Supporting information

Carlsson et al. 10.1073/pnas.1103125108

SI Materials and Methods

Sample Preparation. The serum samples were biotinylated using a protocol previously optimized for labeling the serum proteome (1). All serum samples were labeled, using EZ-Link Sulfo-NHS-LC-Biotin (Pierce), as previously described (1). Briefly, 50-μL serum aliquots were centrifuged and diluted 1:45 in PBS, resulting in a final protein concentration of about 2 mg/mL. Sulfo-NHS-biotin then was added to a final concentration of 0.6 mM, and the samples were incubated on ice for 2 h. Unconjugated biotin was removed by dialysis against PBS at 4 °C for 72 h. Finally, the samples were aliquoted and stored at −20 °C before use.

Production and Purification of Recombinant Antibody Fragments. One hundred thirty-five human recombinant single-chain Fv (scFv) antibody fragments were selected against 65 antigens, mainly against immunoregulatory proteins, using a phage display library (Table 2) (2). The selection criteria were stringent, using standard operating procedures, to ensure the correct antibody specificity. All scFv probes were produced in 100-mL *Escherichia coli* cultures and purified from either expression supernatants or periplasmic preparations using affinity chromatography on Ni^{2+}-NTA agarose (Qiagen). Elution was performed using 250 mM imidazole followed by extensive dialysis against PBS. The protein concentration was determined by measuring the absorbance at 280 nm, and the purified scFv was stored at 4 °C until further use.

Fabrication and Processing of Antibody Microarrays. The production and handling of the antibody microarrays was performed according to a previously optimized set-up (1, 3). Briefly, the scFv microarrays were fabricated, using a noncontact printer (Biochip Arrayer1; Perkin-Elmer Life and Analytical Sciences), which deposits ∼330 pL per drop using piezo technology. The scFv antibodies were arrayed by spotting two drops at each position; the first drop was allowed to dry before the second drop was dispensed. Black polymer MaxiSorp microarray slides (NUNC A/S) were used for solid support, and each antibody was arrayed in eight replicates to ensure adequate statistics. To assist in the grid alignment during the subsequent quantification, a row containing Alexa Fluor 647-conjugated streptavidin (2 μg/mL) was spotted as the top row in all eight subarrays that constituted the array. The slides were blocked with 5% (wt/vol) fat-free milk powder (Semper AB) in PBS overnight and then were placed in a ProteinArray Workstation (Perkin-Elmer Life and Analytical Sciences), where they were washed for 4 min at 60 μL/min with 0.05% Tween-20 in PBS. Thereafter, 75 μL of the labeled sample, diluted 1:2 (total dilution 1:90) in 1% (wt/vol) fat-free milk powder and 1% (vol/vol) Tween 20, was injected and agitated over the array every 15 s for 60 min. After another 4-min wash, the arrays were incubated for 60 min with 350 μL of 1 μg/mL Alexa Fluor 647-conjugated streptavidin in PBS, with 1% (wt/vol) fat-free milk powder and 1% Tween 20. Finally, after a last washing step, the arrays were dried under a stream of nitrogen gas and were scanned with a confocal microarray scanner (ScanArray Express; Perkin-Elmer Life and Analytical Sciences) at 5-μm resolution, using three different scanner settings.

The ScanArray Express software version 4.0 (Perkin-Elmer Life & Analytical Sciences) was used to quantify the intensity of each spot, using the fixed-circle method. The local background was subtracted, and to compensate for possible local defects, the two highest and two lowest replicates were excluded automatically. In all further data analysis, each data point represents the mean value from the remaining four replicate spots. For protein analytes displaying saturated signals, values from lower scanner settings were used.

Microarray Data Normalization. Chip-to-chip normalization of the dataset was performed using a semiglobal normalization approach, similar to the normalization method used for DNA microarrays. To find a scaling factor (4–6), the coefficient of variation (CV) was calculated for each antibody, and the 15% of antibodies displaying the lowest CV values over all samples were identified, corresponding to 20 analytes. The normalization factor N_i was calculated by the formula $N_i = S_i/\mu$, where S_i is the sum of the signal intensities of the 20 analytes for each sample, and μ is the average of S_i from all samples. To normalize samples, all antibody intensities in a sample were divided by its normalization factor N_i.

