Proteomics data description

Eva Borras, Mariona Bustamante, Juan Ramon González, Léa Maitre, Eduard Sabido

27/06/2017

1. Introduction

Targeted mass spectrometry (MS) via SRM and the antibody-based multiplexed platforms such as Luminex have emerged as an alternative to traditional ELISA measurements of defined protein sets. The main advantage of these techniques is the capacity for faster and cost-efficient assay development and their ability to quantify multiple proteins in parallel (multiplexing) at a low limit of detection and high accuracy. Luminex was applied to obtain quantitative measures of candidate proteins in HELIX.

2. Assay description

A set of 43 proteins were selected a priori based on the literature and on the Luminex kits commercially available from Life Technologies and Millipore.

Three kits were selected for the subsequent analyses, which assessed a total 50 measurements that represent 43 unique proteins. The kits namely were the human kits named Cytokines 30-plex (Cat #. LHC6003M), Apoliprotein 5-plex (LHP0001M) and Adipokine 15-plex (LHC0017M), and based on magnetics beads (Life Technologies). Table 1 below describes the proteins measured in each kit, which range from cytokines to hormones. The following 7 proteins were repeated between two kits: IL1beta, IL6, IL8, IL10, MCP1, HGF, TNFalfa.

Cytokines 30plex	Apoliprotein 5plex	Adipokine 15plex
FGF	ApoA1	IL 1beta
IL 1beta	ApoB	IL 6
G CSF	ApoE	IL 8
IL 10	Adiponectin	IL 10
IL 13	CRP	MCP 1
IL6		Leptin
IL 12		SAA
RANTES		HGF
Eotaxin		Insulin
IL 17		Lipocalin 2
MIP 1alfa		TNF alfa
GM CSF		BAFF
MIP 1beta		Resistin
MCP 1		C-peptide
IL 15		Serpine1/PAI-1
EGF		
IL 5		
HGF		
VEGF		
IFN gamma		
IFN alfa		
IL 1RA		
TNF alfa		
IL 2		
IL 7		
IP 10		
IL 2R		
MIG		
IL 4		
IL 8		

Table 1. Proteins targeted in each of the three Magnetic Human Luminex Kits from Life Technologies, namely named Cytokines 30-plex (Cat #. LHC6003M), Apoliprotein 5-plex (LHP0001M) and Adipokine 15-plex (LHC0017M).

3. Laboratory processing

The Luminex System is a multiplex bead-based system that combines microspherebased bioassays with digital signal processing. A Luminex xMAP assay configuration consists of a suspension array where specific capture moieties are covalently coupled to the surfaces of internally dyed microspheres. After completion of assay incubations with a detection reagent, the beads are separated within a Luminex analyzer and interrogated with two LEDs – one for classification of the bead identity and the other for quantification of bound reporter fluorophore. First, Luminex color-codes tiny beads, called microspheres, into 100 distinct sets. Each bead set can be coated with a reagent specific to a particular bioassay, allowing the capture and detection of specific analytes from a sample. Within the Luminex analyzer, LEDs excite the internal dyes that identify each microsphere particle, and also any reporter dye captured during the assay. Many readings are made on each bead set, further validating the results. In this way, xMAP technology allows multiplexing of up to 100 unique assays within a single sample, both rapidly and precisely.

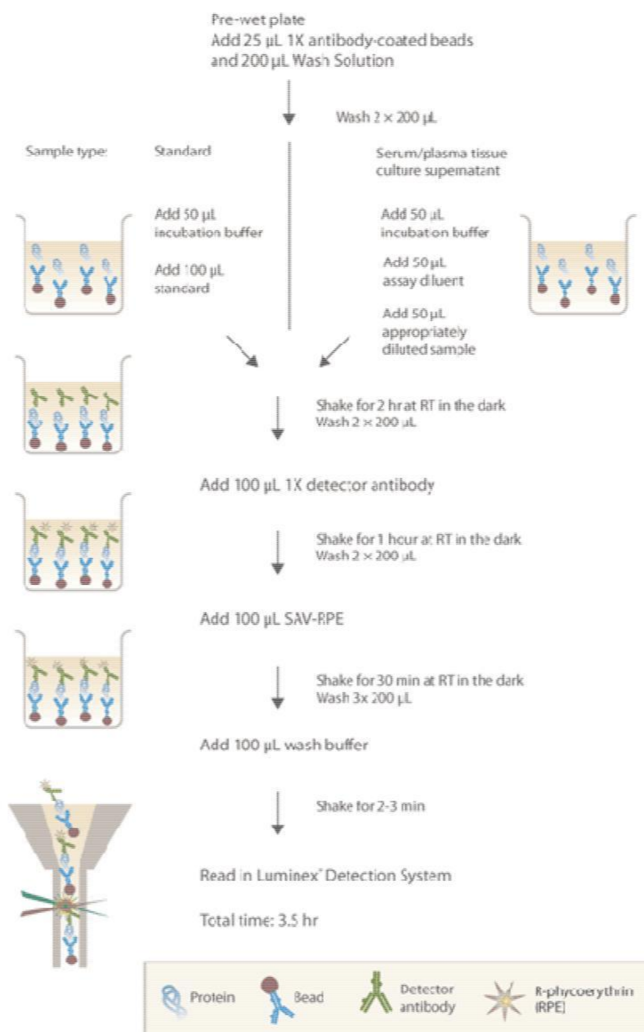

Figure 1 Luminex xMAP assay protocol extracted from the Luminex vendor reference documentation.

