EBioMedicine xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Effects of Long-Term Storage Time and Original Sampling Month on Biobank Plasma Protein Concentrations

Stefan Enroth [a,*], Göran Hallmans [b], Kjell Grankvist [c], Ulf Gyllensten [a]

[a] Department of Immunology, Genetics, and Pathology, Biomedical Center, SciLifeLab Uppsala, Uppsala University, SE 75108 Uppsala, Sweden
[b] Department of Biobank Research, Umeå University, SE 90187 Umeå, Sweden
[c] Department of Medical Biosciences, Clinical Chemistry, Umeå University, SE 90185 Umeå, Sweden

ARTICLE INFO

Article history:
Received 4 July 2016
Received in revised form 19 August 2016
Accepted 24 August 2016
Available online xxxx

Keywords:
Plasma proteins
Biobank
Covariate
Storage time
Sampling month
Proximity extension assay

ABSTRACT

The quality of clinical biobank samples is crucial to their value for life sciences research. A number of factors related to the collection and storage of samples may affect the biomolecular composition. We have studied the effect of long-time freezer storage, chronological age at sampling, season and month of the year and on the abundance levels of 108 proteins in 380 plasma samples collected from 106 Swedish women. Storage time affected 18 proteins and explained 4.8–34.9% of the observed variance. Chronological age at sample collection after adjustment for storage-time affected 70 proteins and explained 1.1–33.5% of the variance. Seasonal variation had an effect on 15 proteins and month (number of sun hours) affected 36 proteins and explained up to 4.5% of the variance after adjustment for storage-time and age. The results show that freezer storage time and collection date (month and season) exerted similar effect sizes as age on the protein abundance levels. This implies that information on the sample handling history, in particular storage time, should be regarded as equally prominent covariates as age or gender and need to be included in epidemiological studies involving protein levels.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The stability of biomolecules in clinical biobanks is essential to many applications in life sciences. The short-term stability is important in order to handle batch effects in the laboratory analyses, while the long-term stability is essential for collection of samples for rare-diseases, for monitoring individual variance in longitudinal studies over extended time periods, and for retrospective studies with samples stored over different time-spans or storage methods. Biobanks should strive to maintain the biomolecular composition of the sample at the time of collection, in order to allow for unbiased comparisons of samples obtained using different methods, collected at different time points, stored over different time-spans, and subjected to different number of freeze-thaw cycles. In previous studies of solid tissue or plasma samples that have been stored frozen, parameters such as sample handling and short-term effects of storage regimes on blood samples before freezing have been examined (Skogstrand et al., 2008). Plasma and serum cytokines concentrations have been found to increase with storage time before separation and freezing. It has also been shown that concentrations of three of the investigated proteins in sets of brain tissue were significantly affected by freezer storage time (Harish et al., 2011). The number of freeze-thaw cycles can have an effect on different proteins (Lee et al., 2015). In addition, other factors, such as time of day, weekday and general intra-individual variation have also been described (Stenemo et al., 2016). However, there have been no examinations of a large set of proteins for these parameters.

Sample collection for a cohort is often carried out at the time of diagnosis, time of treatment or time of screening of healthy individuals, and not confined to a certain time of the year. This implies that environmental factors, such as pollen concentration in the spring, may cause downstream effects on proteins in immune response pathways. For instance, plasma levels of soluble intercellular adhesion molecule-1 (sICAM-1) and soluble E-selectin have been shown to be increased in inhalation allergic children compared to healthy controls (Reich et al., 2003). Also, seasonal differences in variables such as sunlight and ultra violet (UV)-radiation could affect the abundance levels of circulating proteins. This is particularly relevant in parts of the world where there are large differences in amount of solar radiation between different parts of the year. UV-radiation causes inflammatory responses in the skin and several cytokines are affected by UV-radiation. Specifically, production of C-X-C motif chemokine ligand 5 (CXCL5) have been shown to be up-regulated in human skin (Reichert et al., 2015) in response to UV-radiation and changes in plasma levels of several interleukins have been shown in vivo in mouse models (Vostalova et al., 2013), suggesting that seasonal variation is likely to be measurable in human plasma. In order to investigate the effect of storage time in freezer, chronologic

