

Biological Variation of the Platelet Proteome in the Elderly Population and Its Implication for Biomarker Research^{*S}

Wolfgang Winkler‡§, Maria Zellner§¶, Michael Diestinger¶, Rita Babeluk¶, Martina Marchetti‡, Alexandra Goll||, Sonja Zehetmayer||, Peter Bauer||, Eduard Rappold**, Ingrid Miller‡‡, Erich Roth¶, Günter Allmaier‡, and Rudolf Oehler¶§§

Knowledge about the extent of total variation experienced between samples from different individuals is of great importance for the design of not only proteomics but every clinical study. This variation defines the smallest statistically significant detectable signal difference when comparing two groups of individuals. We isolated platelets from 20 healthy human volunteers aged 56–100 years because this age group is most commonly encountered in the clinics. We determined the technical and total variation experienced in a proteome analysis using two-dimensional DIGE with IPGs in the pI ranges 4–7 and 6–9. Only spots that were reproducibly detectable in at least 90% of all gels ($n = 908$) were included in the study. All spots had a similar technical variation with a median coefficient of variation (cv) of about 7%. In contrast, spots showed a more diverse total variation between individuals with a surprisingly low median cv of only 18%. Because most known biomarkers show an effect size in a 1–2-fold range of their cv, any future clinical proteomics study with platelets will require an analytical method that is able to detect such small quantitative differences. In addition, we calculated the minimal number of samples (sample size) needed to detect given protein expression differences with statistical significance. *Molecular & Cellular Proteomics* 7:193–203, 2008.

Biomarkers are used to differentiate between different biological states and to monitor disease progress or the success of medical treatment. To fulfill these tasks, there are not only strict requirements for each biomarker candidate in terms of selectivity and specificity but also for the precision of the analytical process. Therefore knowledge on the extent of var-

iation between the individuals within each group is crucial for the study design as well as the selection of the method of measurement. The total variation experienced between individuals within one group is the result of the biological variation caused by factors like sex, age, genetic background, lifestyle, or health status and the technical variation introduced by the applied sample handling and the method of measurement itself. This study investigated the extent of variation in human blood platelets. Platelets are responsible for the maintenance of vascular integrity and are also involved in inflammation and wound healing (1). They are anucleate cytoplasmic fragments released from megakaryocytes in the bone marrow. During platelet biogenesis, organelles, especially the mitochondria and the α - and dense granules, are actively transported into the platelets. Furthermore platelets receive a certain set of mRNAs during their biogenesis from the megakaryocytes and are still capable of protein synthesis and processing (2). Several proteomics studies have focused on different aspects in platelet research using 2D¹ electrophoresis followed by mass spectrometry. These studies provided 2D maps of the total platelet proteome (3, 4), the platelet secretome (5), and the processes during platelet activation via analysis of phosphoproteins (6, 7). But to our knowledge there are no studies published until now that analyze quantitative differences in protein expression levels between groups of individuals. To separate platelet proteins we used 2D DIGE with cyanine dyes and minimal labeling (8). On each gel, a sample was separated together with an internal standard (a mixture of equal amounts of all samples of the respective experiment). We determined the extent of variation in protein expression levels experienced between individuals of a group of 20 healthy volunteers. On the basis of the results a sample size calculation for future platelet proteomics studies was made.

¹ The abbreviations used are: 2D, two-dimensional; V_{norm} , background-corrected intragel-normalized fluorescence volume; SA, standardized abundance; PMF, peptide mass fingerprint; cv, coefficient of variation; FDR, false discovery rate; 2DE, two-dimensional gel electrophoresis; apoA-I, apolipoprotein A-I.

From the ‡Institute of Chemical Technologies and Analytics, Vienna University of Technology, A-1060 Vienna, Austria, ¶Department of Surgery and ||Section of Medical Statistics, Medical University of Vienna, A-1090 Vienna, Austria, **Geriatric centre of the Otto Wagner Hospital, A-1140 Vienna, Austria, and ‡‡Institute of Medical Chemistry, Department of Natural Sciences, University of Veterinary Medicine, A-1210 Vienna, Austria

Received, March 26, 2007, and in revised form, October 11, 2007
Published, MCP Papers in Press, October 25, 2007, DOI 10.1074/mcp.M700137-MCP200

EXPERIMENTAL PROCEDURES

Blood Sampling—With approval of the local ethics committee, blood was drawn from 20 healthy fasting volunteers (six male and 14 female; average age, 77.3 ± 12 years). Blood was drawn without stasis from the antecubital vein into vacuum tubes (Greiner Bio-One, Kremsmünster, Austria) containing 0.129 mol/liter sodium citrate as anticoagulant (mixing ratio, 1:9 with blood). The first drawn tube was discarded.

2D Sample Preparation—The platelet isolation and the protein precipitation were carried out as we described previously (9). Briefly centrifugation of blood samples followed by size exclusion chromatography of the supernatant was used to obtain a pure plasma protein-free platelet suspension. Out of this suspension, the platelet proteins were precipitated using trichloroacetic acid. The precipitated proteins were washed in acetone (Merck) enriched with 20 mM DTT (Roche Diagnostics) and resolubilized using 70 μ l of 2D sample buffer (7 M urea, 2 M thiourea, 4% CHAPS, 30 mM Tris-HCl, pH = 8.5) per 100 million platelets. The internal standard was made by pooling equal amounts of numerous different platelet samples including all samples of this study. Labeling with fluorescent cyanine minimal dyes (CyDyes[®], GE Healthcare) was carried out according to the manufacturer's instructions. However, the ratio of CyDye to protein was reduced to 7 pmol/ μ g, giving the same results/patterns as the originally described ratios. The internal standard was always labeled using Cy2. The samples were labeled alternately with Cy3 and Cy5.

