Cancer genetics-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer

Igor Cima[a,1], Ralph Schiess[b,1], Peter Wild[c], Martin Kaelin[d], Peter Schüffler[e], Vinzenz Lange[b], Paola Picotti[b], Reto Ossola[b], Arnoud Templeton[d], Olga Schubert[a], Thomas Fuchs[e], Thomas Leippold[f], Stephen Wyler[f], Jens Zehetner[a], Wolfram Jochum[g], Joachim Buhmann[e], Thomas Cerny[d], Holger Moch[c,2], Silke Gillessen[d,2], Ruedi Aebersold[b,h,2], and Wilhelm Krek[a,2]

[a]Institute of Cell Biology and [b]Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule Zurich, 8093 Zurich, Switzerland; [c]Institute of Surgical Pathology, University Hospital Zurich, 8091 Zurich, Switzerland; [d]Department of Medical Oncology, [f]Department of Urology, and [g]Institute of Pathology, Kantonsspital St. Gallen, 9007 St. Gallen, Switzerland; [e]Department of Computer Science, Eidgenössische Technische Hochschule Zurich, 8092 Zurich, Switzerland; and [h]Faculty of Science, University of Zurich, 8057 Zurich, Switzerland

Edited by Owen N. Witte, The Howard Hughes Medical Institute, University of California, Los Angeles, CA, and approved January 12, 2011 (received for review September 25, 2010)

A key barrier to the realization of personalized medicine for cancer is the identification of biomarkers. Here we describe a two-stage strategy for the discovery of serum biomarker signatures corresponding to specific cancer-causing mutations and its application to prostate cancer (PCa) in the context of the commonly occurring phosphatase and tensin homolog (*PTEN*) tumor-suppressor gene inactivation. In the first stage of our approach, we identified 775 *N*-linked glycoproteins from sera and prostate tissue of wild-type and *Pten*-null mice. Using label-free quantitative proteomics, we showed that *Pten* inactivation leads to measurable perturbations in the murine prostate and serum glycoproteome. Following bioinformatic prioritization, in a second stage we applied targeted proteomics to detect and quantify 39 human ortholog candidate biomarkers in the sera of PCa patients and control individuals. The resulting proteomic profiles were analyzed by machine learning to build predictive regression models for tissue *PTEN* status and diagnosis and grading of PCa. Our approach suggests a general path to rational cancer biomarker discovery and initial validation guided by cancer genetics and based on the integration of experimental mouse models, proteomics-based technologies, and computational modeling.

serum biomarkers | mass spectrometry | Pten conditional knockout mouse model

Molecular and genetic biomarkers play a paramount role in clinical oncology. They can help predict who will develop cancer or detect the disease at an early stage. Biomarkers also can guide treatment decisions and help identify subpopulations of patients who are most likely to respond to a specific therapy (1, 2). However, the noninvasive detection and prognostic evaluation of a specific tumor by the analysis of indicators in body fluids such as serum remains a formidable challenge. Novel biomarkers represent today an urgent and critical medical need.

Serum has long been considered a rich source for biomarkers (3). However, the discovery of serum biomarkers has been technically challenging and ineffectual for reasons that include the particular and variable composition of the serum proteome and its enormous complexity (4). As the genetic alterations that cause cancer are becoming better understood, one strategy to overcome the limitations of the traditional serum proteome comparisons is to use the knowledge about specific cancer-causing mutations and the underlying disrupted signaling pathways to guide the discovery of novel cancer serum biomarkers.

The tumor-suppressor gene phosphatase and tensin homolog (*PTEN*) is one of the most commonly inactivated genes in human cancer and has been identified as lost or mutated in several sporadic cancers, including endometrial carcinoma, glioblastoma, breast cancer, and prostate cancer (5). An established consequence of *PTEN* inactivation is the constitutive aberrant activation of the PI3K-signaling pathway that drives uncontrolled cell growth, proliferation, and survival (6, 7). It is expected that specific signaling pathway-activating mutations such as *PTEN* loss will produce changes in the surface and secreted proteomes of the affected tissue (8), and, in principle, these changes should be detectable as discrete biomarker signatures in the serum. Based on this conceptual consideration, we developed a two-stage strategy for the discovery and initial validation of serum biomarkers in humans based on a mouse model of prostate cancer (PCa) progression caused by *Pten* inactivation.

Results

In the first stage of our approach, we identified PCa candidate biomarkers by applying a large-scale quantitative proteomic screen to detect and quantify *N*-linked glycoproteins that differ in their amount in the prostate tissue and sera of prostatic *Pten*-deficient PbCre4-*Pten*$^{fl/fl}$ (*Pten* cKO) and littermate control animals (9) (Fig. 1*A* and Fig. S1 *A–C*). The choice of an experimental mouse model as entry point for the identification of candidate biomarkers was guided by the possibility of collecting tissue samples from a genetically defined and homogeneous population in which variables such as environmental factors, age, and tumor type and stage are controlled and standardized. We selectively analyzed *N*-glycosylated proteins to maximize their subsequent detectability in the serum (10) and to focus on a subproteome that is enriched for validated serum biomarkers. In fact, 30 of the 38 protein biomarkers currently used in the clinic are glycosylated (11).

Author contributions: I.C., R.S., W.J., J.B., T.C., H.M., S.G., R.A., and W.K. designed research; I.C., R.S., P.W., M.K., P.S., V.L., P.P., R.O., A.T., O.S., T.L., S.W., and J.Z. performed research; I.C., R.S., P.W., M.K., P.S., V.L., and T.F. analyzed data; and I.C., R.S., S.G., R.A., and W.K. wrote the paper.

