml of 2.5 M NaCl was added. Fifty milliliters of ethanol was added to the eluate (1 ml) and stored at -20°C for 24 hr. After centrifugation at 5,000 $\times g$ for 15 min, the pellet was reconstituted in 160 μl of water and divided in two aliquots of 80 μl , and both were lyophilized. One aliquot of the extracted GAGs was treated with chondroitinase ABC, and the second one was submitted to heparinase treatment. Unsaturated disaccharides generated by the treatment of extracted GAGs with enzymes were fluorotagged with 2-aminoacridone ([Coppa et al., 2011](#); [Volpi and Maccari, 2005a, 2005b](#)) and separated by capillary electrophoresis equipped with laser induced fluorescence according to the previous reported method ([Volpi et al., 2014](#)).

Eighteen independent GAG properties were measured in each sample (either plasmatic or urinary): CS concentration, HS concentration, and fractions of disaccharide composition for both CS and HS. The charge is the sum over all sulfated disaccharide fractions. Principal component analysis was performed on available GAG properties for three cases: only plasmatic, only urinary, or both plasmatic and urinary (combined). Principal component analysis was implemented using R-package *ade4* ([Dray and Dufour, 2007](#)) (centering was performed by the mean). All measurements are available in [Table S3](#).

Biomarker Design

To design the biomarkers in the only plasmatic or in the only urinary case, we used Lasso penalized logistic regression ([Tibshirani, 1996](#)) with leave-one-out cross-validation to select those GAG properties that are most predictive of the clinical outcome (i.e., mCCRCC versus healthy) at the optimal Lasso penalty value. This was calculated using the *glmnet* R-package ([Friedman et al., 2010](#)) as the penalty value for which the cross-validation error was within 1 SE of the minimum. The biomarkers were built as the ratio between the sum of the GAG properties robustly predictive of mCCRCC over the sum of the GAG properties robustly predictive of healthy state. Each property value was normalized using the respective regression coefficient (rounded to the nearest rational number). The biomarker for the combined case was taken as the mean of the so-designed plasmatic and urinary biomarkers. The highest density interval (HDI) for the mean difference in biomarker scores between mCCRCC versus healthy was calculated using Bayesian estimation under the following assumptions: scores are sampled from a *t*-distribution of unknown and to be estimated normality (i.e., degrees of freedom); high uncertainty on the prior distributions; the marginal distribution is well approximated by a Markov chain Monte Carlo sampling with no thinning and chain length equal to 100,000. The estimation was performed using *BEST* ([Kruschke, 2013](#)) (the above assumptions are reflected by the default parameters). Bayesian estimation was preferred over the widely used *t* test since it provides a robust and reliable estimation of mean difference even under uncertainty of the underlying score distribution for the two groups (that is the case when the number of samples is limited) ([Nuzzo, 2014](#)).

Accuracy Metrics

For each biomarker (plasma, urine, or combined), we evaluated its performance in the binary classification of a sample as either mCCRCC or healthy at varying threshold scores by deriving the receiver operating characteristic (ROC) curves. We measured the accuracy of each biomarker as the area under the curve (AUC) of its ROC curve (AUC is 1 for a perfect classifier and 0.5 for a random classifier). We selected as a potential cutoff value for a given biomarker the score for which the positive predictive value was maximum; i.e., a sample whose biomarker score is above this cutoff value has the maximum probability of being mCCRCC. The ROC curves were calculated using the *pROC* R-package ([Robin et al., 2011](#)), while the optimal cutoff

was calculated using the OptimalCutpoints R-package (Lopez-Raton et al., 2014).