* Corresponding author at: Stefan Enroth, Biomedical Centre (BMC), Box 815, SE 75108 Uppsala, Sweden.
E-mail address: stefan.enroth@igp.uu.se (S. Enroth).

http://dx.doi.org/10.1016/j.ebiom.2016.08.038
2352-3964/© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: Enroth, S., et al., Effects of Long-Term Storage Time and Original Sampling Month on Biobank Plasma Protein Concentrations, EBioMedicine (2016), http://dx.doi.org/10.1016/j.ebiom.2016.08.038

age at sampling, sampling month, and season (sunlight), we determined the abundance level of 122 plasma proteins using the Protein Extension Assay (PEA) in 106 women sampled at 380 occasions spanning from 1988 to 2014.

2. Materials and Methods

2.1. Samples

380 female plasma samples from the Västerbotten intervention program (Hallmans et al., 2003) (VIP, N = 208) and Mammography Screening Project (MA, N = 172) were obtained from the Umeå Biobank in Sweden. The plasma samples were drawn between 1988 and 2014 and donated by 106 women. The plasma was treated with ethylenediaminetetraacetic acid (EDTA) separated and frozen within 1 h after sampling, followed by long time storage in −80 °C. In the VIP cohort almost all samples has been taken after overnight fasting. The samples analyzed had not previously been thawed.

2.2. Protein Abundance Measurements

Plasma protein abundance levels were quantified using the Proximity Extension Assay (PEA) (Assarsson et al., 2014) multiplex assays from Olink Proteomics AB, Uppsala, Sweden. In brief, for each observed protein, a pair of oligonucleotide-labeled antibodies probes bind to the targeted protein, and if the two probes are in close proximity, a PCR target sequence is formed by a proximity-dependent DNA polymerization event and the resulting sequence is subsequently detected and quantified using standard real-time PCR. Data is then normalized and transformed using internal extension controls and inter-plate controls, to adjust for intra- and inter-run variation. The final assay read-out is given in Normalized Protein eXpression (NPX), which is an arbitrary unit on log2-scale where a high value corresponds to a higher protein expression. Each protein has specific lower limit of quantification (LLOQ) and an upper limit of quantification (ULOQ) between which a 1-unit increase in NPX correspond to a two-fold increase in protein concentration. Internal controls specify a run-specific lower limit of detection (LOD) as 3 times the standard deviation over background. The manufacturers webpage (www.olink.com) contains information on protein-specific detection limits and records of technical validations of intra- and inter-variability (in terms of coefficients of variation (CV) of each analyzed protein. Samples were analyzed for the complete Proseek Multiplex Inflammation panel I and 20 proteins each from the Proseek Multiplex CVD I panel and Proseek Multiplex Oncology II panels (http://www.olink.com/proseek-multiplex/). The plasma samples were analyzed, quality-controlled and pre-processed into NPX at Olink Proteomics AB, Uppsala, Sweden. In total, 122 unique proteins were quantified out of which 10 were overlapping between the Inflammation panel and either of the other panels.

2.3. Statistical Analysis

All statistical analyses were carried out in R (R Development Core Team, 2014). Effect of storage time on protein levels was examined by fitting a linear model with protein abundance levels as response and whole years since original sampling date as variable. Significance was evaluated by ANOVA-analysis. Proteins where storage time was found to be a significant influence was adjusted by removing the contribution of storage time using the beta (β)-coefficient from the linear models as described above. Contribution of individual age to protein abundance levels was analyzed in the same way using storage-time adjusted protein levels as input. Monthly and seasonal variation in protein levels was examined using storage time and age-adjusted protein levels using a two-sided Wilcoxon ranked-sum test. Contribution of monthly sun hours to protein levels were examined by fitting a linear model with protein levels as response and sun hours as variable. This analysis was repeated with a 1–11 month shift in sun hours, allowing for a possible delay in protein levels depending on sun activity. Significance was evaluated by ANOVA analysis. Calculations of the variance explained by a variable to the response, were carried out using the 'summary'-function applied to the fitted linear models.