2D Electrophoresis and Image Acquisition—Twenty-four-centimeter pH 4–7 IPG DryStrips (GE Healthcare) were passively rehydrated for 14 h at room temperature within a mixture containing 33 μ g of labeled sample and 33 μ g of labeled internal standard, and the volume was brought up with rehydration solution (7 M urea, 2 M thiourea, 4% CHAPS), DTT, and ampholytes pH 4–7 (Serva, Heidelberg, Germany) to a final volume of 450 μ l with a final concentration of 70 mM DTT, 0.5% ampholytes. Isoelectric focusing was carried out on an IPGphor device (GE Healthcare) until 30 kV-h were reached. Twenty-four-centimeter pH 6–9 IPG DryStrips were passively rehydrated for 14 h at room temperature with 450 μ l of rehydration solution containing 150 mM DTT and 2% ampholytes pH 6–9 (Serva). A mixture containing 25 μ g of labeled samples, 25 μ g of labeled internal standard, 150 mM DTT, 2% ampholytes pH 6–9, and the volume of 2D sample buffer to reach 55 μ l was applied to the DryStrip via cup loading, and IEF was performed with simultaneous loading of 3.5% DTT from the cathode until 30 kV-h were reached. Prior to SDS-PAGE the IPG strips were equilibrated for 20 min in equilibration solution (50 mM Tris-HCl, 6 M urea, 30% glycerol, 2% SDS, pH = 8.8) containing 10 mg/ml DTT followed by equilibration for 15 min in equilibration solution containing 25 mg/ml iodoacetamide (Sigma). Thereafter SDS-PAGE was performed on homogeneous gels (12.5% acrylamide) using the Ettan DALT six equipment (GE Healthcare). Electrophoresis conditions were: 35 V for 1 h, 50 V for 1.5 h, and finally 110 V for 16.5 h. Then gels were scanned using a Typhoon TRIO scanner (GE Healthcare) with a resolution of 100 μ m. Sensitivity (photomultiplier voltage) was adjusted in a prescan with 1000- μ m resolution so that no channel had saturated spots (exception: the actin spot in pI range 4–7 was always saturated). After scanning, the gels were silver-stained and stored in 1% acetic acid at 4 °C until protein identification. Spot detection was performed on the gel images using the DeCyder software module Differential In-gel Analysis (version 6.00.28) setting the target spot number to 2800 (pH range 4–7) and 2500 (pH range 6–9), respectively. Saturated spots (peak height above 65,000) were excluded from further analysis. Next an exclusion filter for spots with a slope greater than 1.4 or a volume below 25,000 was applied. Finally all spots that had been removed by exclusion filters were manually reviewed to ensure no false exclusion. The fluorescent signal of a spot was background-corrected and nor-

malized by the DeCyder software (version 6.00.28) according to the manual. Briefly for background correction the 10th percentile pixel value on the spot border was subtracted. For intragel normalization a histogram was plotted for spot frequency against \log_{10} spot volume, a Gaussian distribution was fitted, and the center of this curve was used as a correction factor for each spot. The resulting normalized volume was denoted as V_{norm_g} . Each gel was added to the appropriate workspace and group and matched against the master gel using the DeCyder module Biological Variation Analysis (version 6.01.02). Before the matching process was started the first time, up to 250 landmarks were defined all over the gel within high density regions where the algorithm fails to give correct matches. Subsequently the matching process was started followed by reviewing all matches that were classified by the algorithm as a level 1 match (automatic matches with high confidence) and confirming all correct matches. The cycle of matching, reviewing, and confirming the matches was repeated until no new level 1 matches were found. Using the DeCyder module XMLToolbox, the spot raw data values, especially the normalized volume V_{norm_g} for each gel, were exported from DeCyder and analyzed using a spreadsheet software.

Western Blotting—For Western blotting proteins were separated on a 2D gel as described above and transferred to a nitrocellulose membrane by electroblotting. Expression of specific proteins was revealed with goat antibodies against human fibrinogen (G11, Biomedica, Burlingame, CA). Bound antibodies were detected with a peroxidase-conjugated anti-goat IgG antibody (Sigma) in the presence of SUPERSIGNAL[®] substrate (Pierce).

Protein Identification—The spots of interest were excised manually and subjected to in-gel digestion (10) using trypsin (from bovine pancreas, modified; sequencing grade; Roche Diagnostics). The digests were desalted and concentrated using ZipTips (C₁₈, Millipore, Bedford, MA) and analyzed on a MALDI-TOF/reflectron TOF instrument (TOF² (11, 12); Shimadzu Biotech, Manchester, UK) applying the thin layer preparation technique (13) using α -cyano-4-hydroxycinnamic acid (Sigma-Aldrich) as matrix. All mass spectra were recorded in positive ion reflectron mode with delayed extraction by accumulating up to 500 single unselected laser shots. Internal calibration was performed using the singly charged peptides from tryptic autodigestion at m/z 659.38, 805.42, 1153.57, and 2163.06 (monoisotopic masses). The lists of monoisotopic m/z values (signal to noise ratio of a signal had to be above 3:1, and a plausible isotopic pattern had to be observable) for peptide mass fingerprint (PMF) searches were manually derived from unsmoothed mass spectra without base-line subtraction, omitting those values related to tryptic autodigestion, contamination from keratins (14), and matrix cluster ions (15). Peak annotation was carried out automatically using software provided by the instrument manufacturer (Launchpad, version 2.7.1.20060929; Shimadzu Biotech). Typically spectra showing a resolution of 9000–12,000 (full-width half-maximum, m/z 2163.06) and an average mass accuracy of ± 40 ppm could be observed. The m/z lists were submitted to the publicly available, web-based versions of PMF search engines ProFound (16) and MASCOT (version 2.1, January/February 2007) (17) to search the Swiss-Prot protein sequence database (version 51.5, 15,580 sequences assigned to *Homo sapiens*) (18) (MASCOT only) and the non-redundant database from the National Center for Biotechnology Information (NCBI) (version 20061013, 151,998 sequences for *H. sapiens*) using the following settings: fixed modification, carbamidomethylation on Cys; variable modifications, oxidized methionine, protein N-terminal acetylation, and formation of pyroglutamic acid on peptide N-terminal Glu; missed cleavages, 1; species, *H. sapiens*; m/z tolerance, ± 200 ppm; molecular weight, value estimated from 2D gel $\pm 33\%$, ProFound only; and isoelectric point, value estimated from 2D gel ± 2 , ProFound only. To confirm the PMF results, PSD and high energy CID experiments (collision gas,

TABLE I
List of individuals and labeling strategy
f, female; m, male.