Analysis of Covariance

The analysis of covariance was performed using logistic regression on the clinical outcome (mccRCC versus control) on selected covariates among those reported in the clinical and dietary information in Table S3. These covariates were selected using Lasso penalized logistic regression with leave-one-out cross-validation as the most predictive of the clinical outcome at the optimal Lasso penalty value (chosen as described in Biomarker design). These covariates are age, weekly consumption of pasta and rice, and use of alcohol. Next, we performed logistic regression on the clinical outcome based on the combined score and the four selected covariates. The significance of each coefficient was tested using the Wald z-statistics for the hypothesis that the corresponding parameter is zero. The same procedure was followed to check the effect of systemic therapy as covariate, but using only plasma samples to regress the clinical outcome (since only for such sub-cohort there were enough patients that did not undergo any systemic therapy). In this case, either only one covariate was used to indicate the presence or absence of undergoing therapy or a second covariate to account for the specific effect of sunitinib was added. Logistic regression was implemented adopting the Firth bias-reduction method using the brglm R-package. The performance of the two alternative models for logistic regression (either combined score + age + weekly consumption of pasta + weekly consumption of rice + use of alcohol; or combined score) was evaluated according to the minimum Kullback-Leibler divergence criterion by calculating the Akaike's information criterion (AIC) for the models and deriving the model probability in terms of AIC weights (Wagenmakers and Farrell, 2004).

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2016.04.056>.

AUTHOR CONTRIBUTIONS

Conceptualization, F.G. and J.N.; Software, F.G.; Validation, F.G., N.V., H.N., and M.E.J.; Formal Analysis, F.G.; Investigation, F.G., H.N., and M.E.J.; Resources, N.V., U.B., A.R., M.M., U.S., and S.L.; Data Curation, F.G.; Writing – Original Draft, F.G.; Writing – Review & Editing, F.G., J.N., I.N., N.V., M.E.J., H.N., U.B., U.S., and S.L.; Visualization, F.G.; Supervision, J.N. and I.N.; Project Administration, F.G. and J.N.; Funding Acquisition, J.N.

ACKNOWLEDGMENTS

We wish to thank Maria A. Pinhal (Federal University of São Paulo) for valuable discussions and Sakda Khoomrung (Chalmers University of Technology) for sample preparation. We thank the research nurses Cristina Magro and Orejeta Diamanti (Istituto Oncologico Veneto IOV – IRCCS) and Kristina Welinder (Sahlgrenska University Hospital) for performing blood draws and storing samples. We also thank Dina Petranovic and Adil Mardinoglu (Chalmers University of Technology) for critically reviewing the article. This work was financially supported by the Knut and Alice Wallenberg Foundation.

Received: November 19, 2015

Revised: March 11, 2016

Accepted: April 14, 2016

Published: May 12, 2016

REFERENCES

Afratis, N., Gialeli, C., Nikitovic, D., Tsegenidis, T., Karousou, E., Theocharis, A.D., Pavão, M.S., Tzanakakis, G.N., and Karamanos, N.K. (2012). Glycosaminoglycans: key players in cancer cell biology and treatment. *FEBS J.* *279*, 1177–1197.

Agren, R., Mardinoglu, A., Asplund, A., Kampf, C., Uhlen, M., and Nielsen, J. (2014). Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* *10*, 721.

Albiges, L., Oudard, S., Negrier, S., Caty, A., Gravis, G., Joly, F., Duclos, B., Geoffrois, L., Rolland, F., Guillot, A., et al. (2012). Complete remission with tyrosine kinase inhibitors in renal cell carcinoma. *J. Clin. Oncol.* *30*, 482–487.

Anower-E-Khuda, M.F., Matsumoto, K., Habuchi, H., Morita, H., Yokochi, T., Shimizu, K., and Kimata, K. (2013). Glycosaminoglycans in the blood of hereditary multiple exostoses patients: half reduction of heparan sulfate to chondroitin sulfate ratio and the possible diagnostic application. *Glycobiology* *23*, 865–876.

Batista, L.T., Matos, L.L., Machado, L.R., Suarez, E.R., Theodoro, T.R., Martins, J.R., Nader, H.B., Pompeo, A.C., and Pinhal, M.A. (2012). Heparanase expression and glycosaminoglycans profile in renal cell carcinoma. *Int. J. Urol.* *19*, 1036–1040.

Bordbar, A., Monk, J.M., King, Z.A., and Palsson, B.O. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* *15*, 107–120.

Coppa, G.V., Gabrielli, O., Buzzega, D., Zampini, L., Galeazzi, T., Maccari, F., Bertino, E., and Volpi, N. (2011). Composition and structure elucidation of human milk glycosaminoglycans. *Glycobiology* *21*, 295–303.

Creighton, C.J., Morgan, M., Gunaratne, P.H., Wheeler, D.A., Gibbs, R.A., Gordon Robertson, A., Chu, A., Beroukham, R., Cibulskis, K., Signoretti, S., et al. (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* *499*, 43–49.