2.4. Ethical Considerations

The study has been approved by the Regional Ethics Board in Umeå, Dnr 2013-314-32M and 2012-229-31M.

3. Results

3.1. Study Design

The Västerbotten Intervention Program (Hallmans et al., 2003) invites participants at 10-year intervals starting from the age of 30 or 40, while women conduct mammography check-ups at intervals normally spaced 18–24 months apart. Because of the different sampling-strategies in the VIP and the MA-cohorts, the VIP-samples allows for study of the impact of storage time independent of the actual age of the individual, by restricting the analysis to a given age-span. The samples from the MA-cohort have a more uniform distribution of ages and, when corrected for storage time, are suited for analysis of the contribution of chronological age at sampling to protein levels. Among the samples used here, the oldest was collected in 1988 and the most recent in 2014 (Fig. 1A). Out of the 106 women, 96 had >1 sample, and the maximum age-range for a single woman was 33.5 years (first sample at age 40.3 and latest at 73.8 years) (Fig. 1B). Ages of participants in the VIP ranged from 29 to 70 and in MA from 42 to 73. Overall the samples were collected over the entire year, with the exception of July with 0 samples (Fig. 1C). July is traditionally the time of summer vacation in Sweden and no samples have been collected during this month.

3.2. Protein Abundance Levels

Relative abundance levels for 132 proteins were quantified using the Proximity Extension Assay (PEA) in 380 plasma samples from 106 Swedish women in the Umeå-area from the VIP and MA-cohorts. Each of the 132 assays had a lower limit of detection specified by the manufacturer and here, 14 of the 132 (10.6%) assayed proteins had >80% of the individual measurements below the detection limit. This is similar to the pattern seen previously when analyzing protein levels from these panels in non-disease cohorts (Enroth et al., 2015a; Enroth et al., 2014). These 14 proteins (IL2RB, IL1a, IL2, TSLP, RA1, PD-L1, IL24, ARTN, TNF, IL20, IL33, IL4, LIF and NRTN) were excluded from further statistical analysis. Moreover, 10 of the 118 remaining proteins were replicated over two different panels (Material and Methods). Each pair of replicated assays was significantly correlated ($p < 8.0 \times 10^{-53}$, Spearman's Rho), with a mean correlation coefficient (R^2) of 0.78. The highest correlation was found for C-X-C motif chemokine ligand 11 (CXCL11) with $R^2 = 0.93$ ($p < 2.2 \times 10^{-209}$), while the lowest was found for the Cluster of differentiation 40 protein (CD40) with $R^2 = 0.62$ ($p < 1.4 \times 10^{-77}$) (Supplementary Fig. 1).

3.3. Correlation with Storage Time

Since the majority of samples in our data have been sampled on multiple occasions, it is difficult to separate the effect of the ageing of the individuals from the storage time in the freezer. Therefore, we restricted the analysis of storage time to 92 individuals' aged 50 ± 0.5 years. Due to the sampling strategy of the VIP-study, this age-interval contains the highest number of samples (Fig. 1A). The storage time distribution among these individuals does not differ from the rest of the samples ($p > 0.4$, Wilcoxon ranked sum test), and is thus representative of the complete material. Analysis of coefficients of variance (CV) for all

Please cite this article as: Enroth, S., et al., Effects of Long-Term Storage Time and Original Sampling Month on Biobank Plasma Protein Concentrations, EBioMedicine (2016), http://dx.doi.org/10.1016/j.ebiom.2016.08.038

Download English Version:

https://daneshyari.com/en/article/8439232

Download Persian Version:

https://daneshyari.com/article/8439232

Daneshyari.com