Data Analysis. ***Classifier calibration and independent testing.*** To reduce sample-to-sample variations, differences between $t = 0$ and $t = 3$- to 6-mo protein expressions were used as independent variables for a support vector machine (SVM) classifier. We trained and tested SVMs with a leave-one-out procedure, using $n = 38$ samples and initially all $M = 135$ variables. Because the number of antibodies exceeded the number of samples, we needed to eliminate the antibodies with low impact on the predictions to avoid fitting to noise because of random correlations. This procedure was followed for every leave-one-out sample using a backward-elimination procedure. Performance was measured by the Receiver Operating Characteristics (ROC) area. The ranked antibody lists resulting from each SVM model then were fused into a consensus signature that was used with a frozen SVM in an independent test set. The different steps are described in some detail below, and the full procedure is summarized in Fig. S3.

Leave-one-out procedure. The principle for the leave-one-out procedure for N samples belonging to two classes is to train and test an SVM using leave-one-out cross-validation, i.e., to train the SVM on all but one sample and to test the resulting model using the omitted sample. The test sample is assigned a decision value using the trained SVM model and is put back into the training set; then the next sample is left out and used as a test sample. The procedure is repeated until each sample has been assigned a decision value once using all other samples as training set. The decision values for all samples are used to create an ROC curve, and the area under the curve (AUC) is calculated. The obtained area serves as an estimation of the expected area for this sample set.

Backward variable elimination. To reduce the number of antibodies used for the classification, we combined the leave-one-out procedure with a backward-elimination process for the antibodies. This process also produces a ranking of the antibodies with the purpose of assigning low ranks to randomly correlated antibodies.

The process is described in Fig.S3, which illustrates steps 1–7 discussed here. Starting with M antibodies, M datasets are created. In each dataset one antibody is replaced with a constant value, which is the average value of that antibody across all samples (step 1). To evaluate each antibody's importance for classification in the current dataset, an SVM leave-one-out procedure (as described above) is made for each of the M datasets. Subsequently, M ROC curves are created using the SVM output, and the dataset from which the ROC area generated had the smallest decrease (which could be negative) compared with the original is identified. The antibody set to a constant value in that dataset is identified and eliminated (step 2). The datasets now contain $M-1$ antibodies, and therefore $M-1$ new datasets are created, each having one of the remaining antibodies replaced with its average. The leave-one-

out testing procedure together with the SVM evaluation using constant antibodies is repeated, in effect eliminating the next antibody carrying the least information. The procedure is continued until only one antibody remains, resulting in a rank order for each antibody's importance in the classification of the samples currently in the dataset (step 3). This information can be used to build antibody subpanels of any desired length (step 4). To evaluate the predicting power of such an antibody panel, it is not possible to test a new model using the same samples in an unbiased way. Furthermore, randomly correlated antibodies still may obtain a high rank. Therefore, an additional leave-one-out loop (step 5a) is added in which one sample is removed before the initiation of the backward-elimination procedure, which therefore can be used as a test sample (step 5b). This outermost loop is iterated, leaving out each sample as a test set once and using the remaining samples to produce an antibody rank list. This rank list then will be used to pick a subpanel of antibodies with the highest rank and to train a single SVM model, which is tested on the test sample. The result of the outermost leave-one-out process is a list of decision values for all samples for any given subpanel length. The corresponding ROC area serves as a test of the performance using antibody subpanels of the given size, a test that can be used to estimate the number of antibodies required to make an adequate classification in the dataset.

Consensus antibody signature. The backward-elimination procedure results in the same number of antibody rank lists as the number of samples. To produce a single ranking order, the information from each run is concatenated into a consensus list by assigning each antibody a score (step 6) based on its average survival in the elimination rounds, where the antibody with the highest average survival is ranked as the most important. Finally, a new SVM is trained using the 21 antibodies with the best scores and is tested on the new, independent, dataset (step 7).

