SomaLogic

SOMAscan™

Technical White Paper

SOMAscan™ Proteomic Assay

Technical White Paper

Introduction

The SOMAscan™ assay is a highly multiplexed, sensitive, quantitative and reproducible proteomic tool for discovering previously undetected biomarkers for drug discovery, pre- clinical and clinical drug development, and clinical diagnostics, across a wide range of important diseases and conditions. The SOMAscan assay measures 1129 protein analytes in only 150 µL of serum, plasma or cerebrospinal fluid, or equally small amounts of a variety of other biological matrices. The assay offers exceptional dynamic range, quantifying proteins that span over 8 logs in abundance (from femtomolar to micromolar), with low limits of detection (38 fM median LOD in buffer) and excellent reproducibility (5.1 median %CV).

The SOMAscan proteomic assay is enabled by a new generation of protein-capture SOMAmer® (Slow Off-rate Modified Aptamer) reagents. SOMAmer reagents are constructed with chemically modified nucleotides that greatly expand the physicochemical diversity of the large randomized nucleic acid libraries from which the SOMAmer reagents are selected. The SOMAscan assay measures native proteins in complex matrices by transforming each individual protein concentration into a corresponding SOMAmer reagent concentration, which is then quantified by standard DNA techniques such as microarrays or qPCR. The assay takes advantage of SOMAmer reagents' dual nature as both protein affinity-binding reagents with defined three-dimensional structures, and unique nucleotide sequences recognizable by specific DNA hybridization probes. The assay is performed under Good Laboratory Practice (GLP)-like quality systems, and can be run under GLP when necessary. To date, the SOMAscan assay has been applied successfully to biomarker discovery and validation in many pharmaceutical research and development projects, diagnostics discovery and development projects, and academic research projects.

SOMAmer reagents are discovered using robust SELEX technology with proprietary chemical modifications

SOMAmer reagents are single stranded DNA-based protein affinity reagents that benefit from aptamer technology developed over the past 20 years (Ellington and Szostak, 1990) (Tuerk and Gold, 1990). The more recent proprietary innovation incorporates chemically modified nucleotides (Fig. 1) that mimic amino acid side chains, expanding the chemical diversity of standard aptamers and enhancing the specificity and affinity of protein-nucleic acid interactions (Gold *et al.*, 2010). These modified nucleotides are incorporated into nucleic acid libraries used for the iterative selection and amplification process called SELEX (Systematic Evolution of

Ligands by EXponential Enrichment) from which SOMAmer reagents are selected (Vaught *et al.*, 2010) (Eaton, 1997) (Davies *et al.*, 2012). Repeatedly, by using this novel, proprietary ASELEX process, SomaLogic has generated SOMAmer reagents to proteins that had been resistant to selection with unmodified nucleic acids (ACTG traditional aptamers) (Gold *et al.*, 2010). A key advantage of this artificial selection process is that conditions can be tailored to select for the desirable properties of specificity and slow off-rate as well as to mimic the assay conditions under which the reagents will be used.

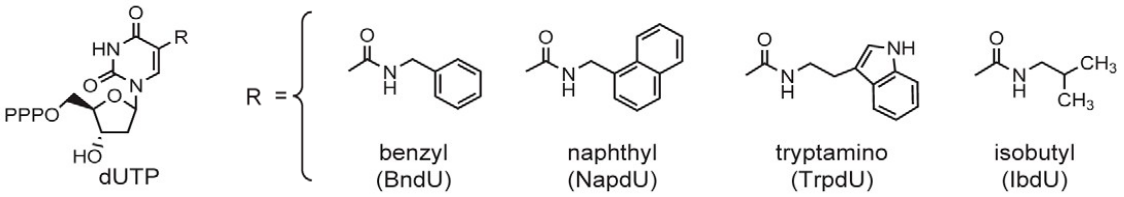
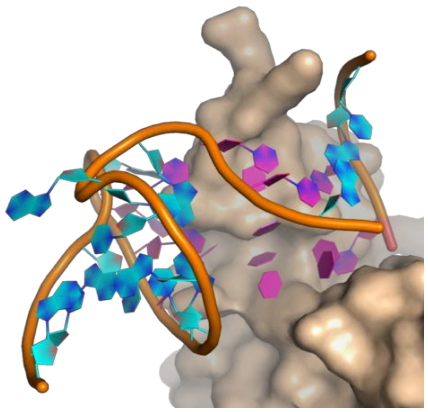


Figure 1. Modified nucleotides. Nucleotide triphosphate analogs modified at the 5-position (R) of uridine (dUTP); 5-benzylamino- carbonyl-dU (BndU); 5-naphthyl-methylaminocarbonyl- dU (NapdU); 5-tryptaminocarbonyl-dU (TrpdU); and 5-isobutylaminocarbonyl-dU (iBudU).

SOMAmer reagents are selected against proteins in their native folded conformations and are therefore generally found to require an intact, tertiary protein structure for binding. As such, unfolded and denatured, and therefore presumably inactive proteins, are not detected by SOMAmer reagents. Three co-crystal structures have been solved for SOMAmer reagents bound to their cognate protein target by X-ray crystallography and indicate that the modified nucleotides contribute extensively to intra-molecular folding of the SOMAmer reagent and to inter-molecular binding to cognate targets (Davies *et al.*, 2012). In Figure 2, a model derived from the X-ray crystal structure of PDGF-BB SOMAmer reagent binding to PDGF-BB demonstrates that the interactions between the SOMAmer reagent and its cognate protein are mainly mediated via the modified nucleotides. These structures show very specific interactions between SOMAmer reagent and target with binding site dimensions of 1100-1200 A^2 similar to antibody-antigen interactions (Ramaraj *et al.*, 2012).

Figure 2. X-ray crystal structure of a SOMAmer reagent to PDGF-BB binding to PDGF- BB. The DNA backbone of the SOMAmer reagent is shown in gold, the blue and teal indicate natural bases and pink the modified benzyl appendage on uracil binding to PDGF-BB (brown).



The dissociation kinetics of a large subset of SOMAmer reagents binding to their respective targets have been determined using a solution-phase radiolabeled binding assay (Gold *et al.*, 2010), and a subset confirmed using the Biacore Flexchip surface plasmon resonance biosensor. Association and dissociation rate constants and calculated equilibrium dissociation constant (K_D) for 10 SOMAmer-protein pairs measured on the Biacore Flexchip are summarized in Table 1. Biosensor results confirm slow dissociation off-rates, ranging from 10^{-4} to 10^{-5} s^{-1} that correlate well with dissociation rate constants measured by solution-phase filter binding assays. Lead SOMAmers are required to have a K_D of 1nM or better unless they are against highly abundant proteins. In total, SOMAmer reagents are analogous to high-quality antibodies that recognize intact tertiary protein structures, with the notable exception that they are made out of nucleic acids, leading to several advantages over antibodies such as tailored *in vitro* selection conditions, chemical synthesis, storage stability, and detection using sensitive and advanced DNA detection methods.

Table 1. Kinetic measurements for SOMAmer binding to cognate protein by Biacore Flexchip and radiolabel filter binding assay.

Protein Target	Biacore Flexchip[1]			Filter Assay[2]
	k_a,$M^{-1}s^{-1}$	k_d,s^{-1}	K_D, nM	K_D, nM
4-1BB ligand	5.5×10^5	1.6×10^{-4}	0.300	1.00
TNF sR-I	4.9×10^5	4.3×10^{-5}	0.090	0.07
Rab GDP dissociation inhibitor beta	4.7×10^6	1.7×10^{-5}	0.004	0.01
Thrombin	1.4×10^6	2.2×10^{-4}	0.200	0.75
VEGF	5.4×10^5	4.9×10^{-5}	0.090	0.09
IgE	4.0×10^6	5.0×10^{-4}	0.100	2.50
sL-Selectin	3.1×10^5	5.9×10^{-4}	1.900	0.30
4-1BB	1.7×10^5	8.1×10^{-5}	0.500	0.90
Cystatin C	5.0×10^4	1.5×10^{-4}	3.000	2.70
Transferrin	1.2×10^4	6.5×10^{-5}	5.400	18.00

[1] SOMAmer-target association and dissociation kinetics were measured using SOMAmer reagents immobilized onto neutravidin-coated biosensor chips.

[2] Equilibrium binding constants were determined using radiolabeled SOMAmer reagents equilibrated with increasing amounts of target protein. Bound and free SOMAmer reagents were separated by filtration and captured SOMAmer reagents quantified by phosphorimaging (Gold *et al.*, 2010).

SOMAmer affinity reagents to over 1100 proteins

SOMAmer reagents have been created for over 1100 protein targets that cover a diverse set of molecular functions, including several known disease and physiology associations (Fig. 3A). Targets to date extensively cover major gene families including receptors, kinases, growth factors

and hormones, and span a diverse collection of secreted, intracellular and extracellular proteins or domains. The distributions of current SOMAmer reagents are shown in Figures 3B and 3C.

