Verification and Validation of the SOMAscan® Assay 1.3k

Introduction

The SOMAscan Assay is optimized for protein biomarker discovery. Both the individual SOMAmer® reagents included in the assay and the assay itself are validated for the relative quantification of human protein levels in endogenous matrix.

Discovery is enabled by the simultaneous measurement of a broad range of protein targets. The current version of the SOMAscan Assay 1.3k measures >1,300 protein analytes, providing robust and reproducible discovery of biomarker signatures, with performance characteristics unmatched by other proteomics technologies.

Inherent in any highly multiplexed assay is the possibility of false discoveries. Consequently, once discoveries are made using the SOMAscan assay, each biomarker signature should be further validated for its intended use, such as a diagnostic, prognostic, or predictive test. Such validation of individual measurements may include specificity, parallelism, and dilution linearity, as appropriate for the intended application.

The SOMAscan Assay 1.3k increases the number of measured analytes from the previous version SOMAscan Assay containing 1,129 to >1,300, with no modification to the assay protocol. The characterization of SOMAscan Assay 1.3k consists of two parts, reproducibility and reference range generation. This note describes the methods and analysis used to verify/validate the new SOMAscan Assay 1.3k performance.

Methods

Reproducibility Study

The reproducibility of the SOMAscan Assay 1.3k was assessed in human serum and plasma by analyzing 11 individuals as single measurements and three individual samples in triplicate over three independent runs (same buffer lots, but different operators and two different robotic systems). Included in each of the three runs was a set of controls (seven replicates of the calibrator sample, two QC samples each assayed in duplicate, and one blank sample) routinely analyzed in the assay. These data allow for the calculation of intra- and inter-run and total assay precision as well as other assay metrics defined below.

Reference Range Study

Serum and plasma samples from 166 individuals spanning different ethnicities, ages and genders were analyzed using the SOMAscan Assay 1.3k. This data set was used to define reference ranges for each measured analyte for a population of healthy individuals. In addition, these data were used to compute matrix limits of detection (mLOD) on a per SOMAmer reagent basis.

Analysis

All analyses were completed separately for both human serum and plasma measurements.

Reproducibility Study

The three samples run in triplicate across three independent runs were used to compute the total assay coefficients of variation (CV) as well as the intra- and inter-assay contributions to the CV. Spearman correlations for the fourteen individuals were computed between the three independent runs. In addition, the measurements from the fourteen individuals and the assay CVs were used to assess population F-statistics. The blank sample measurements, along with the median signal from the fourteen individuals, allow for an evaluation of signal-to-noise for the assay.

Reference Range Study

Reference ranges were computed by constructing an empirical cumulative distribution function (CDF) for each analyte measurement within the population of 166 healthy individuals. The lower and upper reference ranges in RFU were directly obtained from the CDFs where CDF(lower range RFU) = $\alpha/2$ and CDF(upper range RFU) = $1 - \alpha/2$.

mLODs for a matrix (either serum or plasma) were computed from these data when the data are either:

1. consistent with a log-normal distribution and the correlation between serum and plasma measurements is low and correlations between repeated runs within the matrix from the reproducibility data is low, or
2. the sum of two log-normal distributions and the lower RFU measurements have low correlation in plasma and serum.

In either case, the mLOD is computed from the log-normal fit by

$$\log(mLOD) = \mu \pm z_{0.975}\sigma$$

where $z_{0.975}$ is the 97.5% quantile of the normal distribution with mean and standard deviation . For the second case, the fit describing the low correlation RFUs is used to compute the mLOD.

Results

The median total CVs are comprised of almost equal amounts of intra-assay and inter-assay variation. The distribution of total CVs is slightly higher for plasma than for serum, with medians of 4.6% and 2.9%, respectively. Reference ranges in normal healthy individuals were determined for all analytes in SOMAscan Assay 1.3k.