Sample	Age	Sex	Dye	
			pl 4–7	pl 6–9
	yr			
1	100	f	Cy5	Cy5
2	93	m	Cy5	Cy5
3	89	f	Cy5	Cy5
4	88	f	Cy5	Cy5
5	87	f	Cy5	Cy3
6	87	f	Cy3	Cy3
7	84	f	Cy3	Cy5
8	82	f	Cy3	Cy5
9	82	f	Cy5	Cy5
10	82	f	Cy3	Cy3
11	75	f	Cy5	Cy5
12	73	f	Cy5	Cy3
13	73	f	Cy3	Cy3
14	72	m	Cy3	Cy3
15	71	f	Cy5	Cy5
16	66	f	Cy3	Cy3
17	64	m	Cy3	Cy3
18	62	m	Cy5	Cy5
19	59	m	Cy5	Cy5
20	56	m	Cy3	Cy3

helium; isolation width, precursor ion, ± 5 Da; $E_{\text{lab}} = 20$ keV) were performed on selected tryptic peptides using the same sample preparation technique as described above. Mass spectra were recorded in positive ion mode with delayed extraction by accumulating up to 3000 single unselected laser shots. Typically spectra showing a resolution of 400–500 (full-width half-maximum; m/z range, 500–1500) and an average mass accuracy of ± 0.5 Da were acquired. The m/z values for MS/MS ion searches were manually derived as average values from smoothed spectra (Savitsky-Golay algorithm (19) as implemented by the instrument manufacturer). Signals showing a signal to noise ratio of above 2:1 were chosen to be submitted to the publicly available, web-based version of MASCOT MS/MS search engine to search the Swiss-Prot and NCBI non-redundant database using the same restrictions as described for PMF searches except instrument type was MALDI-TOF/TOF and m/z tolerance was set to ± 1 Da for product ions and ± 0.25 Da for the precursor ion. A hit was considered to be significant if the scores obtained for PMF and MS/MS data clearly exceeded the significance threshold ($p < 0.05$) of each algorithm. In case of multiple hits for the same set of m/z values, these sequences were checked for redundancy by generating alignments with ClustalW (20). In all cases PMF and MS/MS data were sufficient for significant protein identification as there was no emphasis laid on further discrimination of isoforms.

Statistical Methods—Arithmetic mean, standard deviation (σ), and coefficient of variation (cv) are shown for the description of the data. Spearman rank correlation coefficients were computed to show the coherence between two measures. For the graphical illustration, histograms and scatter plots are given. We performed asymptotic sample size calculations controlling the false discovery rate (FDR), defined as the expected fraction of erroneously rejected hypotheses under all rejected hypotheses (21). Here the asymptotic power per hypothesis of the two-sided two-sample t test depending on the sample size per group (starting with two observations per group) for different effect sizes is given. An asymptotic FDR (22) of 5% targeted at 500 and 1500 candidate proteins was investigated where 10 and 20 of the proteins are effective, respectively. For small power values an approximation using the Bonferroni level was performed because the asymptotic FDR does not yield reliable results.

RESULTS

Total Variation between Individuals—To evaluate the extent of variation between individuals, platelet protein extracts of 20 healthy volunteers were covalently labeled by fluorescent tags with either Cy3 or Cy5 (see Table I). These two dyes are commonly used in DIGE studies to investigate two samples concomitantly on a single gel together with a Cy2-labeled internal standard. To mimic that approach and to include eventual dye-specific variances it was decided to use both dyes in the present study. Each sample was mixed with a Cy2-labeled internal standard and separated on 40 2D gels: 20 in the pl range 4–7 and 20 in the pl range 6–9. Typical gels are shown in Fig. 1 (A and B). About 1500 spots were detected in each gel. Because the internal standard was always the same protein mixture it was expected that all gels would show the same set of spots in the Cy2 channel. However, every gel contained several spots that were seen only on a few other gels. These spots were mainly low abundance and close to the detection limit. Although this set of spots might also represent interesting proteins it was decided to focus on reproducible spots because this allows meaningful and rep-

resentative statistic analysis regarding spot variation (a detailed analysis of incidence and relevance of poorly reproducible spots would go beyond the scope of the present work). Only spots that were present in the internal standard in at least 18 of 20 gels (90%) were included in the present analysis (484 protein spots in the acidic pl range and 424 protein spots in the alkaline pl range). The abundance of these spots varied from a V_{norm_g} of 0.37 to 229 (median 2.90) for acidic proteins and from 0.37 to 380 (median, 3.76) for alkaline proteins (see Fig. 2, A and B). The great majority of spots showed a V_{norm_g} below 50. For intergel normalization the V_{norm_g} values of the Cy3 and Cy5 channels were divided by their corresponding Cy2 signal on the same gel resulting in a “standardized abundance” (SA) value for each spot (see Fig. 2, C and D). Then the average of the SA over the 20 subjects and its cv were calculated. To calculate SA and cv values it was necessary to export the raw V_{norm_g} data from DeCyder. The software package itself gives only the average SA values. Fig. 3 (A and C) shows that the cv values are distributed over a broad range. Seven acidic and three alkaline spots even have a cv above 1. On the other hand, 95% of acidic spots have a cv below 0.60, and 95% of alkaline spots have a cv below 0.50. The median cv is 0.182 and 0.186, respectively, for acidic and alkaline spots. Thus, a typical platelet protein spot varies between different individuals by about 18%. The cv values show neither a dependence on pl (Fig. 4, A and B) nor a dependence on molecular weight (Fig. 4, C and D) as underlined by the low Spearman rank correlation coefficients.