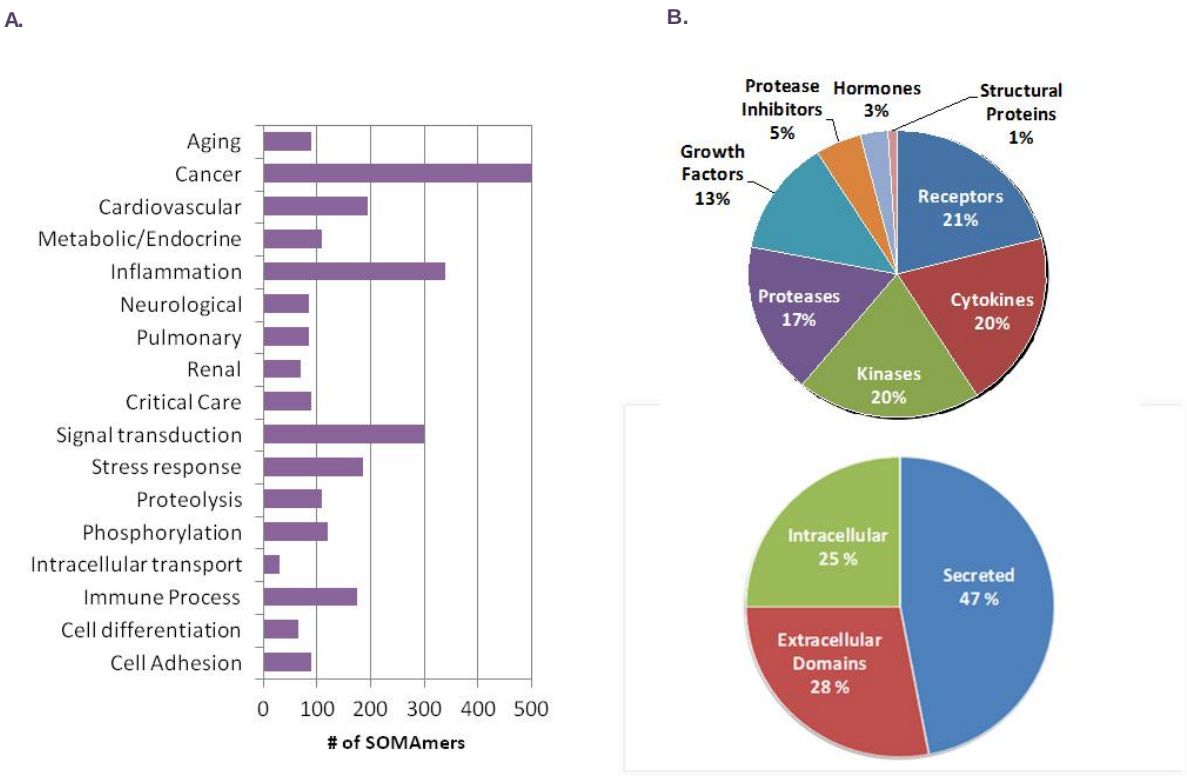


Figure 3. Characterization of SOMAmer reagents in the SOMAscan Assay.
(A) The SOMAmer reagents comprising the SOMAscan bind proteins in numerous disease areas and physiological processes. (B) SOMAmer reagents have been generated to various classes of proteins such as kinases, receptors and even hormones. (C) SOMAmer reagents target secreted, intracellular and extracellular domains of proteins.

Lead SOMAmer reagents are chemically synthesized, stable and rigorously analyzed

After identification using SELEX technology, the SOMAmer reagents are chemically synthesized, purified and analyzed by methods including ultra-high performance liquid chromatography (UPLC), capillary gel electrophoresis (CGE) and mass spectrometry (MS). Extensive functional analysis ensures consistent high performance of the SOMAmer reagents as quantitative affinity reagents. Nearly half of the SOMAmer reagents used in the current SOMAscan assay have been evaluated for cross reactivity to related proteins with 50% similarity or better. Two examples demonstrating SOMAmer specificity are in Appendix C and several more examples and details can be found in Gold *et al.*, 2010.

The Assay Steps

The SOMAscan assay quantitatively transforms the proteins present in a biological sample into a specific SOMAmer-based DNA signal (Fig. 4). A SOMAmer-protein binding step is followed by a series of partitioning and wash steps that converts relative protein concentrations into measurable nucleic acid signals that are quantified using DNA detection technology, which for the SOMAscan assay with 1129 SOMAmer reagents is by hybridization to custom DNA microarrays. Assays with smaller numbers of SOMAmer reagents (i.e., 1 – 100, called "SOMApanels"), have been quantified by either qPCR or Luminex beads using sequences complementary to the SOMAmer reagent sequences. Assay details are provided in Appendix A and in Gold *et al.*, 2010. Whichever detection method is used, the readout in relative fluorescent units (RFU) is directly proportional to the amount of target protein in the initial sample, as informed by a standard curve generated for each protein-SOMAmer pair.

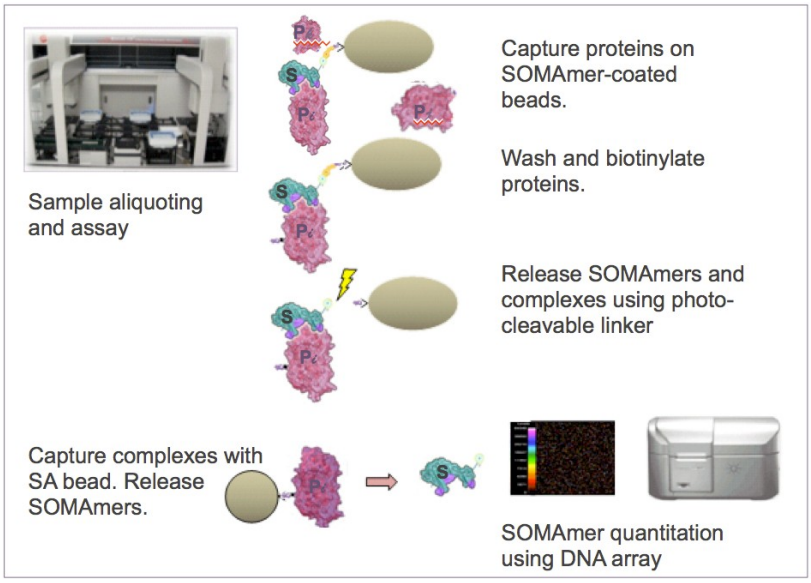


Figure 4. General steps of the SOMAscan assay. Samples are diluted and added to respective SOMAmer mixes using a BiomekFxP (Keeney *et al.*, 2009). After the proteins and SOMAmers equilibrate the solution is placed back on the deck of the BiomekFxP for the assay steps as described in the text (SA = streptavidin). The SOMAmer reagents are eluted and quantified using DNA microarrays.

Achieving the 10^8 dynamic range: SOMAmer reagent mixes

Each SOMAmer reagent was characterized by multiplexed protein standard curves in buffer with the proteins used in SELEX. A summary of the metrics from this experiment (LoD, LLoQ, ULoQ) are shown in Table 2. The median detection range across all SOMAmer reagents is 4.2 logs. The large dynamic range of SOMAscan results from the detection range of each SOMAmer reagent in combination with three serial dilutions of the sample. The least concentrated dilution is designed to detect the most abundant proteins (~µM concentration in 100% sample), and the most concentrated solution is designed to detect the east abundant proteins (fM to pM concentration in 100% sample).