Conflict of interest statement: W.K., R.A., S.G., T.C., I.C., and R.S. are named as coinventors on a patent application related to the use of serum signatures for the diagnosis of prostate cancer. W.K., T.C., R.A., and R.S. are cofounders of ProteoMediX, a spin-off company with the aim of marketing products related to the diagnosis of prostate cancer. W.K., R.A., T.C., and R.S. own shares in ProteoMediX.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1] I.C. and R.S. contributed equally to this work.

[2] To whom correspondence may be addressed. E-mail: holger.moch@usz.ch, silke.gillessen@kssg.ch, aebersold@imsb.biol.ethz.ch, or wilhelm.krek@cell.biol.ethz.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1013699108/-/DCSupplemental.

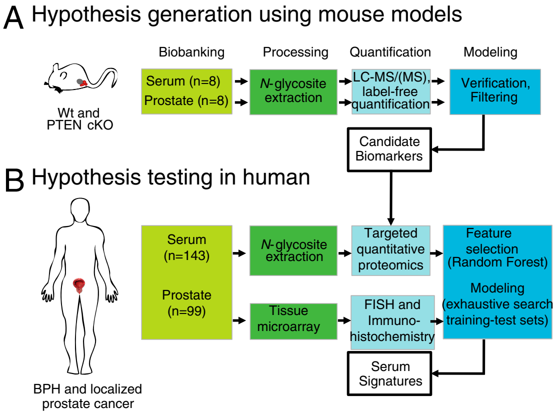
A Hypothesis generation using mouse models
Biobanking
Processing
Quantification
Modeling
Wt and PTEN cKO
Serum (n=8)
Prostate (n=8)
N-glycosite extraction
LC-MS/(MS), label-free quantification
Verification, Filtering
Candidate Biomarkers
B Hypothesis testing in human
Serum (n=143)
Prostate (n=99)
N-glycosite extraction
Targeted quantitative proteomics
Feature selection (Random Forest)
Modeling (exhaustive search training-test sets)
Tissue microarray
FISH and Immuno-histochemistry
Serum Signatures
BPH and localized prostate cancer

Fig. 1. Translational approach for biomarker discovery and validation workflow. (*A*) Candidate biomarkers are discovered using a genetic mouse model by enriching *N*-linked glycoproteins to sera and freshly isolated perfused prostates from wild-type and *Pten* cKO mice. Tryptic *N*-glycosites then are measured by LC-MS/MS. Identification and quantitation of proteins is performed as described. After a filtering process, candidate biomarkers are selected for the verification phase. (*B*) Verification phase. Highly standardized biobanking and clinical data collection are used for collecting serum and matching tissue samples from patients harboring localized PCa and control patients with BPH. *N*-linked glycoproteins are extracted as in *A*, and selected candidates from the discovery phase are measured by targeted proteomics and ELISA. At the same time, tissues are spotted as microarray and stained for the indicated antigens. Feature selection and modeling then is performed to find novel biomarkers for diagnosis, patient stratification by Gleason score, and *PTEN* status.

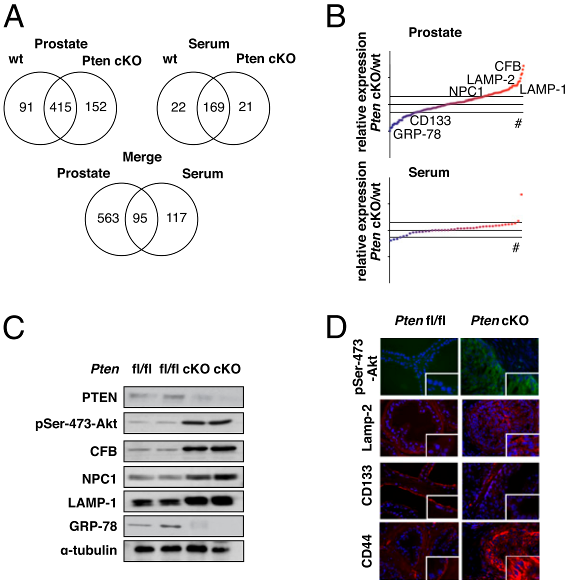

Fig. 2. Murine prostate and serum *N*-linked glycoproteome. (*A*) Venn diagrams of the mouse prostate and serum glycoproteome identified in the wild-type and *Pten* cKO mice indicating proteins commonly detected or detected only in the respective genotype/organs. (*B*) Label-free quantification of the proteins by means of SuperHirn plotted for prostate and serum. Dots indicate the ratio for each protein between the *Pten* cKO and wild-type prostates or sera and indexed from the most down-modulated to the most up-regulated. (*C* and *D*) Previously unknown *Pten*-dependent changes in protein expression are verified by standard cell biology techniques such as Western blot (*C*) and immunofluorescence (*D*).

Purified tissue and serum samples from wild-type and prostate-specific *Pten* cKO animals were subjected to solid-phase extraction of *N*-glycopeptides (SPEG) (12). The enriched *N*-glycosites from each sample were analyzed in duplicate on a linear ion trap quadrupole-Fourier transform (LTQ-FT) mass spectrometer. Glycoprotein enrichment in combination with high-throughput MS at a false-discovery rate ≤1% resulted in the identification of a total of 775 glycoproteins (Fig. 2*A* and Dataset S1). Seventy-seven percent of proteins detected in prostate tissue and up to 92% of proteins identified in serum were manually annotated (13) to be secreted, to reside in/on the plasma membrane, or to belong to the secretory or lysosomal compartments, indicating strong enrichment for the intended target population (14) (Fig. S1*D*). Of the 658 proteins identified in the prostate tissue, 152 (23.1%) were identified exclusively in the *Pten* cKO prostate cancer tissue, and 91 (13.8%) were identified only in the wild-type tissue (Fig. 2*A*).