Analysis of sample size. The observed decision values from the SVM analysis displayed an SD of 1.13 and a delta value (the difference between mean values) between the groups of 1.14. A statistical power analysis with the alpha level (level of significance) set to 0.05 was performed to estimate the number of patients required in a second dataset (the prevalidation cohort). The analysis was performed, using the function "power.t.test" in R (decision values assumed normally distributed as suggested by Shapiro–Wilk testing) and showed that each group must contain a minimum of 8.7 patients (total of 17.4 patients) to reach a power of 80%. The 26 patients subsequently included in our prevalidation set yielded a statistical power of 93%.

Kaplan–Meier Analysis. The SVM decision value was used to divide the test cohort into two groups: a low-risk group (decision value ≤ 0) with 20 patients and a high-risk group (decision value >0) with 18 patients. With a total follow-up time of 5 y, Kaplan–Meier estimates of the tumor recurrence function for the two groups were used (Fig. S5). To indicate the significant difference between the two groups, a log rank test was used.

1. Wingren C, Ingvarsson J, Dexlin L, Szul D, Borrebaeck CAK (2007) Design of recombinant antibody microarrays for complex proteome analysis: Choice of sample labeling-tag and solid support. *Proteomics* 7:3055–3065.
2. Söderlind E, et al. (2000) Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nat Biotechnol* 18:852–856.
3. Ingvarsson J, et al. (2007) Design of recombinant antibody microarrays for serum protein profiling: Targeting of complement proteins. *J Proteome Res* 6:3527–3536.
4. Hamelinck D, et al. (2005) Optimized normalization for antibody microarrays and application to serum-protein profiling. *Mol Cell Proteomics* 4:773–784.
5. Ingvarsson J, et al. (2008) Detection of pancreatic cancer using antibody microarray-based serum protein profiling. *Proteomics* 8:2211–2219.
6. Carlsson A, et al. (2008) Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays. *Eur J Cancer* 44:472–480.

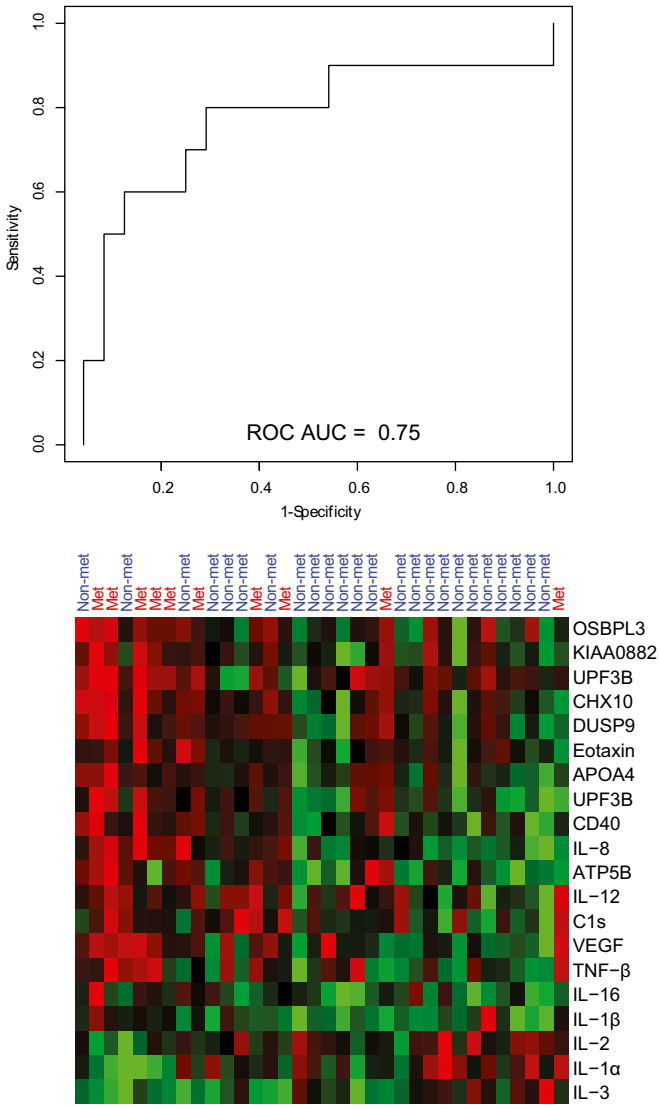

Fig. S1. SVM analysis for prediction of metastatic breast cancer using the 12-mo samples. Analyte velocities were fed to an SVM, which, using a leave-one-out cross-validation procedure, was calibrated to classify the patients. (*Upper*) The analysis yielded an AUC of 0.75. (*Lower*) The heat map shows all analytes displaying a Wilcoxon *P* value of <0.05, with red indicating an increase and green indicating a decrease in biomarker velocity.