Table 2. Summary of the total (intra- plus inter-run) %CV in plasma and upper limits of quantification (ULOQ), lower limits of quantification (LLOQ), limits of detection (LOD) and quantification range (difference between ULOQ and LLOQ) for SOMAmer reagents in buffer in the 1129-plex SOMAscan assay. *(Note: Values for the SOMAmer reagents were determined as described in Appendix B.)*

Percentile of SOMAmers	%CV (total in plasma)	LOD (M)	LOD (pg/mL)	LLOQ (M)	ULOQ (M)	Quantification Range (log)
5%	≤ 2.9%	≤ 3.7×10^{-15}	≤ 0.1	≤ 1.1×10^{-14}	≥ 7.2×10^{-9}	≥ 5.3
25%	≤ 3.9%	≤ 1.2×10^{-14}	≤ 0.4	≤ 3.2×10^{-14}	≥ 2.3×10^{-9}	≥ 4.6
50%	≤ 5.1%	≤ 3.8×10^{-14}	≤ 1.6	≤ 9.8×10^{-14}	≥ 9.9×10^{-10}	≥ 4.2
75%	≤ 7.3%	≤ 2.4×10^{-13}	≤ 9.5	≤ 4.4×10^{-13}	≥ 4.5×10^{-10}	≥ 3.5
95%	≤ 12.6%	≤ 3.3×10^{-12}	≤ 194.0	≤ 3.0×10^{-12}	≥ 1.4×10^{-10}	≥ 2.2

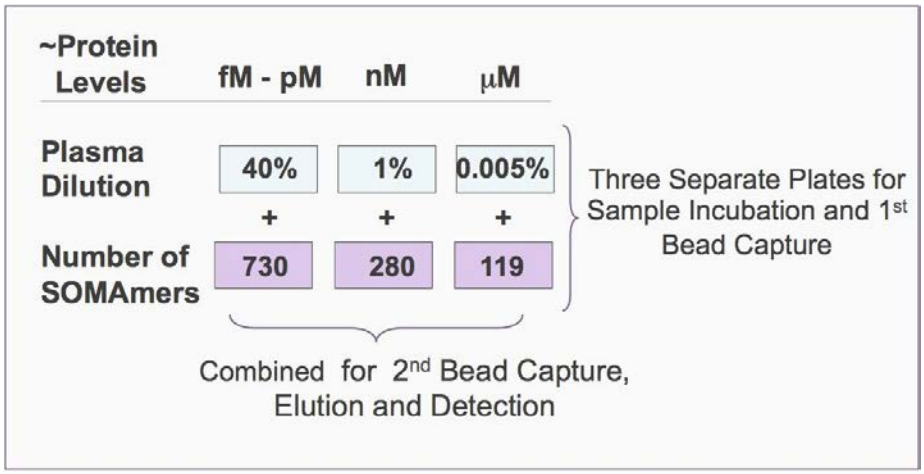


Figure 5. SOMAscan assay dynamic range. Custom SOMAmer reagent mixes are prepared for plasma, serum, or other matrices to achieve optimal detection in mixtures with a large range of concentrations. Shown here is the dilution distribution of SOMAmer reagents for SOMAscan with plasma.

SOMAscan Assay Characterization

The SOMAscan assay has been characterized for reproducibility, limits of quantification (LOQ) and limit of detection (LOD), a summary of which is in Table 2, and experimental details in Appendix A. The SOMAscan assay has excellent reproducibility, with half of the SOMAmers having a median %CV of 5.1 or lower and only 5% of the SOMAmer reagents having a median %CV of 12.6 or higher in plasma. The sensitivity is excellent and equivalent to more sensitive antibodies, with half of the SOMAmer reagents binding to their target with a LOD in buffer of 38 fM or less and 5% of the SOMAmer reagents having LOD less than 4 fM. The upper and lower limits of quantification, determined using precision profiles, as detailed in Appendix B, yield large quantification ranges for each individual SOMAmer reagent. The median quantification range across the 1129 SOMAmer reagents in the assay is $10^{4.2}$ M (Table 2). All values were determined in the multiplex assay, profiling 1129 analytes simultaneously.

Sample types

The SOMAscan platform is amenable to proteomic profiling of numerous matrices, allowing for discovery of biomarkers that may translate across all phases of drug discovery and basic research, and target organ compartments that may store the most disease-relevant biomarker information. Currently, human plasma (EDTA, citrate, heparin or ACD), serum, and cerebrospinal fluid (CSF) are the most well-characterized matrices, and only 150 µL is required to run them in the SOMAscan assay. Other biological matrices such as cell culture supernatant, cell and tissue lysates, synovial fluid, and bronchoalveolar and nasal lavage have been used successfully (Table 3, and see "Applications" Figs. 6 - 10). Additionally, the SOMAscan assay has been used to detect differntial expression in drug-treated preclinical xenograft models. SOMAmer reagents generated to pure human proteins have varying degrees of cross reactivity to non- human orthologs and therefore the assays can be used to identify differntial expression of some analytes in non-human samples, including non-human primates and rodents. We are developing optimized protocols for many of the "non-standard" matrices, and are open to discussing additional sample types with collaborators. Specific protocols for various sample preparations are available on request.

Table 3. SomaScan sample matrices

Tier	Tier Name	Matrices	
		Species	Sample Types
1	**"Qualified" -** For intended purpose in biomarker discovery. Performance in SOMAscan extensively evaluated with assay parameters developed for matrix specific optimization.	Human	EDTA-Plasma, Serum, CSF, Cell lysate[A], Heparin plasma, Conditioned media
2	**"Developed" -** Performance in SOMAscan using existing parameters has been characterized.	Human	Lymphocytes, ACD/Citrate plasma
		Dog	EDTA Plasma
3	**"Research" -** Performance in SOMAscan using existing parameters has been minimally characterized or not at all. These matrices are run as standalone experiments requiring internal control.	Human	Nasal lavage, Wound fluid[B], Urine[B], BAL[A], Sputum (Hargreave with/or without DTT)[B], Synovial fluid[B], Exosomes[A], Tears in buffer, Tissue homogenate (tumor, muscle, mucosa)[A]
		Non-Human Primate	Plasma, Serum, CSF, Aqueous Humor
		Rat	EDTA Plasma[B], Serum[B], CSF[B], Muscle[A], BAL[A], Citrate Plasma
		Mouse	EDTA Plasma[B], Serum[B], Cell lysate[A], Conditioned media , BAL[A], Tissue homogenate (xenograft, muscle) [B]
		Dog	Serum, Synovial fluid[B], Urine[B]
		Cat	Plasma, Serum
4	**"Not tested" -** Matrix not tested on platform.	All	Everything not mentioned

[A] This sample type requires a total protein quantification prior to assay

[B] Matrices that require additional pre-assay sample preparation

SomaLogic Quality Systems

The SOMAscan assay is run under the SomaLogic Quality System (QS) and SomaLogic is prepared to run samples under GLP when required. The assay is performed in a facility that contains both access and environmental control. Equipment within the facility is maintained, calibrated, and operated in compliance with controlling Standard Operating Procedures (SOPs). Equipment and associated software is validated for its intended use in support of the SOMAscan assay. Method validation has been completed for processes that could impact the performance of the SOMAscan assay. SOPs cover the incoming receipt, inspection, and release of raw materials to assure that the materials used in the production of assay reagents or directly in the assay maintain the performance requirements established during the development of the SOMAscan assay. Reagent expiration dating, including SOMAmer reagents, calibrators, controls, and Master Mix retest dating, is set with all available background information and extended following review of data collected in the Stability Program.

Applications of the SOMAscan assay

Comparative interrogation of the human proteome in healthy and diseased tissues can offer deep insights into the biology of disease and to the discovery of new drug targets.
Elucidation of protein response to drug in pre-clinical samples can provide valuable insight into mechanism of action (MOA) and ultimately biomarkers that may translate into the clinic. In clinical samples proteomic profiling can lead to novel, highly specific biomarkers for diagnostics and drug treatment, enabling the realization of personalized medicine. The high throughput SOMAscan assay with low sample volume requirements is designed to address proteomic profiling from basic research to pharmaceutical development.

Pre-clinical Applications

The SOMAscan proteomic assay can be a powerful tool in pre-clinical research from uncovering new drug targets, elucidating MOA, target validation and identifying off-target effects in cell and animal model systems, examples of which follow.

Translational medicine example:
NSCLC cell lines and xenograft mouse model drug treatment

Erlotinib (Tarceva®) is a well-studied cancer drug approved for second/third line treatment of non-small cell lung cancer (NSCLC) (Shepherd *et al.*, 2005), that was used as a model to illustrate the utility of the multi-analyte SOMAscan assay in finding biomarkers that translate from pre-clinical cell and animal models to the human. In this study two NSCLC cell lines, HCC827, known to be erlotinib sensitive with IC_{50} < 1µM, and H1299, an erlotinib-resistant line with an IC_{50} of >5 µM, were treated in culture with escalating doses of erlotinib for 24 hours. The cells were lysed and profiled in the SOMAscan assay. Simultaneously, HCC827 cells were implanted in NOD/SCID mice to generate an NSCLC tumor xenograft. The animals were then dosed with 0, 50 or 100 mg/kg of erlotinib for 72 hours and tumor samples taken at 8, 24 and 72 hours post

treatment, lysed and profiled in the SOMAscan assay (Figs. 6, 7 and 8). Ninety-three analytes changed upon erlotinib treatment in the sensitive cell line HCC827 (p-value < 10^{-4} – Fig. 6A) whereas only three analytes changed in response to 10µM erlotinib treatment in the resistant cell line H1299 (p-value < 10^{-4} – Fig. 6B). Fifty-one analytes changed in response to erlotinib in the tumor xenografts (p-value < 10^{-4} – Fig. 6C), 21 of which were in common with the HCC827 cellular response to erlotinib (Fig. 7). Two analytes, $p27^{KIP1}$ (a cyclin-dependent kinase inhibitor), and CDK2/cyclin A changed in the HCC827 cell culture and tumor in a manner consistent with the previously reported observation (Ling *et al.*, 2007) that erlotinib induces G1/S phase arrest via an increase in $p27^{KIP1}$ transcript and a reduction in CDK2 activity (Fig. 8). A study is underway in humans receiving Tarceva® as treatment to compare biomarkers in these pre-clinical models to response in humans. These pre-clinical data demonstrate the utility of the SOMAscan assay to find proteomic signatures of drug pharmacology leading to mechanism of action, pathway redundancy, and uncover possible undesired off-target, toxic effects as well as potentially identifying biomarkers that translate into the clinic, in a set of efficient and well-controlled experiments.