To detect *Pten*-dependent changes in the *N*-glycosite profiles of prostate tissue and serum respectively, we compared the liquid chromatography (LC)-MS feature maps of the corresponding samples using the SuperHirn software (15). The ion chromatograms for each *N*-glycosite were used subsequently as the basis for relative label-free quantification of 352 proteins (213 from tissue, 105 from serum, and 34 from both samples). The relative amounts of 68 proteins differed significantly in the tissue of *Pten* cKO mice compared with their age-matched controls. In the matched sera only 12 proteins with significantly altered abundance were detected (Fig. 2*B* and Dataset S2), a result that is in agreement with previous ineffective direct serum–proteome comparisons. Spectral counting (16) confirmed the quantitative data derived from the ion counts and identified a further 43 proteins with significantly altered abundance between *Pten* cKO and wild-type prostate tissue (Dataset S2). Moreover, immunoblotting and immunofluorescence microscopy of selected glycosylated proteins verified the quantitative MS data (Fig. 2 *C* and *D*). Interestingly, our tissue glycoprotein screen identified differential expression of proteins known to be associated with differentiation and the stem cell phenotype (Fig. 2*D*). Further analyses of differentiation and stem cell markers confirmed this hypothesis (Fig S1 *D* and *E*), in agreement with previous reports (17) that implicate a role for PTEN in differentiation and stem cell homeostasis during PCa progression. The candidate biomarker list was analyzed further based on a series of three criteria: *Pten* dependency, prostate specificity, and detectability in serum. This analysis resulted in a list of 126 proteins (Dataset S2), which is expected to contain one or more specific candidate biomarker signatures that mirror *PTEN*-loss in human PCa.

Therefore, in the second stage of our approach we tested whether *PTEN*-inactivation in human PCa is associated with a specific serum signature (Fig. 1*B*). Between the years 2004 and 2007, using a standardized protocol, we collected prostate tissue samples of consenting patients who underwent biopsy, radical prostatectomy, or transurethral resection because of PCa and matched sera from a single source. We collected sera from a total of 143 patients. As control group (n = 66; median age = 65.7 y; range, 50.8–90.26 y), we selected patients with histologically confirmed benign prostatic hyperplasia (BPH). The PCa group (n = 77; median age = 67.7 y; range, 49.1–89.4 y) had histologically confirmed localized PCa (locPCa). Patients with other malignancies or with chronic or acute inflammatory conditions and patients with advanced prostate cancer were excluded from our analysis. From 99 patients (BPH, n = 40; locPCa, n = 59) we had also access to the corresponding prostate tissue samples. Of these, 92 samples (BPH, n = 40; locPCa, n = 52) were spotted on a tissue microarray (referred to thereafter as TMA-P92) for genetic and immunohistochemical analyses (Fig. S2*A*).

We analyzed the epithelial *PTEN* status by dual-color FISH on the TMA-P92 by calculating the percentage of epithelial cells that lost at least one *PTEN* gene copy number on each spot. To this end we compared *PTEN* gene copy numbers with total chromosome 10 copy numbers per cell using commercially available fluorescently labeled DNA probes for cytoband 10q23.3 and region 10p11.1~q11.1, respectively. We also stained sections from the TMA-P92 with antibodies reporting the activation state of the PI3K-pathway including phospho-serine (pSer)-473-Akt and stathmin (18). Seventy-two percent of prostate cancers displayed focal loss of *PTEN* gene copy numbers compared with the control group, indicating deletion of one or both alleles of *PTEN* (Fig. 3*A*, and Table S1) in at least 20% of the cells analyzed. This result is in agreement with previous reports using the same technique (19, 20). As expected, a significant fraction of these cancers demonstrated PI3K-pathway activation, as evidenced by the increased staining of pSer-473-Akt and stathmin (Fig. 3 *B* and *C*).

Next, we analyzed serum samples from these patients by using *N*-glycosite extraction followed by targeted quantitative MS via selected reaction monitoring (SRM) (21). To this end, we used a hybrid quadrupole/linear ion trap mass spectrometer (22) to detect and quantify 57 *N*-glycosites, corresponding to 49 candidate protein biomarkers present on the list of prioritized candidates. The absolute serum concentrations of these proteins were determined using stable isotope-labeled reference peptides as external standards. Of the 57 targeted peptides, 36 peptides representing 33 different proteins were detected consistently and quantified in 80–105 patients (Dataset S3). The median concentration of the measured proteins varied from 320 μg/mL to 5.5 ng/mL, indicating that our approach allows the quantification of protein concentrations in sera along six orders of magnitude. The median concentration of various measured proteins was in the concentration range of prostate-specific antigen (PSA), a widely used diagnostic serum biomarker for prostate cancer. For nine proteins, we also established ELISAs, which confirmed the validity of the SRM approach for two proteins and provided independent quantitative data for the other proteins that were not detected by SRM, thus resulting in a total of 39 proteins that were quantified consistently (Dataset S3). Next, we used this dataset to select the best candidate biomarkers and to build predictive models for the discrimination between normal and aberrant *PTEN* status. First, we applied the random forest (RF) classifier algorithm (23) for variable ranking and subsequent selection. RF is particularly well-suited in this regard, because it does not assume that the data are linearly separable. Moreover, the selection of the top-ranked variables reduces the dimensionality of the feature space and the computing time, thus allowing a subsequent exhaustive screening of the best models.