Cuperlovic-Culf, M., Cormier, K., Touaibia, M., Reyjal, J., Robichaud, S., Belbraouet, M., and Turcotte, S. (2016). H NMR Metabolomics analysis of renal cell carcinoma cells: effect of VHL inactivation on metabolism. *Int. J. Cancer* *138*, 2439–2449.

Dray, S., and Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* *22*, 1–20.

Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., and Parkin, D.M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* *127*, 2893–2917.

Fernández-Vega, I., García, O., Crespo, A., Castañón, S., Menéndez, P., Astudillo, A., and Quirós, L.M. (2013). Specific genes involved in synthesis and editing of heparan sulfate proteoglycans show altered expression patterns in breast cancer. *BMC Cancer* *13*, 24.

Finley, D.S., Pantuck, A.J., and Beldegrun, A.S. (2011). Tumor biology and prognostic factors in renal cell carcinoma. *Oncologist* *16* (Suppl 2), 4–13.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* *33*, 1–22.

Gatto, F., Nookaew, I., and Nielsen, J. (2014). Chromosome 3p loss of heterozygosity is associated with a unique metabolic network in clear cell renal carcinoma. *Proc. Natl. Acad. Sci. USA* *111*, E866–E875.

Gatto, F., Miess, H., Schulze, A., and Nielsen, J. (2015). Flux balance analysis predicts essential genes in clear cell renal cell carcinoma metabolism. *Sci. Rep.* *5*, 10738.

Gupta, K., Miller, J.D., Li, J.Z., Russell, M.W., and Charbonneau, C. (2008). Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review. *Cancer Treat. Rev.* *34*, 193–205.

Hakimi, A.A., Reznik, E., Lee, C.H., Creighton, C.J., Brannon, A.R., Luna, A., Aksoy, B.A., Liu, E.M., Shen, R., Lee, W., et al. (2016). An integrated metabolic atlas of clear cell renal cell carcinoma. *Cancer Cell* *29*, 104–116.

Jackson, R.L., Busch, S.J., and Cardin, A.D. (1991). Glycosaminoglycans: molecular properties, protein interactions, and role in physiological processes. *Physiol. Rev.* *71*, 481–539.

Jerby, L., and Ruppin, E. (2012). Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clin. Cancer Res.* *18*, 5572–5584.

Jonasch, E., Futreal, P.A., Davis, I.J., Bailey, S.T., Kim, W.Y., Brugarolas, J., Giaccia, A.J., Kurban, G., Pause, A., Frydman, J., et al. (2012). State of the science: an update on renal cell carcinoma. *Mol. Cancer Res.* *10*, 859–880.