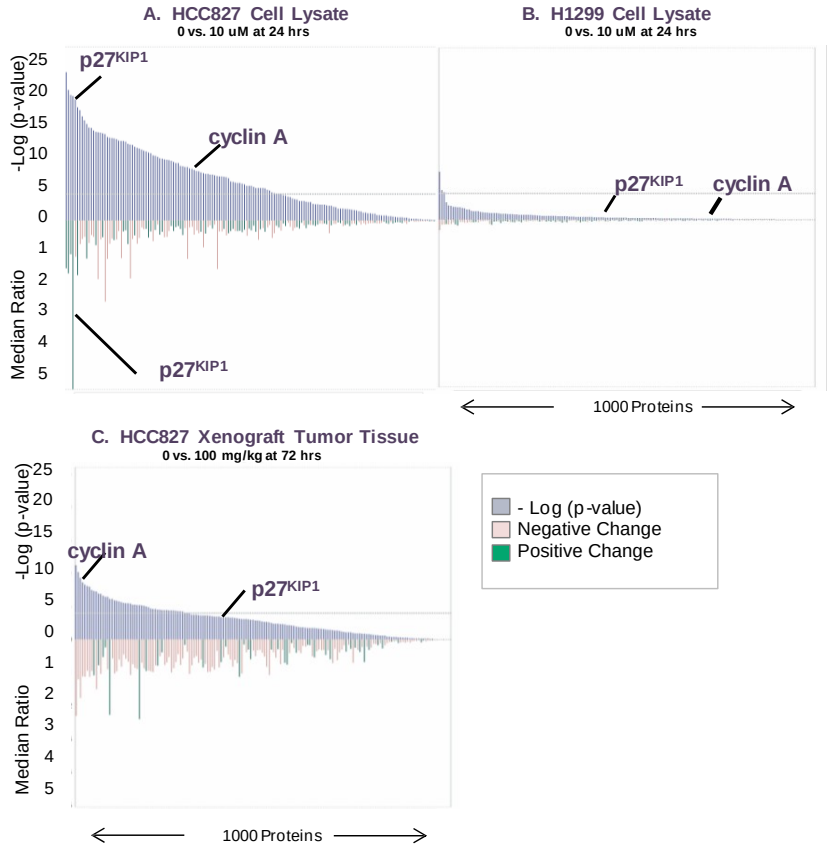


Figure 6. Profiling response to erlotinib in NSCLC cell lines HCC827, H1299 and tumor tissue derived from a HCC827-derived xenograft using SOMAscan assay. Ratio of median RFU (median ratio, bottom *y-axis*) for each analyte in the SOMAscan assay (*x-axis*) from cells lysed after treatment with 0 relative to 10 µM erlotinib for 24 hrs for erlotinib sensitive cells HCC827 (A) or erlotinib resistant cells H1299 (B).

A decrease in signal upon 10 µM erlotinib treatment is indicated by pink bars, an increase by green bars. The negative log of the p-value of the median ratio is indicated by gray bars on the upper *y-axis*. Location of analytes cyclin A and $p27^{KIP1}$ are indicated.

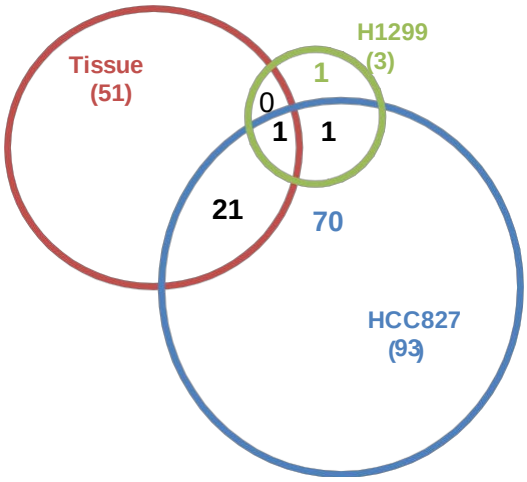


Figure 7. Venn diagram of analytes whose levels changed in response to erlotinib treatment in H1299 or HCC827 cells and tumor tissue derived from HCC827 xenograft animals (with p-value <10^{-4}).

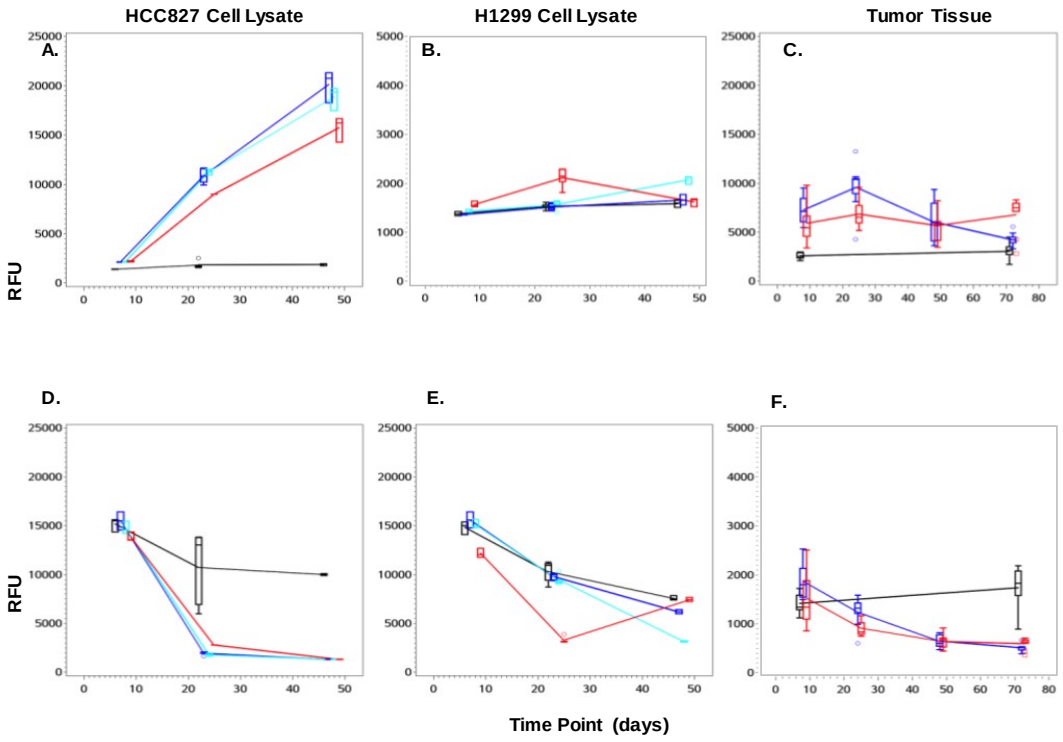


Figure 8. $p27^{KIP1}$ and CDK2/cyclin A response to erlotinib treatment in HCC827, H1299 cells and tumor tissue from HCC827-derived xenografts. $p27^{KIP1}$ (A, B, C) or CDK2/cyclin A (D, E, F) levels post treatment with 0 (black), 0.1 (blue), 1 (aqua) or 10µM (red) erlotinib for 24hrs in HCC827 (A, D) or H1299 (B, E) cells or tumor tissue from HCC827 xenograft animals (C, F) after treatment with 0 (black), 50 mg/kg (blue) or 100mg/kg (red) erlotinib.

New target discovery: Tumor tissue lysates

In addition to cell culture lysates, lysates of tissue biopsies of fresh/frozen tissues have been run in the SOMAscan assay, uncovering novel potential drug targets for NSCLC (Mehan *et al.*, 2012). Figure 9 shows the outcome of homogenized lung tissue samples from surgical resections obtained from eight NSCLC patients diagnosed with pathology- confirmed NSCLC stages IA - IIIB run through an earlier version of SOMAscan that contained SOMAmer reagents to 813 analytes. Three samples were obtained from each resection: tumor tissue sample, adjacent non-tumor tissue (within 1 cm of the tumor) and distant uninvolved lung tissue (furthest edge of the resection from the tumor). Overall, the signals generated by most analytes were similar in adjacent and distant tissue (Fig. 9A); only one analyte exhibited more than a two-fold difference between the two control samples. In contrast, comparison of tumor tissues with non-tumor tissue (adjacent or distant) identified 11 proteins (1.3%) with greater than four-fold differences and 53 proteins (6.1%) with greater than two-fold differences (Figs. 9B and 9C).