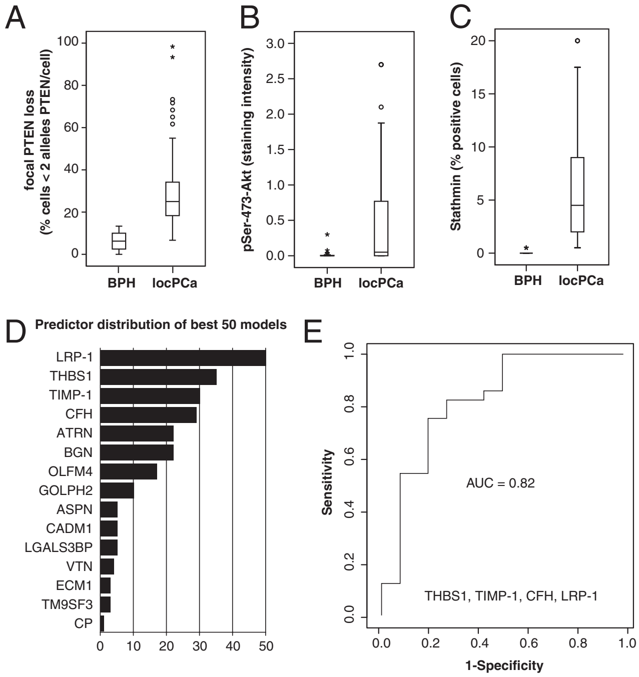

Fig. 3. *PTEN* score status predictors modeling. (*A*) *PTEN* FISH boxplot for BPH and locPCa cases. Focal *PTEN* loss indicates the percent of cells with reduced *PTEN* FISH signals compared with the centromere signal in the analyzed epithelial cells ($n = 75$). (*B*) pSer-473-Akt immunostaining boxplot. The score indicates the average staining intensity on a scale of 0–3 multiplied by the percentage of positive epithelial cells in BPH and locPCa tissues ($n = 92$). (*C*) Boxplot showing the staining intensity of stathmin, a marker for the PTEN/PI3K signature on BPH and locPCa tissues ($n = 92$). Numbers indicate the percentage of positive epithelial cells. (*D*) Predictor variable distribution for genetic *PTEN* status. The predictive importance of LRP-1, THBS1, TIMP-1, CFH, Attractin (ATRN), BGN, OLFM4, Golgi membrane protein 1 (GOLPH2), ASPN, Cell adhesion molecule 1 (CADM1), Galectin-3-binding protein (LGALS3BP), Vitronectin (VTN), ECM1, Transmembrane 9 superfamily member 3 (TM9SF3), and Ceruloplasmin (CP) selected by random forests and subsequent bootstrapped exhaustive search is based on the frequency in which the candidates appear in the best 50 predicting models. (*E*) ROC curve for the best predicting signature for tissue *PTEN* status ($n = 54$). In boxplots (*A*, *B*, and *C*), the line within the box indicates the median value, the box spans the interquartile range, whiskers extend to data extremes, and asterisks are outliers >3× interquartile range.

We selected the 20 top-ranked variables resulting from 100 RFs and screened for all logistic regression models to predict focal loss of *PTEN* by combining one to five serum proteins. This screening resulted in 21,699 different models, which were validated by 100-fold bootstrapping (24). For each model we calculated the median area under the receiver operating characteristic (ROC) curve (AUC), thereby identifying the best regression models that are able to predict significantly aberrant *PTEN* status from an overlapping data set comprising 54 patients derived from the *PTEN* FISH analysis of 82 patients and the SRM and ELISA quantification of sera from 105 patients (*PTEN* focal loss <20%: $n = 26$; PTEN focal loss ≥20%: $n = 28$) (Fig. S2*B*). The signature comprising thrombospondin-1 (THBS1), metalloproteinase inhibitor 1 (TIMP-1), complement factor H (CFH), and prolow-density lipoprotein receptor-related protein 1 (LRP-1) could predict correctly 78% of cases belonging to patients having aberrant or normal *PTEN* status with a sensitivity of 79.2% and specificity of 76.7% [AUC = 0.82; $P = 5.49*10E-5$; 95% confidence interval (CI) = 0.704–0.936] (Fig. 3*E*). Taken together, these results suggest that the reduction of *PTEN* gene copy number in prostate cancer led to a measurable perturbation of the serum proteome. Moreover, they demonstrate the usefulness of computational variable selection using RF followed by exhaustive regression model screening as a valid approach to extract information on candidate biomarkers.