A number of spots that showed an abundance (V_{norm_g}) in the range of the median spot volume were selected for protein identification by MALDI mass spectrometry. This abundance

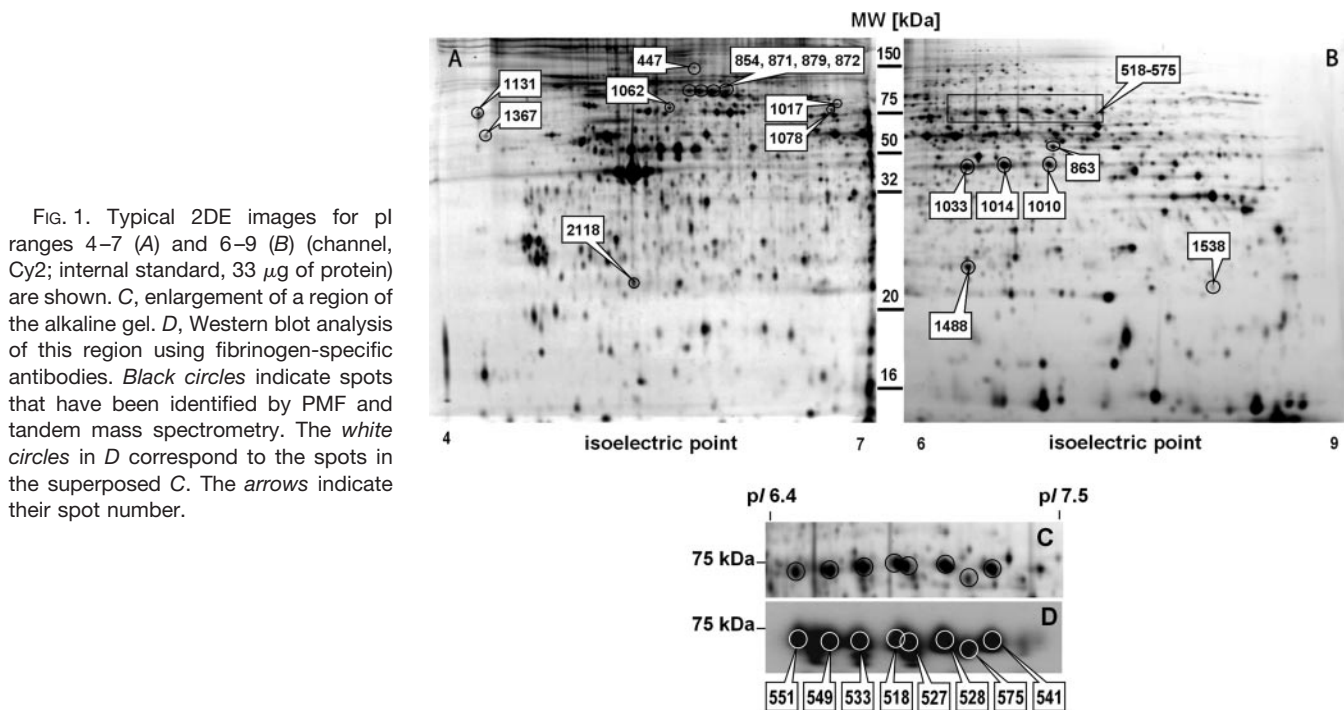


FIG. 1. Typical 2DE images for pI ranges 4–7 (A) and 6–9 (B) (channel, Cy2; internal standard, 33 μ g of protein) are shown. C, enlargement of a region of the alkaline gel. D, Western blot analysis of this region using fibrinogen-specific antibodies. Black circles indicate spots that have been identified by PMF and tandem mass spectrometry. The white circles in D correspond to the spots in the superposed C. The arrows indicate their spot number.

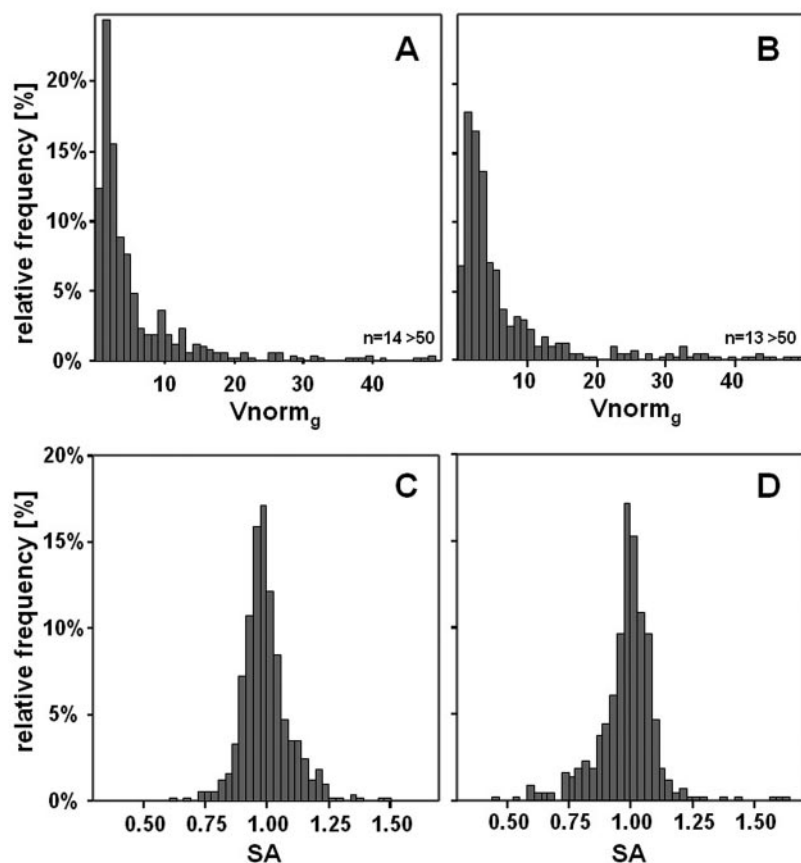


FIG. 2. Frequency distribution of mean spot volume over the 20 gels (expressed as $V_{norm,g}$; A and B) and of the mean SA (C and D) for the pI-ranges 4–7 (A and C; $n = 484$ spots) and 6–9 (B and D; $n = 424$ spots).