- Kreuger, J., and Kjell n, L. (2012). Heparan sulfate biosynthesis: regulation and variability. *J. Histochem. Cytochem.* **60**, 898–907.
- Kruschke, J.K. (2013). Bayesian estimation supersedes the t test. *J. Exp. Psychol. Gen.* **142**, 573–603.
- Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29.
- Li, B., Qiu, B., Lee, D.S., Walton, Z.E., Ochocki, J.D., Mathew, L.K., Mancuso, A., Gade, T.P., Keith, B., Nissim, I., et al. (2014). Fructose-1,6-bisphosphatase opposes renal carcinoma progression. *Nature* **513**, 251–255.
- Lopez-Raton, M., Cadarso-Suarez, C., Rodriguez-Alvarez, M.X., and Gude-Sampedro, F. (2014). OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J. Stat. Softw.* **61**, 1–36.
- Mannello, F., Maccari, F., Ligi, D., Canale, M., Galeotti, F., and Volpi, N. (2014). Characterization of oversulfated chondroitin sulfate rich in 4,6-O-disulfated disaccharides in breast cyst fluids collected from human breast gross cysts. *Cell Biochem. Funct.* **32**, 344–350.
- Mannello, F., Maccari, F., Ligi, D., Santi, M., Gatto, F., Linhardt, R.J., Galeotti, F., and Volpi, N. (2015). Breast cyst fluid heparan sulphate is distinctively N-sulphated depending on apocrine or flattened type. *Cell Biochem. Funct.* **33**, 128–133.
- Mardinoglu, A., and Nielsen, J. (2015). New paradigms for metabolic modeling of human cells. *Curr. Opin. Biotechnol.* **34**, 91–97.
- Mardinoglu, A., Gatto, F., and Nielsen, J. (2013). Genome-scale modeling of human metabolism - a systems biology approach. *Biotechnol. J.* **8**, 985–996.
- McCarthy, K.J., and Wassenhove-McCarthy, D.J. (2012). The glomerular basement membrane as a model system to study the bioactivity of heparan sulfate glycosaminoglycans. *Microsc. Microanal.* **18**, 3–21.
- Mikami, T., and Kitagawa, H. (2013). Biosynthesis and function of chondroitin sulfate. *Biochim. Biophys. Acta* **1830**, 4719–4733.
- Miner, J.H., Towler, D., Chen, F., Kopan, R., and Hruska, K. (2011). Organogenesis of the kidney glomerulus: focus on the glomerular basement membrane. *Organogenesis* **7**, 75–82.
- Mizumoto, S., Takahashi, J., and Sugahara, K. (2012). Receptor for advanced glycation end products (RAGE) functions as receptor for specific sulfated glycosaminoglycans, and anti-RAGE antibody or sulfated glycosaminoglycans delivered in vivo inhibit pulmonary metastasis of tumor cells. *J. Biol. Chem.* **287**, 18985–18994.
- Moch, H., Srigley, J., Delahunt, B., Montironi, R., Egevad, L., and Tan, P.H. (2014). Biomarkers in renal cancer. *Virchows Arch.* **464**, 359–365.
- Nilsson, H., Lindgren, D., Mandahl Forsberg, A., Mulder, H., Axelson, H., and Johansson, M.E. (2015). Primary clear cell renal carcinoma cells display minimal mitochondrial respiratory capacity resulting in pronounced sensitivity to glycolytic inhibition by 3-Bromopyruvate. *Cell Death Dis.* **6**, e1585.
- Nuzzo, R. (2014). Scientific method: statistical errors. *Nature* **506**, 150–152.
- Patil, K.R., and Nielsen, J. (2005). Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl. Acad. Sci. USA* **102**, 2685–2689.
- Pe a-Llopis, S., Vega-Rub n-de-Celis, S., Liao, A., Leng, N., Pav a-Jim nez, A., Wang, S., Yamasaki, T., Zhrebker, L., Sivanand, S., Spence, P., et al. (2012). BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.* **44**, 751–759.
- Rini, B.I., Campbell, S.C., and Escudier, B. (2009). Renal cell carcinoma. *Lancet* **373**, 1119–1132.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., and M ller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77.
- Salanti, A., Clausen, T.M., Agerb k, M. ., Al Nakouzi, N., Dahlb k, M., Oo, H.Z., Lee, S., Gustavsson, T., Rich, J.R., Hedberg, B.J., et al. (2015). Targeting human cancer by a glycosaminoglycan binding malaria protein. *Cancer Cell* **28**, 500–514.
- Sawyers, C.L. (2008). The cancer biomarker problem. *Nature* **452**, 548–552.
- Schmidt, E.P., Li, G., Li, L., Fu, L., Yang, Y., Overdier, K.H., Douglas, I.S., and Linhardt, R.J. (2014). The circulating glycosaminoglycan signature of respiratory failure in critically ill adults. *J. Biol. Chem.* **289**, 8194–8202.