To corroborate this analysis, we determined the occurrence of the 15 RF-selected top-ranked variables in the best 50 bootstrapped models (Fig. 3*D*). All proteins have been selected in more than half of all highly predictive models. This approach thus provided the theoretical robustness of discrimination of individual candidate biomarkers described in Fig. 3*E*. To determine whether our signature is significantly linked to the PTEN network, we sought curated knowledge-based connections between our signature and the PTEN network. When tested against 50 random signatures, the PTEN signature identified here showed significantly more direct and indirect connections to the PTEN signaling network, thus providing independent support for our serum signature as a predictor of tissue *PTEN* status (Fig. S3 *A* and *B*). Because *PTEN* loss is causally associated with accelerated PCa progression and aggressiveness, as exemplified by the association between *PTEN* loss of function and Gleason sum score (25, 26), we next asked whether the bioinformatic approach also could extract serum protein signatures reflecting tumor grading. Examination of our TMA-P92 revealed a correlation between *PTEN* loss and Gleason score sum (Fig. S4), confirming previous reports (25, 26). The Gleason grading available for 69 tumors and the corresponding quantitative SRM serum analysis served as the basis for applying the bioinformatics approach outlined above (Fig. S2*B*). Intriguingly, we identified a five-protein signature from an overlapping dataset of 54 patients comprising polypeptide GalNAc transferase-like protein 4 (GALNTL4), fibronectin (FN), zinc-α-2-glycoprotein (AZGP1), biglycan (BGN), and extracellular matrix protein 1 (ECM1) that predicted patients having tumors with a Gleason score <7 or ≥7 with an AUC = 0.788 [*P* = 3.1*10E-4; 95% CI = 0.668–0.907; sensitivity (sens.) = 60.9%; specificity (spec.) = 67.8%] (Fig. 4*B*). The predictive relevance ranking corroborated the composition of the best signature (Fig. 4*A*). These results imply a potential link between aberrant *PTEN* status and the emergence of protein signatures in the serum reporting on tumor grading. Taken together, because only few reports describe serum biomarkers for the stratification of patients based on the grading of the tumors (2, 27, 28), the data suggest an application of our biomarker discovery platform for the prognosis of locPCa, wherein patients with clinically significant or insignificant prostate cancer can undergo stratification for therapy or watchful waiting, respectively (29). Finally, we assessed whether our approach can reveal signatures for PCa diagnosis. As reported previously, we note that the vast majority of the tumors show aberrant focal *PTEN* loss (Fig. 3*A* and Table S1) and altered PI3K signaling (Fig. 3 *B* and *C*). The current method of choice for noninvasive screening of PCa is the blood-based quantification of PSA together with digital rectal examination (DRE). Recent studies showing that PSA, alone or in combination with DRE, is prone to overdiagnosis and has no or very limited beneficial effects on overall survival (30, 31) suggest a strong need for better diagnostic signatures. We thus analyzed a total of 143 sera from 77 PCa patients and 66 controls. Sera from 105 patients were selected as training-validation set (Fig. S2*B*). Machine learning analysis applied to a quantitative data set derived from SRM analysis of the sera of 82 patients identified a four-protein signature comprised of hypoxia up-regulated protein 1 (HYOU1), asporin (ASPN), cathepsin D (CTSD), and olfactomedin-4 (OLFM4) (32). This signature discriminated between BPH and PCa with an AUC = 0.726 (*P* = 0.01; 95% CI = 0.614–0.838; sens. 81%, spec. 57%). PSA measurements resulted in a similar AUC = 0.730 (*P* = 1*10E-6; 95% CI = 0.693–0.871; sens. 78%, spec. 63%). Strikingly, the combination of the four-protein signature with PSA resulted in an AUC = 0.840 (*P* = 5*10E-9; 95% CI = 0.824–0.964; sens. 85%, spec. 79%) (Fig. 4 *C* and *D*). With the aim of testing the reproducibility of our approach, we measured the four-biomarker signature by SRM in an independent test set comprising 38 patients that were not included in the training-validation set. In the test set, the diagnostic signature from an overlapping dataset of 37 patients performed equally as well as the training-validation set, indicating reproducibility and robustness of the test as well as of the measurement procedure (Fig. 4*E*). To exclude confounding variables such as inflammatory conditions as the origin of eventual bias (33) in our analysis, we correlated the single biomarkers comprised in the prognostic and diagnostic signatures mentioned above with clinical parameters of inflammatory state in a subset of patients, independently of the disease status. Specifically, we correlated C-reactive protein (CRP) and the leukocyte count. All the selected variables failed to correlate with either parameter, thus excluding a bias derived from the inflammatory status of the patient at the time of diagnosis (Fig. S5).

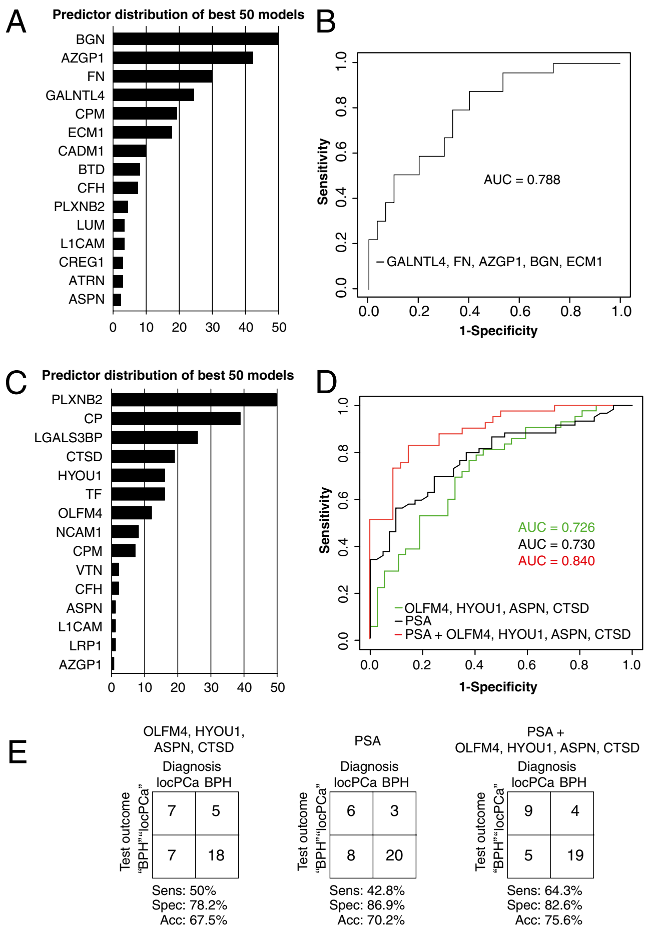

Fig. 4. Candidate biomarkers for diagnosis and Gleason score prediction. (*A*) Predictor variable distribution for Gleason score. The predictive importance for BGN, AZGP1, FN, GALNTL4, Carboxypeptidase M (CPM), ECM1, CADM1, Biotinidase (BTD), Complement factor H (CFH), Plexin B2 (PLXNB2), Lumican (LUM), Neural cell adhesion molecule L1 (L1CAM), Protein CREG1 (CREG1), ATRN, and ASPN selected by random forests and subsequent bootstrapped exhaustive search is based on the frequency in which the candidate appears in the first 50 models. (*B*) ROC curve for the indicated signature for prediction of Gleason score <7 or ≥7 (*n* = 54). (*C*) Predictor variable distribution for diagnosis between locPCa and BPH. (*D*) ROC curve showing the performance for the selected signature (green line), the signature combined with PSA (red line), and PSA alone (black line) (*n* = 105). (*E*) Independent test set for the diagnostic signature. Performance of the identified diagnostic signature (*Left*), PSA alone (*Center*), or the combination of the signature and PSA (*Right*) in a set of patients measured independently and not considered in the training set. Data are presented as confusion matrices with the sensitivity (sens.), specificity (spec.), and accuracy (acc.) indicated for each signature (*n* = 37).