range reflects the great majority of spots. The selected spots were well distributed over the gels and included spots with low, intermediate, and high cv. Table II lists the complete set

of data for the identification of the selected protein spots based on peptide mass fingerprint and MS/MS experiments as well as their total and technical variation. Fibrinogen α

FIG. 3. Frequency distribution of the cv showing the variability for the pI-ranges 4–7 (A, C, and E) and 6–9 (B, D, and F). A and B show the total variation of samples from 20 different individuals. C and D show the technical variation after separation of the very same platelet extract on three different gels. E and F show the technical variation after preparation and electrophoretic separation of platelet proteins from six blood samples taken from the same donor and separated on six different gels. Not all spots found in the total variation experiments could be found in the technical variation experiments: A, n = 484; B, n = 424; C, n = 469; D, n = 412; E, n = 313; F, n = 411.

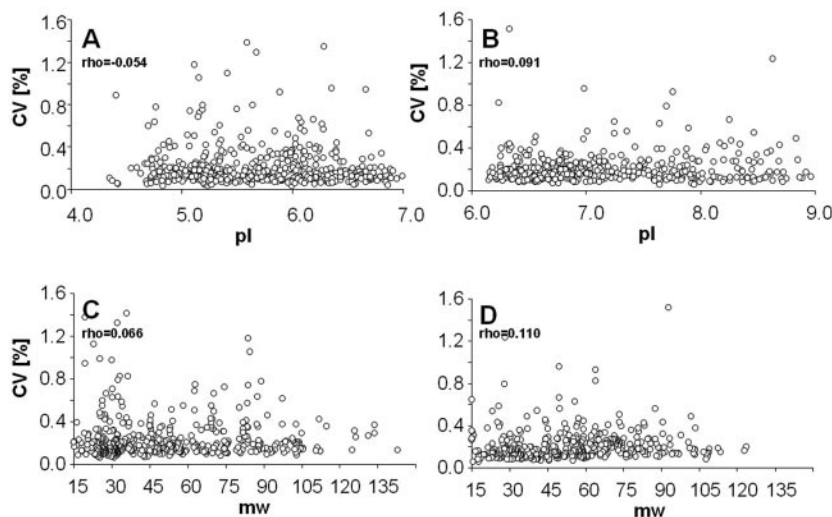
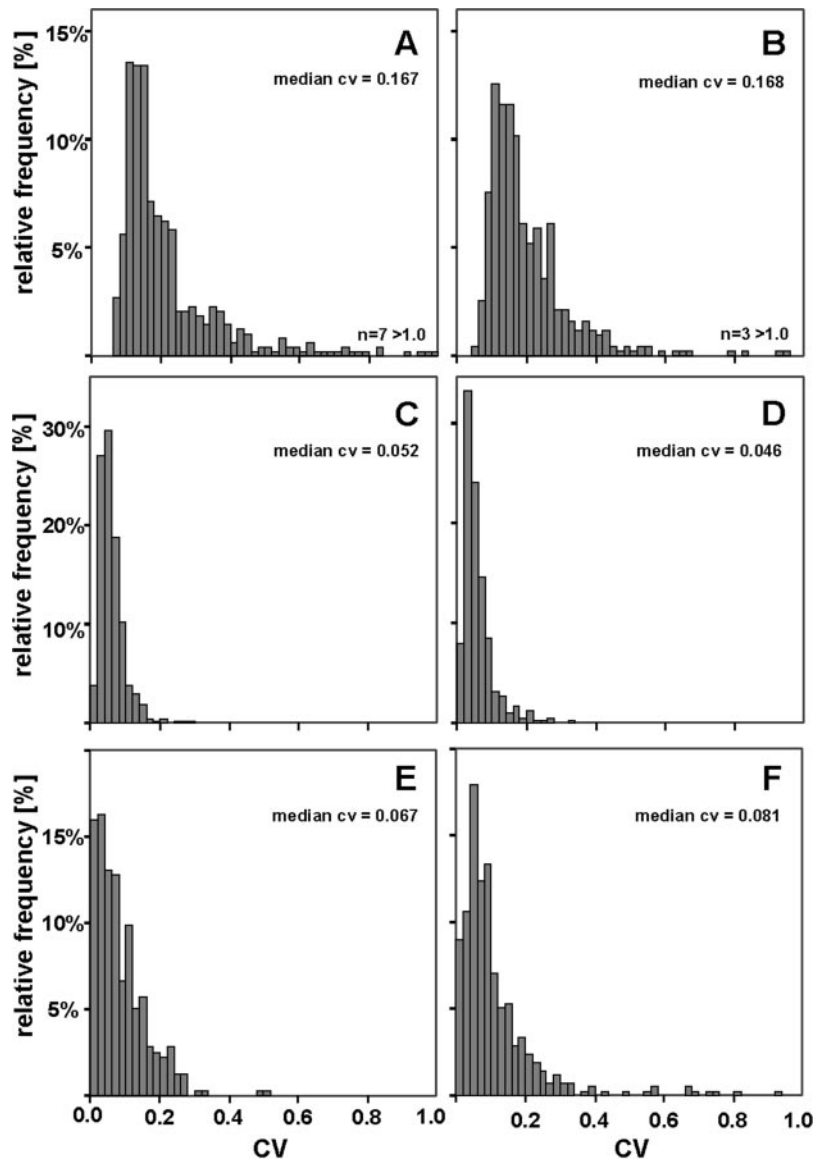


FIG. 4. Scatter plots to visualize the independence of the cv from pI (A and B) and molecular weight (C and D). Spearman rank correlation coefficients (ρ) were calculated to quantify the degree of dependence.