- Schmidt, E.P., Overdier, K.H., Sun, X., Lin, L., Liu, X., Yang, Y., Ammons, L.A., Hiller, T.D., Suffita, M.A., Yu, Y., et al. (2016). Urinary glycosaminoglycans predict outcomes in septic shock and ARDS. *Am. J. Respir. Crit. Care Med.*, Published online February 29, 2016.
- Schulze, A., and Harris, A.L. (2012). How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature* **491**, 364–373.
- Shao, C., Shi, X., White, M., Huang, Y., Hartshorn, K., and Zaia, J. (2013). Comparative glycomics of leukocyte glycosaminoglycans. *FEBS J.* **280**, 2447–2461.
- Silbert, J.E., and Sugumaran, G. (2002). Biosynthesis of chondroitin/dermatan sulfate. *IUBMB Life* **54**, 177–186.
- Silver, D.J., Siebzehnrubl, F.A., Schildts, M.J., Yachnis, A.T., Smith, G.M., Smith, A.A., Scheffler, B., Reynolds, B.A., Silver, J., and Steindler, D.A. (2013). Chondroitin sulfate proteoglycans potently inhibit invasion and serve as a central organizer of the brain tumor microenvironment. *J. Neurosci.* **33**, 15603–15617.
- Smetsers, T.F., van de Westerlo, E.M., ten Dam, G.B., Overes, I.M., Schalkwijk, J., van Muijen, G.N., and van Kuppevelt, T.H. (2004). Human single-chain antibodies reactive with native chondroitin sulfate detect chondroitin sulfate alterations in melanoma and psoriasis. *J. Invest. Dermatol.* **122**, 707–716.
- Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, 1–25.
- Thacker, B.E., Xu, D., Lawrence, R., and Esko, J.D. (2014). Heparan sulfate 3-O-sulfation: a rare modification in search of a function. *Matrix Biol.* **35**, 60–72.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B* **58**, 267–288.
- Ucakturk, E., Akman, O., Sun, X., Baydar, D.E., Dolgun, A., Zhang, F., and Linhardt, R.J. (2016). Changes in composition and sulfation patterns of glycoaminoglycans in renal cell carcinoma. *Glycoconj. J.* **33**, 103–112.
- Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033.
- V remo, L., Nielsen, J., and Nookaew, I. (2013). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res.* **41**, 4378–4391.
- V remo, L., Gatto, F., and Nielsen, J. (2014). Kiwi: a tool for integration and visualization of network topology and gene-set analysis. *BMC Bioinformatics* **15**, 408.
- Volpi, N., and Linhardt, R.J. (2010). High-performance liquid chromatography-mass spectrometry for mapping and sequencing glycosaminoglycan-derived oligosaccharides. *Nat. Protoc.* **5**, 993–1004.
- Volpi, N., and Maccari, F. (2005a). Composition of urinary glycosaminoglycans in a patient with pseudoxanthoma elasticum and familial Mediterranean fever. *Clin. Chim. Acta* **359**, 207–209.
- Volpi, N., and Maccari, F. (2005b). Microdetermination of chondroitin sulfate in normal human plasma by fluorophore-assisted carbohydrate electrophoresis (FACE). *Clin. Chim. Acta* **356**, 125–133.
- Volpi, N., Galeotti, F., Yang, B., and Linhardt, R.J. (2014). Analysis of glycosaminoglycan-derived, precolumn, 2-aminoacridone-labeled disaccharides with LC-fluorescence and LC-MS detection. *Nat. Protoc.* **9**, 541–558.
- Volpi, N., Coppa, G.V., Zampini, L., Maccari, F., Galeotti, F., Garavelli, L., Galeazzi, T., Padella, L., Santoro, L., and Gabrielli, O. (2015). Plasmatic and urinary glycosaminoglycan profile in a patient affected by multiple sulfatase deficiency. *Clin. Chem. Lab. Med.* **53**, e157–e160.
- Wade, A., Robinson, A.E., Engler, J.R., Petritsch, C., James, C.D., and Phillips, J.J. (2013). Proteoglycans and their roles in brain cancer. *FEBS J.* **280**, 2399–2417.

- Wagenmakers, E.-J., and Farrell, S. (2004). AIC model selection using Akaike weights. *Psychon. Bull. Rev.* *11*, 192–196.
- Wahlgren, T., Harmenberg, U., Sandström, P., Lundstam, S., Kowalski, J., Jakobsson, M., Sandin, R., and Ljungberg, B. (2013). Treatment and overall survival in renal cell carcinoma: a Swedish population-based study (2000–2008). *Br. J. Cancer* *108*, 1541–1549.
- Wang, Y., Roche, O., Yan, M.S., Finak, G., Evans, A.J., Metcalf, J.L., Hast, B.E., Hanna, S.C., Wundergem, B., Furge, K.A., et al. (2009). Regulation of endocytosis via the oxygen-sensing pathway. *Nat. Med.* *15*, 319–324.
- Ward, P.S., and Thompson, C.B. (2012). Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer Cell* *21*, 297–308.