Discussion

The present study provides a general framework for rational cancer biomarker discovery. The underlying concept is that activation of cancer-signaling pathways caused, for example, by the inactivation of a defined tumor-suppressor gene is associated with specific protein signatures that can be measured in serum and potentially used to detect disease at an early stage or to derive information about the tumor grade and thus guide treatment decisions. In the past the discovery of serum biomarkers has been technically challenging because of the enormous complexity of the serum proteome and the lack of sensitive discovery-driven measurement technologies (4). Based on these considerations, we implemented a two-stage strategy for biomarker discovery. In the first stage, we generated a list of candidate biomarkers based on information derived from large-scale screens of the glycosylated proteome of a mouse model of PCa progression caused by prostate-specific *Pten* inactivation. This approach identified multiple proteins differentially regulated upon *Pten* inactivation. In the second, hypothesis-testing stage, the candidate proteins were quantified simultaneously by SRM-based targeted MS in sera of PCa patients and integrated with information derived from matched PCa tissue characterized with respect to *PTEN* status and PI3K-pathway activation. Using machine learning algorithms, we then extracted robust patterns suitable for predicting tissue *PTEN* status and for the diagnosis and grading of PCa.

The current standard biomarker for early detection of PCa is PSA. However, the effectiveness of systematic PCa screening with PSA testing remains controversial, in part because a lack of sensitivity and specificity results in considerable overdiagnosis and overtreatment (30, 31, 34). The ability of our four-protein signature for prostate cancer diagnosis to distinguish accurately between locPCa and BPH makes it potentially suited for screening tests by reducing false-positive outcomes and therefore avoiding anxiety and biopsies in men who have an elevated PSA but do not harbor cancer.

Another potential drawback of PSA testing relates to the detection of clinically insignificant prostate cancers in asymptomatic men. Therefore, overdiagnosis, in this context meaning detection of cancer that has no clinical impact on an individual during his lifetime, is a major problem. There are several definitions of insignificant prostate cancer (29); most definitions exclude patients with any Gleason pattern 4 prostate cancer. Our five-protein signature predictive for Gleason score therefore may be suited for improving screening efficacy by reporting which men might harbor insignificant cancers. Such patients might be offered active surveillance instead of immediate treatment. If active surveillance is chosen as the treatment option, repetitive biopsies for detection of prostate cancer progression ultimately could be replaced by a serum test.

The biomarker discovery platform presented here provides an approach for the discovery and validation of biomarkers with the aim of improving the effectiveness of PCa testing and noninvasive diagnostic of prostate cancer. Ideally this approach could avoid overdiagnosis and overtreatment and guide treatment decision.

The simultaneous analysis of a large number of candidate biomarkers by SRM allows the discovery of new potential biomarkers independently from the availability of established immunoassays. However, this emerging technology does not yet allow the analysis of large cohorts of patients. The identified candidate biomarkers thus must be validated further in larger, prospective studies, preferably using standardized immunoassays, which are limited in the amount of proteins analyzed per sample but allow the analysis of much larger cohorts. External independent data sets must be added as well to confirm further the clinical usefulness of the reported signatures.

Because we now are entering an era in which the genetic and epigenetic abnormalities responsible for specific forms of cancer guide the design of molecularly targeting drugs, efficient strategies to evaluate such targeted therapies in patients are critical, especially as more such compounds become available. Biomarkers able to identify reliably the patients who are most likely to benefit from a specific molecularly targeted therapy therefore would have significant clinical benefit. In this regard, an increasing armamentarium of targeting agents that inhibit key components of the PI3K pathway exists, and many of these inhibitors already are in clinical testing (35–37). It is conceivable that signatures reflecting tissue *PTEN* status may aid in selecting suitable patients and provide proof of target modulation by these inhibitors. Thus, a cancer genetics-guided path to biomarker discovery, as described here, may hold the promise for the realization of personalized cancer medicines.

Methods

Experimental Mice. PTEN cKO mice were generated as described in ref. 9. The Zurich cantonal veterinary office approved all animal studies. Details are provided in *SI Methods*.

Glycoprotein Enrichment from Murine Serum. Glycoproteins were enriched from sera and tissue of mice using the protocol published by Zhang et al. (12). Details about the isolation of sera and tissues from mice and the glycoprotein enrichment method are explained in *SI Methods*.

Mass Spectrometry Analysis. Samples were analyzed on a hybrid LTQ-FT mass spectrometer (Thermo Electron) equipped with a nanoelectrospray ion source. Chromatographic separation of peptides was performed on an Agilent 1100 micro HPLC system equipped with a 15-cm fused silica emitter, 150-μm inner diameter, packed with a Magic C18 AQ 5 μm resin (Michrom BioResources). Further details are provided in *SI Methods*.

Protein Identification. Proteins were identified following protocols described in refs. 12 and 36–38. Further details are provided in *SI Methods*.