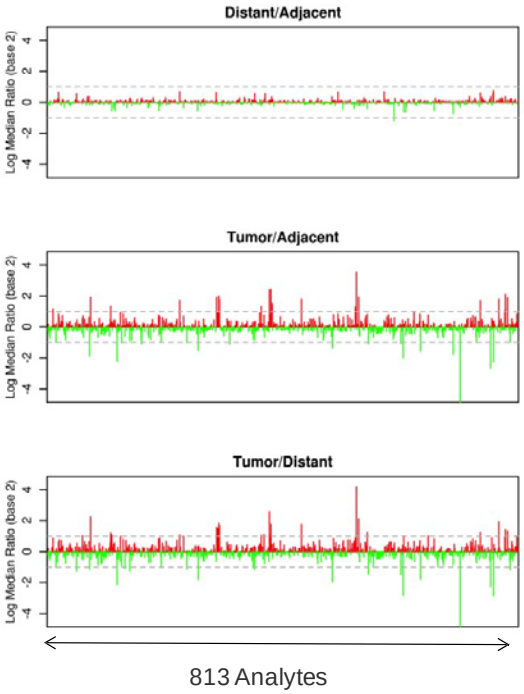


Figure 9. Application of SOMAscan to lung tumor tissue. Frozen lung tumor tissue, tissue adjacent to the tumor and tissue distal to the tumor were profiled across 813 analytes in SOMAscan v1. Differential expression of the 813 analytes are plotted as a $\log_2$ ratio of the median value across the 8 samples of distal tissue relative to adjacent tissue (A), tumor tissue relative to adjacent (B) and tumor relative to distal (C). Red shows analytes that have increased levels in the numerator relative to the denominator and green, analytes that have decreased levels.

The remaining proteins (93.9%) showed relatively small differences between tumor and non-tumor tissue. Some proteins were substantially suppressed while others were elevated in tumor tissues compared to adjacent or distant tissues. About one-third (13/36) of these potential NSCLC tissue biomarkers identified are novel, to the best of our knowledge, and demonstrate the application of SOMAscan to new target discovery. The remaining two-thirds (23/36) have been previously reported as differentially expressed proteins or genes in NSCLC tumor tissue.

SOMAmer reagents as histochemical probes

SOMAmer reagents have been shown to be fast-binding and very specific histochemical staining reagents (Mehan *et al.*, 2012) (Gupta *et al.*, 2011). The SOMAmer reagent against thrombospondin 2, one analyte found to be highly expressed in the tumor relative to adjacent tissue, was used for histological staining of lung tumor tissue (Fig. 10). A strikingly increased level of the protein is seen in the tumor stroma relative to adjacent and distal non-malignant lung tissues. Other SOMAmer reagents showed differential expression histochemical staining consistent with SOMAscan results (Mehan *et al.*, 2012) demonstrating the utility of SOMAmer reagents as general affinity reagents due to the high specificity and slow off-rates.

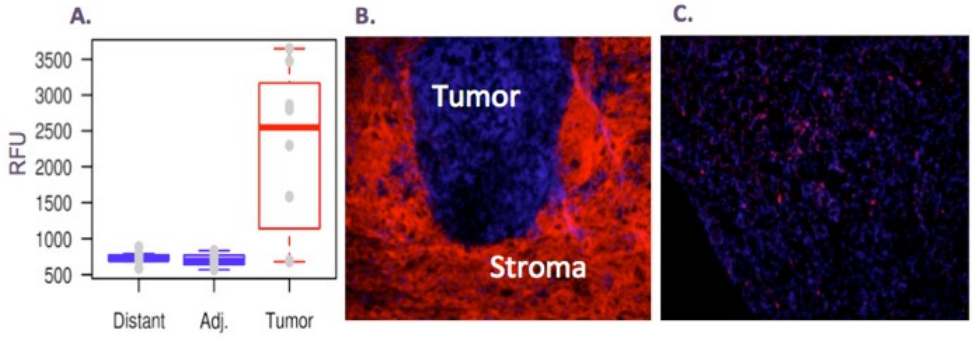


Figure 10. SOMAmer reagents as histochemical probes.
Thrombospondin 2 was discovered to be highly expressed in NSCLC lung tumor frozen tissue sections relative to distal non-malignant tissue by using SOMAscan. (**A**) Data points for eight replicates of tumor, adjacent and distal tissues run through the SOMAscan assay are plotted in boxplots showing the median (line) and 25% and 75% range of the data (box outline). (**B**) and (**C**) Histological staining of frozen tissue sections of tumor and distal non-malignant tissue using the Cy3-labeled SOMAmer reagent to Thrombospondin 2. (**B**) Thrombospondin-2 (red) staining the fibrocollagenous matrix (stroma) surrounding a tumor nest and (**C**) corresponding distal non-malignant lung specimen stained with Thrombospondin-2 SOMAmer reagent (red).

Clinical Diagnostic Discovery

SomaLogic has run thousands of clinical samples through the SOMAscan assay to discover diagnostic signatures to predict disease earlier and more specifically than currently available diagnostic tests. At least ten novel classifiers have passed blinded verification across a variety of diseases ranging from NSCLC to predicting cardiovascular risk. Dozens of studies on drug-treated patient samples have been run in the SOMAscan assay all with novel biomarkers discovered. Here data from patient samples treated with an anti- angiogenic compound, and classifiers from patients with mesothelioma or cardiovascular disease are used as examples of clinical applications of the discovery process fueled by the powerful SOMAscan biomarker discovery tool.

Companion Diagnostic Discovery

The power of the SOMAscan proteomic assay in discovery of a potential companion diagnostic is illustrated in Figure 11. Samples from individuals before and after treatment with an angiogenesis inhibitor were run through an early version of the SOMAscan assay that contained 386 SOMAmer reagents. Five key proteins, perhaps related to pharmacodynamic activity, changed within 4 hours of dosing; two more proteins changed after 2 weeks. One protein related to mechanism of action was confirmed with an ELISA (data not shown). This previously unknown protein would not have been discovered so readily without a highly multiplexed, high-throughput biomarker discovery tool such as the SOMAscan assay.

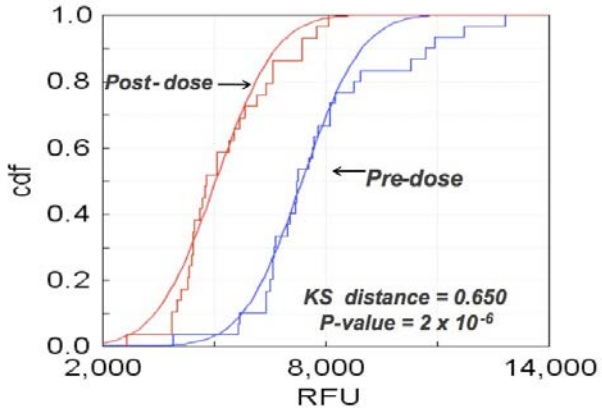


Figure 11. Cumulative distribution function (CDF) of one analyte that showed significant ($p < 2 \times 10^{-6}$) Kolmogorov-Smirnov (KS) distances between pre-dose and two-week dosing.

Clinical Diagnostic Discovery- Mesothelioma

The SOMAscan assay was used to discover biomarkers for the detection of mesothelioma in a high-risk asbestos-exposed population (Ostroff *et al.*, 2012). Mesothelioma has a latency period of up to 30 years after exposure, but once it develops it is an aggressive cancer with a poor prognosis. Early detection of mesothelioma by screening people with an extensive history of asbestos exposure increases the likelihood for effective treatment. Retrospective samples were obtained from 357 mesothelioma cases and asbestos- exposed controls. Thirteen biomarkers were discovered that gave a receiver operator curve (ROC) AUC of 0.99 ± 0.01 in training, 0.98 ± 0.04 in independent blinded verification and 0.95 ± 0.04 in blinded validation studies (Fig. 12), better than any other available diagnostic test, allowing for fewer false positive and more true positive results. This classifier accuracy was maintained in a second blinded validation set with a sensitivity/ specificity of 90%/89% and combined accuracy of 92%. One of these analytes was confirmed by ELISA (Fig. 13) where the SOMAmer reagent displayed better sensitivity, distinguishing the samples with low amounts of the analyte that were grouped together in the ELISA.