TABLE II
List of protein identities

Spot ID ^a	Vol ^b	cv total/tech	Protein name	Acc. no. ^c	PMF		MS/MS		
					Seq. cov.	Peptides ^d	Peptides ^e	Score ^f	
%									
Cytoskeleton									
872 _a	12.1	0.37/0.04	Gelsolin ^{g,h,i} and coagulation Factor XIIIa	P06396, P00488	38, 20	22/19, 16/19	119/45, 48/45	HVWPNEVWQR, EVQGFESATFLGYFK HVYGGELDVQIQR, AVPPNINSNAEEDDLPTVELQGVWPR	146, 228
854 _a	12.5	0.16/0.15	Gelsolin ^g	P06396	25	17/3	211/46	HVWPNEVWQR, IEGSNKVPDPATYGGQFVGGDSYIILYNYR	300
871 _a	41.8	0.13/0.07	Gelsolin ^g	P06396	27	23/2	276	HVWPNEVWQR, EVQGFESATFLGYFK	190
879 _a	60.3	0.12/0.01	Gelsolin ^g	P06396	44	33/19	304/40	EVQGFESATFLGYFK, IEGSNKVPDPATYGGQFVGGDSYIILYNYR	197
Metabolism									
1488 _b	22.0	0.09/0.02	Triose-phosphate isomerase ^g	P60174	68	15/10	199/41	SNVSDAVAQSTR, VPADTEVVCAPPTAYIDFAR	227
1538 _b	2.9	0.17/0.11	Adenylate kinase ^g	P00568	33	7/12	80/54	GFLIDGYPR, ATEPVIAFYEKR, YGYTHLSTGDLR	227
1078 _a	2.3	0.30/0.05	Glycerol-1-phosphate DH	P43304	34	25/15	195/40	MNLAIALTAAR, LAFLNVQAAEEALPR	168
Cell adhesion and platelet aggregation									
541 _b	34.7	0.27/0.07	Fibrinogen α chain ^{g,i,j}	P02671	37	37/15	253/42	SQLQKVPPEWK, HRRHPDEAAFFDTASTGK	139
518 _b	27.0	0.26/0.03	Fibrinogen α chain ^{g,i,j}	P02671	Identified by Western blot				
527 _b	30.8	0.22/0.03	Fibrinogen α chain ^{g,i,j}	P02671	Identified by Western blot				
528 _b	47.8	0.28/0.05	Fibrinogen α chain ^{g,i,j}	P02671	28	26/13	210/48	GDFSSANNRDNTYNR, HRRHPDEAAFFDTASTGK, QFTSSTSYNRGDSSTFESK	89
533 _b	33.5	0.22/0.04	Fibrinogen α chain ^{g,i,j}	P02671	22	22/21	149/43	NPSSAGSWNSGSSGPGSTGNR, GGSTSYGTGSETESPR	122
549 _b	30.5	0.25/0.02	Fibrinogen α chain ^{g,i,j}	P02671	Identified by Western blot				
551 _b	17.7	0.27/0.02	Fibrinogen α chain ^{g,i,j}	P02671	24	22/19	149/55	GDFSSANNRDNTYNR, QFTSSTSYNRGDSSTFESK	115
575 _b	8.7	0.21/0.03	Fibrinogen α chain ^{g,i,j}	P02671	Identified by Western blot				
863 _b	15.6	0.23/0.19	Integrin-linked protein kinase 1 ^{g,i,j}	Q13418	50	22/21	191/44	FSFQCFGR, Acetyl-M _{ox} DDIFTQCF, SVM _{ox} IDEDM _{ox} TAR	142
447 _a	1.6	0.36/0.01	Vinculin ^g	P18206	25	28/7	246/35	AIPDLTAPVAAVQAAVSNLVR, EAFQPEPDPFPPPPDLEQL	152
Other functions									
2118 _a	8.1	1.13/0.02	Apolipoprotein A-I ^g	P02647	39	13/12	120/57	THLAPYSDELRL, QKVEPLRAELQEGAR	105
1131 _a	10.1	0.11/0.08	Calreticulin ^g	P27797	37	15/15	146/46	NVLINKDIR, EPAVYFKEQFLDGGWTSR	212
1017 _a	1.1	0.20/0.03	Dynamin 1-like protein	O00429	31	25/20	203/52	SSVLESVGR, ALQASQIAEIR, GHAVNLLDVPVVAR	185
1010 _b	49.1	0.37/0.12	Pleckstrin ^g	P08567	18	10/14	86/50	GSTLTSPCQDFGKR, QQDHFQAAFLLEER	185
1014 _b	85.0	0.20/0.05	Pleckstrin ^g	P08567	22	6/5	65/55	IREGYLVKK, QQDHFQAAFLLEER	100
1033 _b	49.1	0.45/0.12	Pleckstrin ^g	P08567	43	15/15	126/41	ERTEWIK, QQDHFQAAFLLEER	152
1062 _a	4.2	0.08/0.06	GRP 75, mortalin ^g	P38646	25	19/11	155/45	DAGQISGLNVLRL, AQFEGIVDLIR, LLGQFTLIGIPPAPR	252
1367 _a	4.4	0.08/0.06	Nucleosome assembly protein 1-like 1 ^{g,i,j}	P55209	26	9/14	83/49	FYEEVHDLER, YAVLYQPLFDKR, KYAVLYQPLFDKR	159

^a a, acidic gels; b, alkaline gels.

^b Expressed as Vnorm^g.

^c Accession number from Swiss-Prot database.

^d Number of m/z values matched from an *in silico* digest vs. number of m/z values unmatched.