Label-Free Quantification of Peptide and Protein Ratios. Data from LC-MS runs were converted from raw to the mzXML data format (38) and processed by the software tool SuperHirn as described previously (15). JRatio was used for the calculation and visual assessment of peptide and protein ratios (39). After examining the distribution of the data and assuming normality (Fig S1G), we applied a two-tailed Student's *t* test with unequal variances statistics to assess the significance of a protein fold-change. Protein fold-changes with a nonstringent P value $\leq$0.15 were selected for further analyses. To verify the results obtained by SuperHirn and JRatio, we performed spectral counting analysis (*SI Methods*). Prostate specificity was calculated from gene-expression profiles obtained for the BioGPS database (http://biogps.org) (40) by calculating the ratio of average gene expression from prostate and average gene expression of the remaining tissues. Gene with ratios of 10 or more were considered prostate specific.

SDS/PAGE and Western Blotting. Perfused fresh-frozen prostate tissues were solubilized in RIPA buffer (150 mM NaCl, 10 mM Tris, 0.1% SDS, 1% Triton X-100, 1% deoxycholate, 5 mM EDTA) plus protease inhibitors (1 mM phenylmethylsulfonyl fluoride, 10 mM benzamidine, 10 μg/mL aprotinin). Fifty-microgram protein extracts were resolved on 8–12% SDS/PAGE, blotted on nitrocellulose, and visualized by immunoblotting with the following primary antibodies: anti–phospho-Ser473 AKT (#4058; Cell Signaling Technology), anti-complement factor B (#HPA001817; Sigma Aldrich), anti-KDEL (#ab12223; Abcam), anti–Niemann-Pick C1 (#NB400-148SS; Novus Biologicals), anti–LAMP-1 (clone 1D4B; Developmental Studies Hybridoma Bank), anti–α-tubulin (clone YL 1/2, #ab6160; Abcam).

Immunofluorescence. Five-micrometer cryostat sections on poly-L-lysine slides were fixed in PBS/4% paraformaldehyde for 10 min, washed in PBS, and stained using antibodies against the indicated proteins. Further details are provided in *SI Methods*.

Real-Time PCR Analysis. Prostate tissues from three wild-type and three *PTEN* cKO animals were isolated as described. Total RNA was prepared from powdered tissue using the RNeasy Mini Kit (Qiagen), and cDNA was prepared using random hexanucleotide primers and Ready-to-go you-prime first-strand beads (GE Healthcare). Real-time PCR analysis of cDNA was performed using LightCycler 480 SYBR Green I Master from Roche and specific primers reported in *SI Methods*.

Patients, Sampling, and Handling of Human Sera and Glycoprotein Enrichment. The Ethics Committee of the Kanton St. Gallen, Switzerland, approved all procedures involving human material, and all patients signed an informed consent. For the study we included patients with locPCa and BPH. We excluded from the analysis patients with advanced prostate cancer, infectious or inflammatory diseases, or other malignancies. Eight milliliters of blood were drawn and collected in a serum separator tube containing clot activator and gel (Vacutainer, SSTTM II Advance, REF 367953; Becton Dickinson). Tubes were inverted eight times and centrifuged within 4 h of collection at 4 °C for 10 min at 1,428 × *g*. The serum was divided into five aliquots of 500 μL each and stored at −60 °C or lower until use. Glycoprotein extraction was performed exactly as described for the murine serum.

Targeted MS Analysis Using SRM. We used the absolute quantification of proteins (AQUA) strategy introduced by Gerber et al. (41). Further details are reported in *SI Methods*.

Tissue Microarray Preparation. A tissue microarray was constructed as described (42) using formalin-fixed, paraffin-embedded tissue samples derived from 92 patients (BPH, *n* = 40; locPCa, *n* = 52) with matched serum samples that were used for SRM or ELISA. Details are provided in *SI Methods*.

Immunohistochemistry. Immunohistochemistry was performed using a Ventana Benchmark automated staining system (Ventana Medical Systems) and the following primary antibodies: anti–phospho-Ser473 AKT (dilution 1:150; #ab8932; Abcam) and and anti-Stathmin (dilution 1:50; #3352; Cell Signaling Technology).

FISH. To assess PTEN deletion, we performed dual-color FISH on paraffin-embedded tissue using commercially available fluorescently labeled DNA probes for cytoband 10q23.3 (SpectrumOrange, PTEN locus-specific probe) and region 10p11.1~q11.1 (Spectrum-Green centromere of chromosome 10 probe; LSI PTEN/CEP 10; Abbott Laboratories) according to the manufacturer's instructions. Details are provided in *SI Methods*.

Bioinformatic Analysis of SRM Data. SRM data were normalized and subjected to feature selection using random forest followed by signature modeling using brute force search for all logistic models. AUCs for every model were calculated by bootstrapping to avoid overfitting. Details are provided in *SI Methods*.

ELISA. The concentration of selected candidate biomarkers (Dataset S3) was measured by sandwich or competitive ELISA following the manufacturer's instructions. Details are given in *SI Methods*.

ACKNOWLEDGMENTS. We thank all members of our laboratories for helpful discussions and Corina Fusaro-Graf, Doris Kradolfer, Melanie Sticker, and Manuela Hitz for technical assistance. The project was supported by Grant 3100A0-107679 from the Swiss National Science Foundation (to R.A.), Grant KFP OCS 1262-06-2002 from the Collaborative Cancer Research Project from Oncosuisse (to T.C., S.G., W.K., and H.M.), Smart Inactivity Monitor using Array Based Detectors Grant 213250 from the Future and Emerging Technologies Programme within the Seventh Framework Programme for Research of the European Commission (to P.S., T.F., and J.B.), and by a grant from the Vontobel-Stiftung (to W.K.).