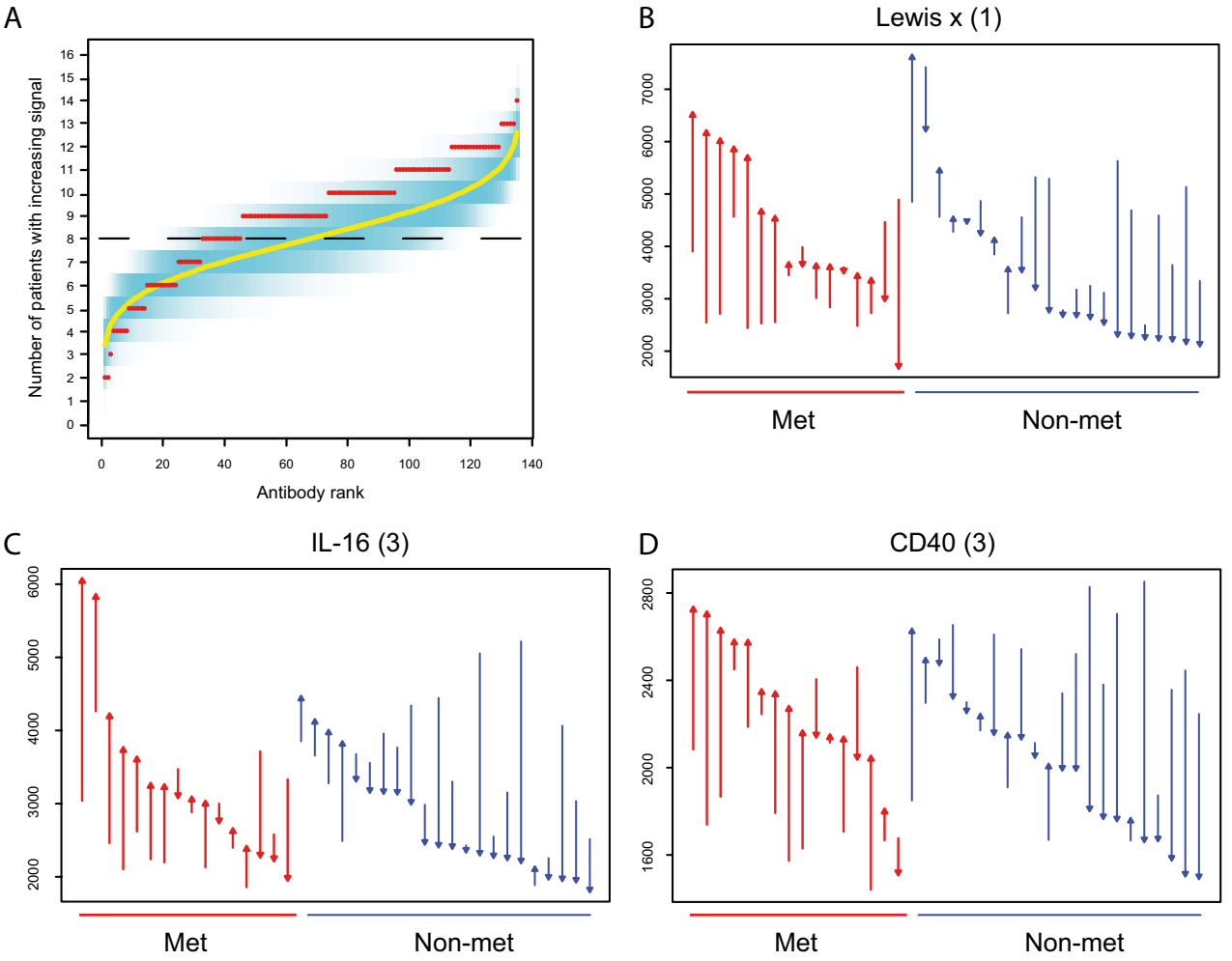

Fig. S2. Analysis of false-discovery rate and analyte velocity. (*A*) The number of patients with higher biomarker signals for the second sample was identified for each antibody. The antibodies were sorted accordingly and plotted (red), with the antibody with increasing signal in the fewest patients to the left. The first and second samples then were recombined randomly, treating the second sample as the first and vice versa, for a random set of patients. The number of patients with higher signal in the "second" sample was identified again, to achieve an estimation of the false-positive rate (FPR) of the analysis. This procedure was repeated 10,000 times. A heatmap indicates the frequency of patients increasing signals for each antibody rank in the permtation analysis, where white indicates nonexisting and dark-blue the most common. The average value for each antibody rank is plotted as a yellow line. From this analysis it is evident that the true combination of the first and second samples gives a result that clearly diverges from the yellow line describing the FPR, not only in the tails of the plot but even more so in the middle, indicating a general up-regulation. Plots show the velocity for Lewis X (*B*), IL-16 (*C*), and CD40 (*D*) during the first 3–6 mo after surgery, with average CV measurements of 3.5%, 3.1%, and 3.5%, respectively. The signal velocity for these analytes was clearly positive for patients eventually developing metastatic breast cancer.