Table 2. Plasma samples used in the Proteomics study classified by cohort (n=1,368).

cohort	panel 1	panel 2	subcohort	Total
INMA	38	40	182	260
RHEA	29	30	167	226
BIB	36	28	168	232
EDEN	28	28	129	185
MOBA			230	230
KANC	31	29	175	235
Total	162	155	1051	1368

4. Final dataset

The final dataset consists of 1343 samples and 36 proteins, distributed as follows:

Period	HELIX		Total
	no	yes	
1X	18	1020	1038
1A	0	150	150
1B	1	154	155
Total	19	1324	1343

Subcohort.all	1188
Subcohort.HELIX	1170
Panel.paired	149

Two ExpressionSet files containing log transformed, imputed and normalized protein levels were created:

- one for the subcohort (1X+1A): proteome_subcohort_v2.Rdata
- one for the panel study (1A+1B): proteome_panel_v2.Rdata

An ExpressionSet is an R object that contains:
- values of the protein levels
- annotation of the omics features: UniProtKB, Gene_Symbol and Gene_Name
- metadata of the samples (laboratory variables (plate) and biological variables)

For more information on ExpressionSet visit https://www.bioconductor.org/packages/devel/bioc/vignettes/Biobase/inst/doc/ExpressionSetIntroduction.pdf.

Raw data can be sent under request.

5. Summary methods

Proteomics laboratory processing

A set of 43 proteins were selected a priori based on the literature and on the Luminex kits commercially available from Life Technologies and Millipore. Three kits were selected for the

subsequent analyses, which assessed a total 50 measurements: Cytokines 30-plex (Cat #. LHC6003M), Apoliprotein 5-plex (LHP0001M) and Adipokine 15-plex (LHC0017M).

1,367 plasma samples from HELIX cohorts were analysed in the CRG/UPF Proteomics Unit with the xMAP and Luminex system and following the manufacturer's protocol. The experimental design was planned to control for batch effects and confounding variables. All samples from 6 cohorts were randomized and blocked by cohort prior measurement to ensure a representation of each cohort in each measurement plate (batch). No distinction was made by panel, nor by sub-cohort, and nor by gender. Nineteen different plates of each kit were needed to measure the whole collection of samples. For protein quantification, an 8-point calibration curve per plate was performed with protein standards provided in the Luminex kit and following the procedures described in the standard procedures described by the vendor. Commercial heat inactivated, sterile-filtered plasma from human male AB plasma (Sigma Cat #. H3667) was used as constant controls to control for intra- and inter-plate variability. Four control samples were added per plate. No duplicate measurements were done for the HELIX samples. All samples—controls and HELIX samples—were diluted ½ for the 30-plex kit, ¼ for the 15-plex kit and 1/2500 for the 5-plex kit.

Data quality control

Raw intensities obtained with the xMAP and Luminex system for each plasma sample were converted to ng/ml (5-plex kit: adiponectine, CRP, APO-A1, APO-B, APO-E) and to pg/ml (15 and 30-plex kits) using the calculated standard curves of each plate and accounting for the dilutions that were made prior measurement. The % of coefficients of variation (% CV) for each protein estimated by plate and then averaged ranged from 3.42% to 36%.

For each protein, the limit of detection (LOD) was determined and the lower and upper quantification limits (LOQ1 and LOQ2, respectively) were obtained from the calibration curves. Seven proteins were removed because they had <30% of measurements in the linear range (3 of them had values <LOD or <LOQ1: IL7, VEGF, GMCSF; and 4 of them had values >LOQ2: Lipocalin2, RANTES, Resistin, SAA). Seven proteins were measured in two different plex and the measure with lower quality was excluded from the analysis.

Twenty four samples were deleted from the study (20 because they had 10 or more proteins below the LOD or above the upper limit of quantification and 4 because they did not had an ethic consent or were excluded from the study for other reasons).

For the 36 proteins that passed the QC, data was log transformed to reach normal distribution. Then, the plate batch effect was corrected by subtracting for each individual and each protein the difference between the overall protein average minus the plate specific protein average. Finally, values below LOQ1 and above LOQ2 were imputed using a truncated normal distribution implemented in the truncdist R package (Nadarajah & Kotz, 2006). This method performs imputation of missing data using random draws from a truncated distribution with parameters estimated using data in the linear range of each protein (e.g between LOQ1 and LOQ2).

The final dataset contains the log transformed, imputed and normalized levels for 36 proteins for 1343 HELIX samples (1038 - 1X, 150 - 1A, and 155 - 1B (149 paired 1A+1B)), 1324 included in the subcohort study.