1. Tainsky MA (2009) Genomic and proteomic biomarkers for cancer: A multitude of opportunities. *Biochim Biophys Acta* 1796:176–193.
2. Ludwig JA, Weinstein JN (2005) Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* 5:845–856.
3. Srinivas PR, Kramer BS, Srivastava S (2001) Trends in biomarker research for cancer detection. *Lancet Oncol* 2:698–704.
4. Rifai N, Gillette MA, Carr SA (2006) Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nat Biotechnol* 24:971–983.
5. Steck PA, et al. (1997) Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nat Genet* 15: 356–362.
6. Maehama T, Dixon JE (1998) The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J Biol Chem* 273: 13375–13378.
7. Stambolic V, et al. (1998) Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell* 95:29–39.
8. Mehrian-Shai R, et al. (2007) Insulin growth factor-binding protein 2 is a candidate biomarker for PTEN status and PI3K/Akt pathway activation in glioblastoma and prostate cancer. *Proc Natl Acad Sci USA* 104:5563–5568.
9. Trotman LC, et al. (2003) Pten dose dictates cancer progression in the prostate. *PLoS Biol* 1:E59.
10. Zhang H, et al. (2007) Mass spectrometric detection of tissue proteins in plasma. *Mol Cell Proteomics* 6:64–71.
11. Schiess R, Wollscheid B, Aebersold R (2009) Targeted proteomic strategy for clinical biomarker discovery. *Mol Oncol* 3(1):33–44.
12. Zhang H, Li XJ, Martin DB, Aebersold R (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 21:660–666.
13. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2:953–971.
14. Roth J (2002) Protein N-glycosylation along the secretory pathway: Relationship to organelle topography and function, protein quality control, and cell interactions. *Chem Rev* 102:285–303.
15. Mueller LN, et al. (2007) SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* 7:3470–3480.
16. Liu H, Sadygov RG, Yates JR, 3rd (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76:4193–4201.
17. Wang S, et al. (2006) Pten deletion leads to the expansion of a prostatic stem/progenitor cell subpopulation and tumor initiation. *Proc Natl Acad Sci USA* 103: 1480–1485.
18. Saal LH, et al. (2007) Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc Natl Acad Sci USA* 104:7564–7569.
19. Sircar K, et al. (2009) PTEN genomic deletion is associated with p-Akt and AR signalling in poorer outcome, hormone refractory prostate cancer. *J Pathol* 218: 505–513.
20. Yoshimoto M, et al. (2006) Interphase FISH analysis of PTEN in histologic sections shows genomic deletions in 68% of primary prostate cancer and 23% of high-grade prostatic intra-epithelial neoplasias. *Cancer Genet Cytogenet* 169:128–137.
21. Addona T, et al. (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* 27(7):633–641.
22. Lange V, Picotti P, Domon B, Aebersold R (2008) Selected reaction monitoring for quantitative proteomics: A tutorial. *Mol Syst Biol* 4:222.
23. Breiman L (2001) Random forests. *Mach Learn* 45:5–32.
24. Efron B, Tibshirani R (1993) *An Introduction to the Bootstrap* (Chapman & Hall, New York).
25. McMenamin ME, et al. (1999) Loss of PTEN expression in paraffin-embedded primary prostate cancer correlates with high Gleason score and advanced stage. *Cancer Res* 59:4291–4296.
26. Dreher T, et al. (2004) Reduction of PTEN and p27kip1 expression correlates with tumor grade in prostate cancer. Analysis in radical prostatectomy specimens and needle biopsies. *Virchows Arch* 444:509–517.
27. Nakanishi H, et al. (2008) PCA3 molecular urine assay correlates with prostate cancer tumor volume: Implication in selecting candidates for active surveillance. *J Urol* 179 (5):1804–1809; discussion 1809–1810.
28. Sreekumar A, et al. (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 457:910–914.
29. Bastian PJ, et al. (2009) Insignificant prostate cancer and active surveillance: From definition to clinical implications. *Eur Urol* 55:1321–1330.
30. Schröder FH, et al.; ERSPC Investigators (2009) Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 360:1320–1328.
31. Andriole GL, et al.; PLCO Project Team (2009) Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* 360:1310–1319.
32. Zhang J, et al. (2002) Identification and characterization of a novel member of olfactomedin-related protein family, hGC-1, expressed during myeloid lineage development. *Gene* 283:83–93.
33. Ransohoff DF (2005) Bias as a threat to the validity of cancer molecular-marker research. *Nat Rev Cancer* 5:142–149.
34. Neal DE, Donovan JL, Martin RM, Hamdy FC (2009) Screening for prostate cancer remains controversial. *Lancet* 374:1482–1483.
35. Engelman JA (2009) Targeting PI3K signalling in cancer: Opportunities, challenges and limitations. *Nat Rev Cancer* 9:550–562.
36. Liu P, Cheng H, Roberts TM, Zhao JJ (2009) Targeting the phosphoinositide 3-kinase pathway in cancer. *Nat Rev Drug Discov* 8:627–644.
37. Sarker D, Reid AH, Yap TA, de Bono JS (2009) Targeting the PI3K/AKT pathway for the treatment of prostate cancer. *Clin Cancer Res* 15:4799–4805.
38. Pedrioli PGA, et al. (2004) A common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotechnol* 22:1459–1466.
39. Schiess R, et al. (2009) Analysis of cell surface proteome changes via label-free, quantitative mass spectrometry. *Mol Cell Proteomics* 8(4):624–638.
40. Su AI, et al. (2002) Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci USA* 99:4465–4470.
41. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA* 100:6940–6945.
44. Kononen J, et al. (1998) Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 4:844–847.