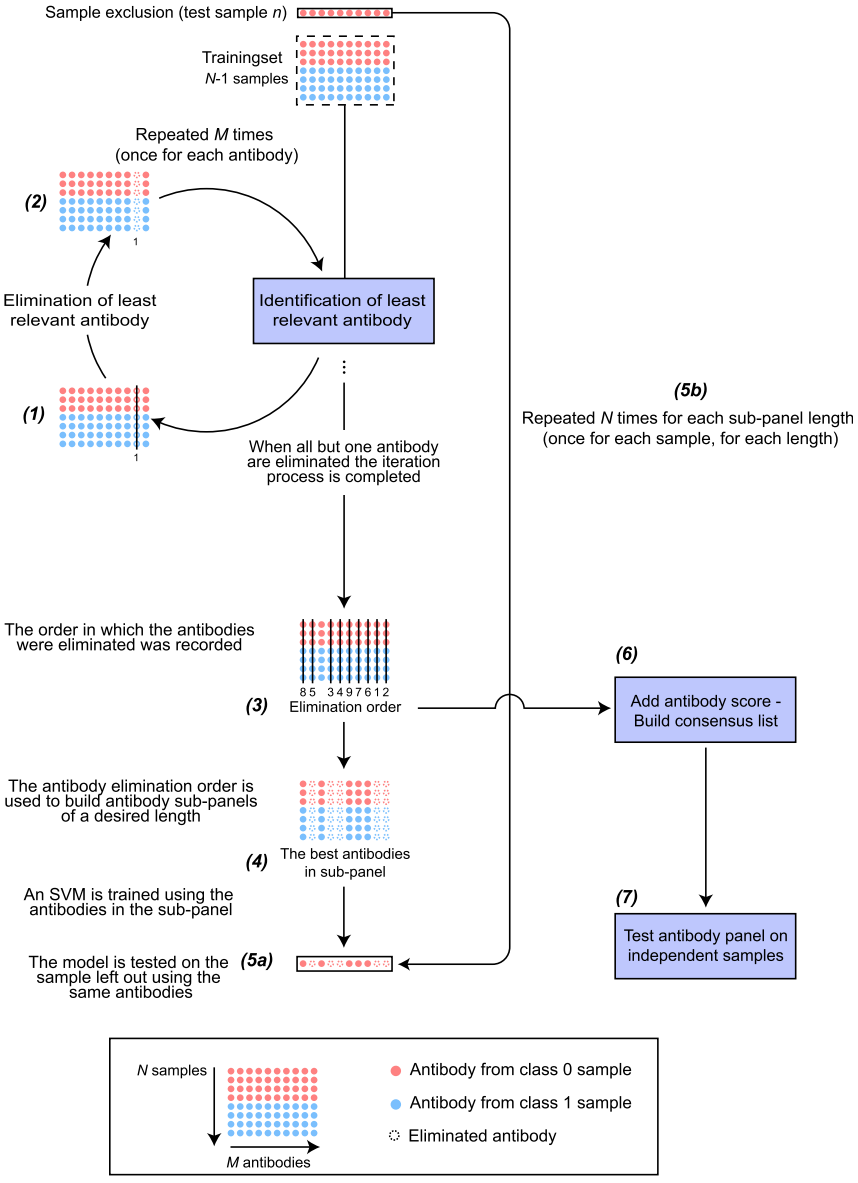

Fig. S3. To reduce the number of antibodies used for the classification, the leave-one-out procedure was combined with a backward-elimination process. This elimination is depicted schematically; numbers refer to steps described in *SI Materials and Methods*.

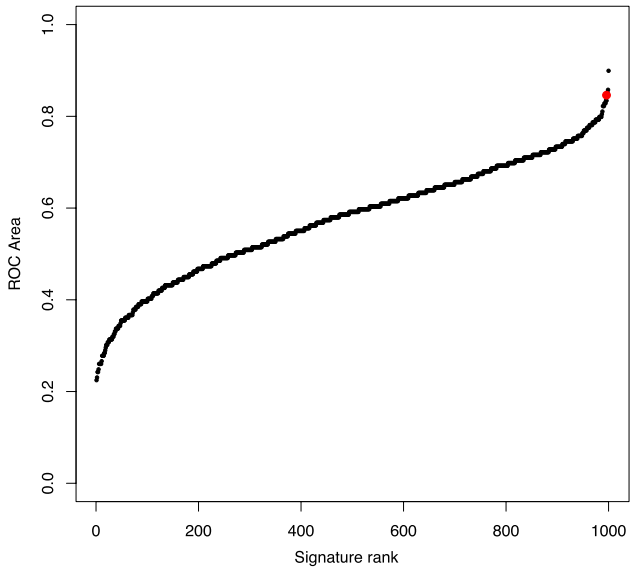

Fig. S4. The effectiveness of the backward-elimination strategy for signature optimization. The performance of the identified 21-biomarker signature was compared with that of 1,000 randomly selected