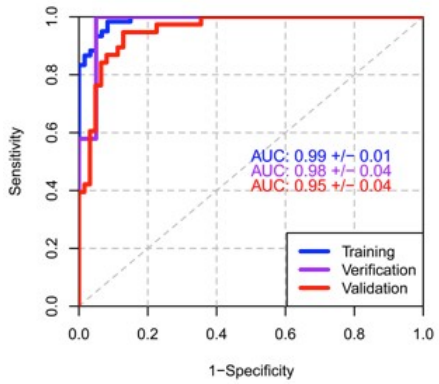


Figure 12. Receiver-operator curve (ROC) for mesothelioma classifier training blinded verification and validation. A classifier composed of 13 analytes distinguishing mesothelioma cases from asbestos-treated controls for training (blue), verification (purple), and validation (red) study ROC curves are plotted with corresponding AUC values and 95% confidence intervals.

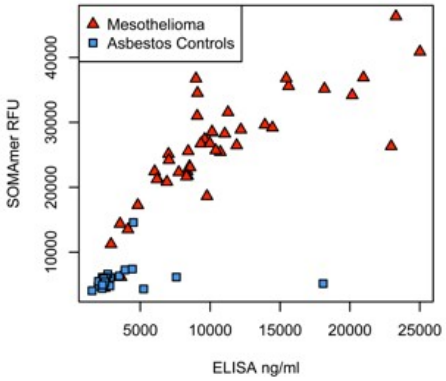


Figure 13. Correlation of ELISA quantification with SOMAmer RFU of an analyte in the mesothelioma classifier. Samples from the training cohort were analyzed for expression of ficolin-2, one of the top analytes in the classifier using a commercially available ELISA kit. Measurements for malignant pleural mesothelioma cases (red triangles) and asbestos-exposed controls (blue squares) are reported as RFU for ficolin-2 SOMAmer directly from the multiplexed SOMAscan results and ng/mL for ELISA measurements from a commercially available kit. Spearman correlation is 0.87.

Predicting cardiovascular risk as a model for clinical trial patient stratification

In collaboration with the University of California, San Francisco, SomaLogic used the SOMAscan assay to discover biomarkers to predict risk of a future cardiovascular (CV) event in 987 individuals with stable coronary heart disease. One plasma sample was taken from these individuals who were then tracked for a median of 6 years; time to hospitalization for myocardial infarction, stroke, transient ischemic attack (TIA), heart failure, or death from any cause was recorded as a primary endpoint (Ruo *et al.*, 2003). The 987 plasma samples were profiled in the SOMAscan assay in just 7 days. From this unbiased search, a ten protein classifier was discovered that yielded a hazard ratio of 10 between the highest and lowest quintiles for subsequent events over a 6 year period following blood draw (Fig. 14A) that reproduced in the blinded verification set (Fig. 14B). The proteins that comprise the classifier are novel for cardiovascular risk stratification yet are fundamentally mapped to renal function, inflammation, insulin resistance and CV hormones all with biologically plausible connections to cardiovascular disease.

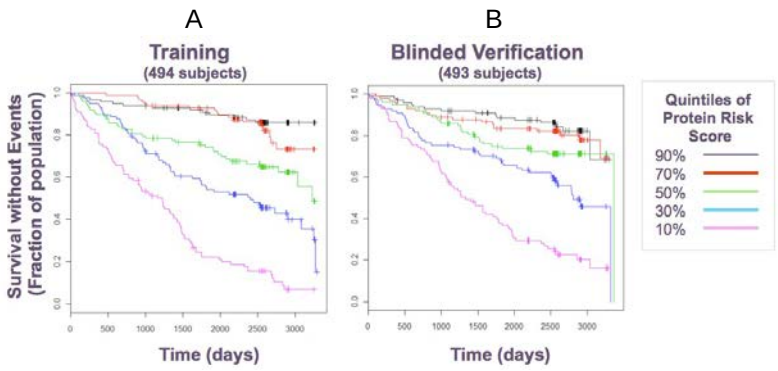


Figure 14. Cardiovascular risk classifier identified using SOMAscan assay on 987 high risk individuals. 494 subjects were used to identify a 10 protein classifier that stratified patients on risk of an event within 6 years (median) (A) that verified in a blinded test on the remaining 493 subjects (B).

This data set was used in a theoretical model to demonstrate the power of patient stratification that can be applied once a robust biomarker classifier is identified. In this model a hypothetical cardiovascular drug with a hazard ratio of 1.5 vs. placebo would show statistically significant difference between drug and placebo within 6 months of clinical trial initiation when 200 subjects with the highest protein-based risk score from the UCSF study were enrolled (Fig. 15B). When 200 individuals selected at random from the UCSF study were included in the trial, no difference between drug and placebo would be observed even after 5 years of treatment (Fig. 15A). This demonstrates the power of biomarkers, which can be readily discovered using the SOMAscan platform, to reduce the time and cost for clinical trials, resulting in more effective treatment sooner to the patients that need it most.

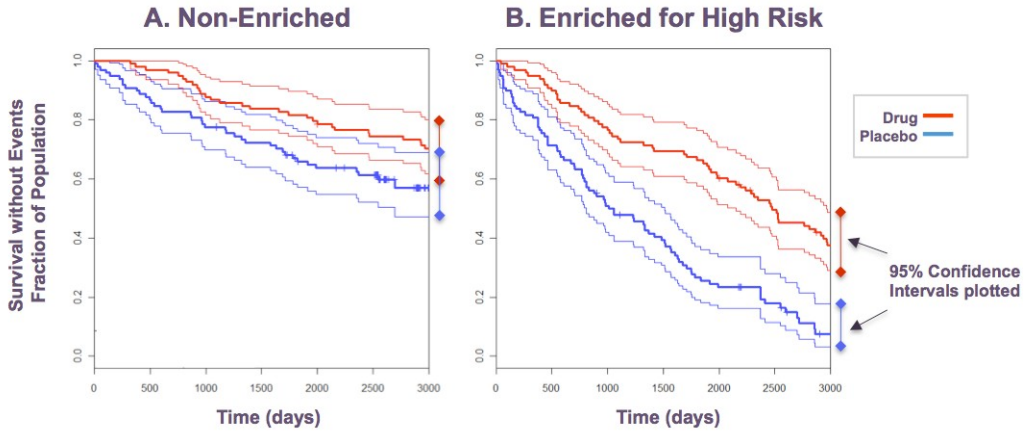


Figure 15. Theoretical model of clinical drug trial success when low risk patients are excluded. A model clinical trial of 200 patients with stable coronary disease but not stratified for risk of a future event showed no difference in effect of a drug after 8 years of treatment (A) whereas 200 patients pre-screened for high risk of a future event showed significant difference within 6 months of drug treatment (B).

Custom SOMApanels

SomaLogic can develop custom SOMApanels (Kraemer *et al.*, 2011) for specific research needs (e.g., particular pathways, classes of proteins). The streamlined SOMApanel format allows for a seamless transition from SOMAscan to a smaller custom multi-analyte assay. Quest Diagnostics has launched the first clinical SOMAmer-based diagnostic test with additional products in the pipeline.

Summary

The SOMAscan assay is a powerful, highly multiplexed platform for discovering novel biomarkers during drug discovery, pre-clinical and clinical drug development, and for the development of clinical diagnostics, across a wide range of clinically important diseases (Table 4 summarizes the SOMAscan assay metrics). SomaLogic's SOMAscan technology provides significant advantages in sample size, cost, time, multiplexing capability, dynamic range, and flexibility of readout over many alternate protein biomarker platforms. It has been used successfully for many sample types and matrices (including non-human primates, mouse, rat, dog, and cat) and is optimized for human analytes. It scales easily to progress from biomarker discovery to focused products without the need for new assay development. This technology can be readily employed as a valuable tool from basic research to drug discovery to diagnostic development.

Table 4. Summary of SOMAscan assay metrics unique human targets

Metric	Condition	Current version of SOMAscan assay
Sensitivity (buffer)	Median LLOQ	100 fM
	Median LOD	38 fM or 1.6 pg/mL
Dynamic Range (buffer)	Over all proteins in serum or plasma	10^8
	Median range per SOMAmer	4.2 logs
Precision	Median Total %CV	5.1%
Sample Volume	Human serum, plasma or CSF (per sample)	150 uL
Multiplex Size	Current number of proteins per sample	1129

References

Dewey T, *et al.* (1995). New Uridine Derivatives for Systematic Evolution of RNA Ligands by Exponential Enrichment. *J. Am. Chem. Soc.* 117:8474-5.

Davies DR, *et al.* (2012). Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. USA* 106:19971-6.

Eaton BE. (1997). The joys of in vitro selection: chemically dressing oligonucleotides to satiate protein targets. *Curr. Opin. Chem. Biol*. 1:10-6.

Ellington AD & Szostak JW. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818-22.

Gold L, *et al.* (2010). Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLOS ONE* 5(12): e15004.

Gupta S, *et al.* (2011). Rapid Histochemistry Using Slow Off-rate Modified Aptamers with Anionic Competition. *Appl. Immunohistochem. Mol. Morphol.* 19:273-8.