^e Mascot probability-based MOWSE (molecular weight search) score calculated for PMF and MS/MS results: Score = $-10 \times \log_{10} p$ where p is the absolute probability that the given hit is a random event. Significance ($p < 0.05$) is reached at scores >54 (PMF) and >28 (MS/MS) when searching the Swiss-Prot protein database; the second number is the score of the second best PMF hit of a non-homologous protein.

^f Peptides analyzed by PSD and high energy CID experiments; charge state, always +1.

^g Position on gel in accordance with that published by Garcia et al. (4).

^h Although in the quantification analysis spot 872 is clearly separated from the nearby focused Factor XIIIa in all gel images, it was not possible to excise only gelsolin.

ⁱ The number of missed cleavages was set to 2 for this protein.

^j Protein identification by Western blotting (see Fig. 1D).

chain was found in a horizontal chain of different spots (Fig. 1B, box). To be sure to identify all variants of fibrinogen α chain a Western blot analysis was performed (Fig. 1, C and D).

Contribution of Technical Variation to the Total Variation—To estimate the portion of biological variation contributing to total variation between samples from different individuals it is essential to know the technical variation. First we determined the technical variation of each spot due to the electrophoretic process. Therefore we labeled platelet proteins from the very same platelet sample with Cy2, Cy3, and Cy5. These extracts were mixed and separated on six different gels: three gels in the pI range 4–7 and three gels in the pI range 6–9. Only spots that were included in the experiment described above were quantified on each gel. This resulted in two SA values for each spot on each gel: one for Cy3 and one for Cy5. Because we analyzed the same sample in all three fluorescence channels these ratios should be theoretically always 1. Any observed deviation from each other was defined as technical variation of the electrophoretic process, which consists of fluorescence staining, 2D electrophoretic separation, and image acquisition and processing. Any potential differential labeling of Cy3 and Cy5 would also be included. To quantify this technical variation the cv of the six SA values was calculated for each spot. The great majority of spots have a very low variation (below 0.1; see Fig. 3, C and D). The median cv is 0.052 for acidic proteins and 0.046 for alkaline proteins. To include also the technical variation of the sample preparation process we analyzed the total variation of six different blood samples taken from the same healthy donor at the same time. After platelet protein preparation three samples were labeled with Cy3, and three samples were labeled with Cy5. Each sample was mixed with a Cy2-labeled internal standard from the first experiment and separated on an a separate gel. After electrophoretic separation of the mixture the same spots were evaluated quantitatively as in the first experiment. The cv values were only slightly higher in comparison with the previous experiment with median values of 0.067 for acidic spots and 0.081 for alkaline spots (see Fig. 3, E and F). Thus, the volume of a typical spot of the very same blood sample varied between different analyses by about 7%.

The variations observed between individuals described above are clearly higher than the variations observed in the analysis of multiple samples from a single subject. However, the distributions of cv values show an overlap. To estimate the proportion of the biological variation in the total variation measured between different individuals the two values were plotted against each other (Fig. 5, A and B). Spots lying close to the 45° line are those that showed in the analysis of different individuals a cv similar to those in the analysis of the same individual. However, the great majority of spots are below the 45° line (85% of the acidic spots and 87% of the alkaline spots). In addition, the very high variation between individuals measured for some spots is not associated with a high tech-

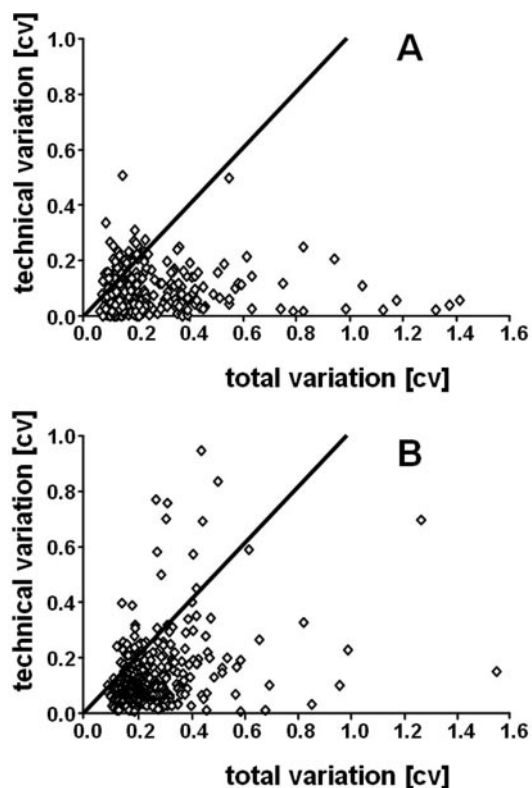


FIG. 5. Scatter plots of technical and total variation for pI ranges 4–7 (A) and 6–9 (B). The cv determined for the total variation has been plotted against the cv for technical variation. Spots below the 45° line show higher total variance than technical variance, which is true for the vast majority of all spots.

nical variation. This indicates that for most spots the technical variation has only a minor contribution to the total variation measured in different biological samples.

Implication for Biomarker Research—Many clinical biomarker studies found disease-associated expression changes (Δ) of biomarkers that are in a 1–2-fold range of its standard deviation (σ) (for a review, see Ref. 23). Similar Δ/σ ratios can also be observed in clinical routine diagnostic. This Δ/σ ratio is referred to as “effect size” (θ). In the DIGE setting used in this study, the standard deviation can be calculated from mean SA and cv by the following equation: $\sigma = cv \times \text{mean SA}$. Because the mean SA is always nearly 1, σ is almost identical to cv. Thus, an effect size (Δ/σ) of 1 of that platelet spot that showed the median cv (0.18) would correspond to an expression change by about 18%. Correspondingly for 50% of all spots the expression change would be lower. However, many spots show a cv much higher than the median value. To include 95% of all spots ($cv \leq 0.50$) an expression change of 50% has to be assumed for an effect size of 1.