signatures. All signatures were trained in the discovery cohort and subsequently tested and evaluated using the patients in the prevalidation cohort. The performance of the random signatures, represented by ROC AUC, was plotted (black) together with that of the original signature (red). The candidate signature outperformed 99.6% of the random signatures.

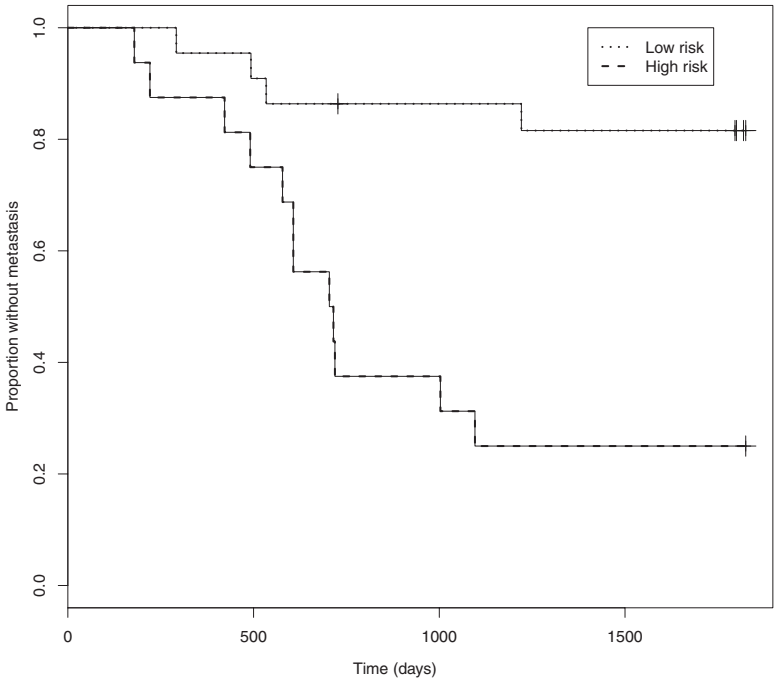

Fig. S5. A Kaplan–Meier plot demonstrating how the velocity-driven biomarker signature classified the patients into high- and low-risk groups.