Keeney TR, *et al.* (2009). Automation of the SomaLogic Proteomics Assay: A Platform for Biomarker Discovery. *J. Lab. Automation* 14:360-6.

Kraemer S, *et al.* (2011). From SOMAmer-Based Biomarker Discovery to Diagnostic and Clinical Applications: A SOMAmer-Based, Streamlined Multiplex Proteomic Assay. *PLOS ONE* 6(10): e26332.

Ling YH, *et al.* (2007). Erlotinib, an Effective Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor, Induces p27KIP1 Up-Regulation and Nuclear Translocation in Association with Cell Growth Inhibition and G1/S Phase Arrest in Human Non-Small-Cell Lung Cancer Cell Lines. *Mol. Pharmacol.* 72: 248-58.

Mehan MR, *et al.* (2012). Protein Signature of Lung Cancer Tissues. *PLOS ONE 7* (4): e35157.

Ostroff RM, *et al.* (2012). Early Detection of Malignant Pleural Mesothelioma in Asbestos-Exposed Individuals with a Noninvasive Proteomics-Based Surveillance Tool. *PLOS ONE* 7(10): e46091.

Ramaraj T, *et al.* (2012). Antigen-antibody interface properties: Composition, residue interactions, and features of 53 non-redundant structures. *Biochim. Biphys. Acta.* 1824:530-2.

Ruo B, *et al.* (2003). Depressive Symptoms and Health-Related Quality of Life. *JAMA* 290:215-21.

Shepherd FA, *et al.* (2005). Erlotinib in previously treated non-small-cell lung cancer. *N. Engl. J. Med.* 353:123-32.

Tuerk C and Gold L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505-10.

Vaught JD, *et al.* (2010). Expanding the chemistry of DNA for in vitro selection. *J. Am. Chem. Soc.* 132:4141-51.

Appendix A

Details of the SOMAscan assay

The first step of the SOMAscan assay is the dilution of a biological sample of interest. The sample dilutions are incubated with the respective SOMAmer reagent mixes pre-immobilized onto streptavidin (SA)-coated beads. The beads are washed to remove all non-specifically associated proteins and other matrix constituents. Proteins that remain specifically bound to their cognate SOMAmer reagents are tagged using an NHS-biotin reagent. After the labeling reaction, the beads are exposed to an anionic competitor solution that prevents non-specific interactions from reforming after they are disrupted (Gold *et al*., 2010). Essentially pure cognate-SOMAmer complexes and unbound (free) SOMAmer reagents are released from the SA beads using ultraviolet light that cleaves a photo-cleavable linker. The photo-cleavage eluate, which contains all SOMAmer reagents (some bound to a biotin-labeled protein and some free), is separated from the beads and then incubated with a second streptavidin-coated bead that binds the biotin-labeled proteins and the biotin-labeled protein-SOMAmer complexes. The free SOMAmer reagents are then removed during subsequent washing steps. In the final elution step, protein-bound SOMAmer reagents are released from their cognate proteins using denaturing conditions. These SOMAmer reagents can then be quantified by standard DNA quantification techniques, which for the SOMAscan with 1129 SOMAmer reagents, is by hybridization to custom DNA microarrays. The Cyanine 3 signal from the SOMAmer reagent is detected on microarrays, phycoerythin for Luminex bead formats and SYBR Green is detected in qPCR.

Data analysis

Normalization procedures have been developed to assure data consistency. In the simplest form, normalization is performed using a set of hybridization control sequences introduced into the assay eluate prior to hybridization and measured independently for each sample array, which corrects for any systematic effects on the data introduced during the hybridization step.
A more robust normalization scheme uses all the SOMAmer reagent signals on a given array to allow for comparison of samples across a plate, within similar groups. It corrects for variation that may be introduced in the course of the SOMAscan assay, including natural variation in initial sample concentration that may occur. Each normalization method computes a single scale factor for each sample that is subsequently applied to the signal on all features within an array.
Calibration is performed to allow for sample measurements across runs. Every plate contains replicates of a calibrator sample that is chosen to match the matrix type of the samples in the study (e.g., serum, EDTA-plasma). The median value is calculated across all the calibrator samples within the study for each SOMAmer reagent. These median values are compared to a previously established reference calibration value that generates a calibration scale factor that is then applied to all measurements for that SOMAmer reagent within the set of samples in the study. Figure A1 demonstrates the effect of calibration for one SOMAmer reagent. Data are reported in relative fluorescent units (RFU) after normalization and calibration.

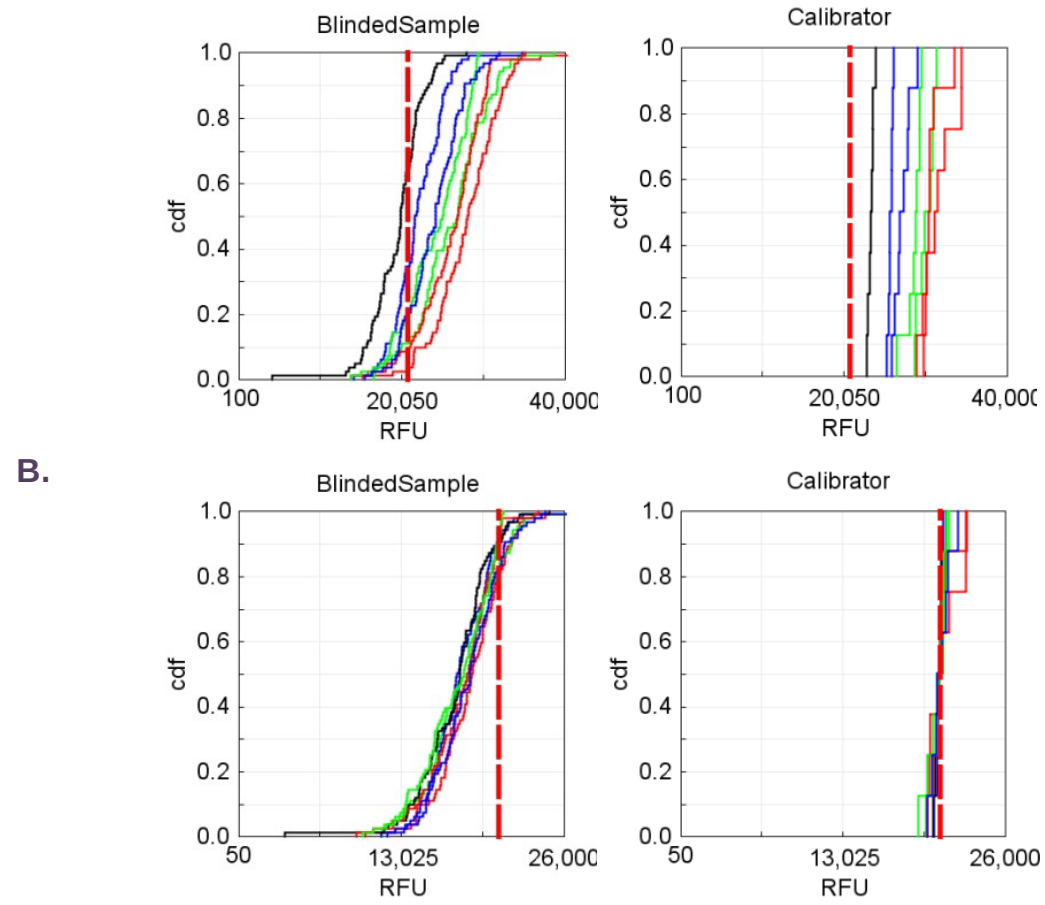


Figure A1. Illustration of using calibration to remove systematic bias between assay runs. (**A**) Cumulative distribution functions (cdf) for a single SOMAmer reagent for a set of samples (left plot) were generated for each of seven independent assay runs, color coded by run. The cdfs for the eight replicate calibrator sample measurements for that SOMAmer are displayed on the right for each run, color coded as on the left. The vertical red bar is the global calibrator reference obtained from a separate independent set of calibrator runs and is the target calibration RFU for this SOMAmer reagent. Note the correlation of shifts between the cdfs for the samples and the calibrator. (**B**) Cumulative distribution functions (cdf) for the set of samples and calibrators in (**A**) after calibration. Note the collapse of the sample distributions to essentially a single distribution after calibration.

Appendix B

Experiments to characterize the SOMAscan assay

Reproducibility

Reproducibility was assessed by running 12 different serum samples and 12 different plasma samples in triplicate, with seven replicates of the associated calibrator sample, two Quality Control samples, and three buffer control wells on three separate runs. The separate runs utilized different operators to simulate typical run-to-run variability. The total %CVs for each SOMAmer reagent was computed across the nine replicates over the three plate runs for all 12 serum or plasma samples after normalization and calibration. The overall median %CVs for intra- assay measurements are generally less than those for total %CV as expected (not shown) and are less than 5% for both serum and plasma. The distribution of the total %CV for all SOMAmer reagents in plasma is in Table 2 of the main text.

Range of quantification

The quantitative performance of this assay has been determined by generating precision profiles for the measured analytes in buffer. Quantification ranges for each analyte in SOMAscan assay were determined in a highly multiplexed experiment run over four separate assay runs (plates) that measured 8-point standard curves spanning 4.2 logs in concentration, typically from 0.3 nM down to 4 fM. Each analyte concentration was measured 3 times per run to determine the assay variation at each concentration. Standard curves were generated by fitting the average Relative Fluorescence Unit (RFU) after normalization and calibration for the 12 replicate measurements at each concentration to a four-parameter logistic curve fit. Two estimates of precision were generated using the %CV of the RFU and of the calculated concentration. 20% CV was chosen to define the upper and lower limits of quantification. A representative standard curve and precision profile is shown in Figure B2. A summary of the distribution of Lower Limit of Quantification (LLOQ), Upper Limit of Quantification (ULOQ) and log range for the SOMAmer reagents in the SOMAscan assay is given in Table 2 in the main text. The median LLOQ for the SOMAmer reagents in the SOMAscan assay is 98 fM, which corresponds to 1 pg/mL for a 10 kDa protein. The median dynamic range of quantification for the SOMAmer reagents is 4.2 logs or $10^{4.2}$.

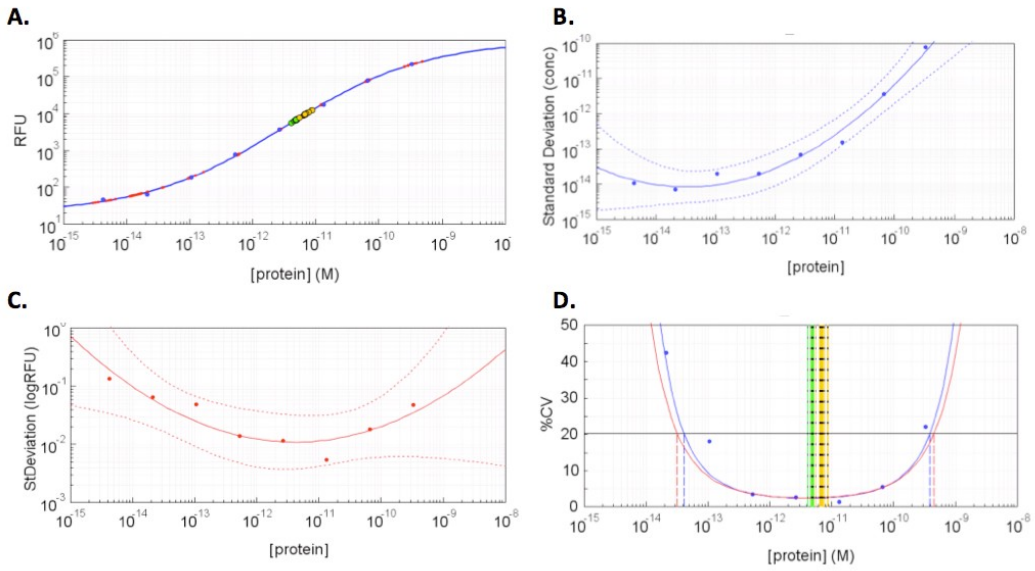


Figure B2. A representative plot from the quantification assessment. (A) The average RFU from the VEGF sR2 SOMAmer reagent at each concentration of input VEGR sR2 protein is denoted by the blue markers and the twelve individual measurements used to compute the average are denoted by the red markers plotted on the four parameter logistic curve fit (solid blue line). The eight replicate serum (orange markers) and plasma (green markers) measurements are denoted on the standard curve. (B) The standard deviation for computed concentration is denoted by the blue markers. The quadratic fit is displayed as a solid blue line and the 95% confidence bands for the fit are displayed as blue dashed lines. (C) The standard deviations for logRFU are denoted by the red markers. The quadratic fit is displayed as a solid red line and the 95% confidence bands for the fit are displayed as red dashed lines. (D) Precision profiles are displayed for the direct computation of the standard deviation of the computed concentrations from B (solid blue line) or using the standard deviations of logRFU from C transformed with the logistic curve fit to standard deviations in concentration (solid red line). Both methods give equivalent results here. The computed %CVs obtained directly from the data are displayed as solid blue markers and the sets of vertical dashed lines correspond to the limits of quantification at 20% CV for the two sets of precision profiles.

Limit of Detection in Buffer

In the limits of quantification experiment described above, four buffer samples were run on each plate to calculate limits of blank (LOB) and limits of detection (LOD). The limit of detection (LOD) was calculated for each analyte using LOD = LOB + 1.645 σ_{low}. Table 2 (main text) displays the percentile distribution for the set of LODs obtained from the SOMAscan assay v3.0. The median LOD is 38 fM, approximately two-fold lower than the median LLOQ. The center fifty-percent of the measurements are between 12 fM – 240 fM. The LOD, ULOQ, LLOQ and quantification ranges for all of the analytes in the array are shown in Table 2 (main text).

Appendix C

SOMAmer Specificity

Over 400 SOMAmer reagents used in the SOMAscan assay have been evaluated for cross reactivity to related proteins using an affinity capture technique very similar to immunoprecipitation. The SOMAmer under evaluation is incubated with either its target protein or with any commercially available protein related to the target by 50% similarity or better. The protein is first incubated with the SOMAmer reagent bound to streptavidin beads. The complex is carried through a series of washes that mimic the first chromatographic step of the SOMAscan assay, with an additional protein labeling step using an AlexaFluorophore. The eluted protein is run on an SDS-PAGE gel and compared to the input protein. Results for affinity capture using SOMAmer reagents to EGFR and ENA-78 are shown in Figures C1 and C2 to demonstrate typical outcomes.

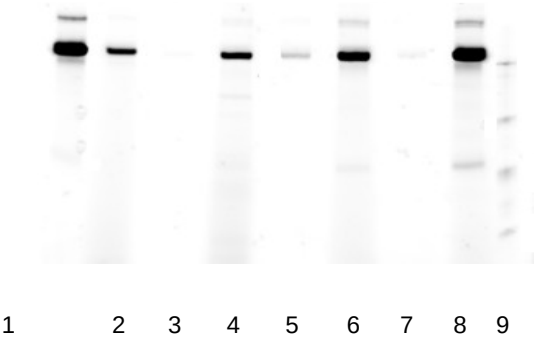
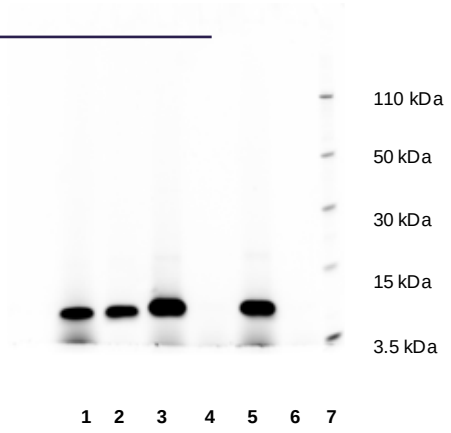


Figure C1. PAGE results of affinity capture assay using SOMAmer to EGFR and purified EGFR and related proteins in buffer. Lanes: (1) EGFR-SOMAmer capture of EGFR, (2) EGFR, (3) EGFR-SOMAmer capture of HER2, (4) HER2, (5)EGFR-SOMAmer capture of HER3, (6) HER3 , (7) EGFR-SOMAmer capture of HER4, (8) HER4, and (9) MW standards.

Figure C2. SDS-PAGE results of affinity capture assay with SOMAmer reagent to ENA-78 using purified ENA-78 and related proteins. Lanes: (1) ENA-78 (2) ENA-78-SOMAmer capture of ENA-78, (3) GCP-2 (4) ENA-78-SOMAmer capture of GCP-2, (5) Gro-α (6) ENA-78-SOMAmer capture of Gro-α, (7) MW standards.



Additional examples of specificity of SOMAmer reagents are provided in Gold *et al.,* 2010. In these experiments proteins were spiked into plasma except for the high abundant proteins LBP and TIG-2, which were captured directly out of plasma, and additional chromatographic steps were employed that mimicked subsequent steps of the assay. In these examples, the resulting eluates are expected to be more representative of the selectivity of SOMAmer reagents to proteins in complex matrices in the context of the assay.

SOMAscan- and SOMAmer-related patents

For a list of our patents, please see the SomaLogic website (www.somalogic.com).

Trademarks

SomaLogic, SomaSciences, and SOMAmer are registered trademarks of SomaLogic, Inc. SOMAscan is a trademark of SomaLogic, Inc.