To minimize the number of false negatives in a biomarker study (which means to reach a high statistical power) it is necessary to include a sufficient number of unique samples (referred to as “sample size”). However, the statistical power of an observed effect size depends also on the number of

spots included in the evaluation due to the problem of multiple testing. In the current study each gel showed nearly 1500 spots. However, not even 500 spots (484 acidic and 424 alkaline spots) were present in at least 90% of all gels. In biomarker research only a small part of these spots is expected to show disease-related changes. We calculated the statistical power per single hypothesis as a function of the sample size for varying effect sizes assumed to be common for all effective proteins. Correction for multiplicity was done according to the method of Benjamini and Hochberg (21). An FDR of 5% for the set of two-sided two-sample tests was chosen. To characterize the influence of the number of spots that are included in the analysis we performed these calculations for 500 spots and for 1500 spots. In addition, we assumed that either (i) 10 spots or (ii) 20 spots are effective. The results are shown in Fig. 6 (for 500 spots (A) and for 1500 spots (B)). When 500 spots are analyzed, an inclusion of 10 patients and 10 controls in a study would allow the detection of an effect size of 2.0 with a power of about 80%. However, an effect size of 1.25 would have only a power below 10%, which is certainly too low. To achieve a sufficient power (e.g. 80%) for an effect size of 1.25 a sample size of 21 per group would be needed. An assumption of 10 effective protein spots instead of 20 (i.e. 10 up- or down-regulated proteins are necessary to function as a biomarker set instead of 20) would increase the sample size to 25 (dashed curve). A further increase can be seen when 1500 spots are analyzed instead of 500. In that case the sample size increases to 26 (for 20 effective spots) and 29 (for 10 effective spots), respectively. These data demonstrate that a good guess of the proportion of effective marker, of the total variation of samples between different individuals, and of the effect size is necessary for an appropriate study design.

DISCUSSION

In this study, we assessed the biological variation of about 900 proteins (484 acidic and 424 alkaline) of human platelets from elderly subjects. We determined the total and the technical variation in a 2D DIGE analysis and evaluated the implications on the design of future biomarker research studies.

We recruited healthy volunteers aged from 56 to 100 years because this group reflects better the patient population in the clinics than younger subjects. It is known that platelets of elderly subjects differ from those of young subjects in several aspects including lower aggregation thresholds to ADP and collagen, higher levels of β -thromboglobulin and platelet factor 4, decreased levels of prostacyclin PGI₂, and increased polyphosphoinositide content (for a review, see Ref. 24). These age-related alterations are not detectable to the same extent in every aged subject. In addition, the health status impairs progressively with age, and our group of volunteers certainly includes subjects with slight, non-apparent illnesses. Thus, the age group used in this study reflects a very heterogeneous population. The present analysis showed that de-

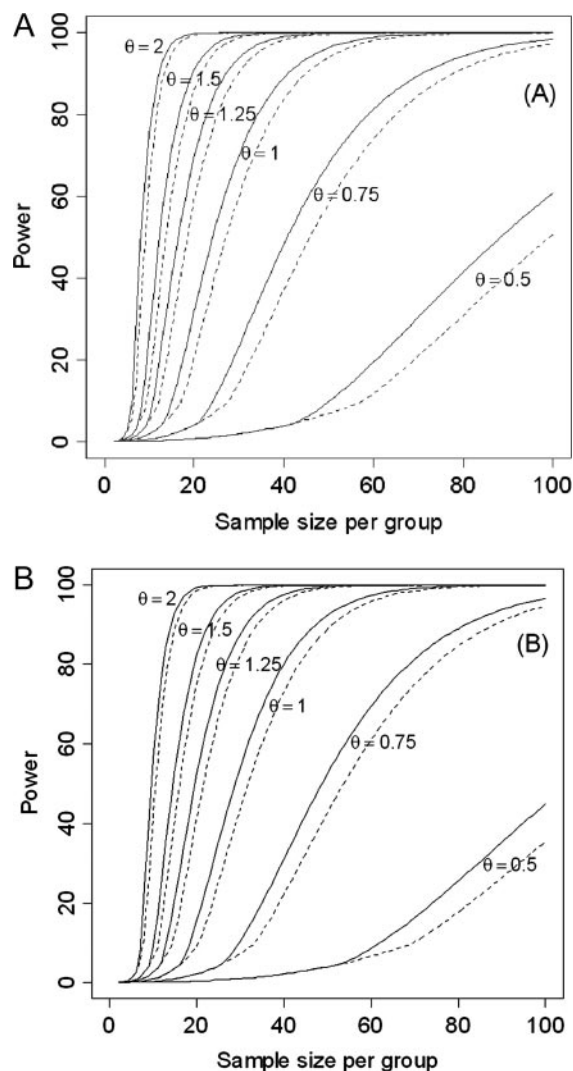


FIG. 6. Power per hypothesis of the two-sided two-sample *t* test depending on the sample size per group (starting with two observations per group) for different effect sizes $\theta = \Delta/\sigma$. An FDR of 5% targeted at 500 (A) and 1500 (B) proteins is investigated. It is assumed that 20 (solid curves) or 10 (dashed curves) markers are effective with a common effect size of $\theta (= \Delta/\sigma)$. Note that for small power values an approximation using the Bonferroni level was performed because the asymptotic FDR does not yield reliable results.

spite this heterogeneity most protein spots have a surprisingly low total variation between individuals. To date the knowledge about interindividual proteome variations in human tissues is very limited. Most proteomics studies were done with non-human samples. The total variation (cv) was found to be in murine primary cell lines 40–46% (technical variation, 23%) (25), in plant tissues 24% (technical variation, 16%) (26) and 40–60% (technical variation, 26%) (27), and in porcine mononuclear cells 34% (technical variation, 15%) (28). With a median total variation of only 18% our results are much lower. This might be explained by the DIGE approach for spot quantification used in our study. The DIGE analysis showed a much lower technical variation than the proteomics methods used in

the other studies (2DE without internal standards). Thus, the internal standard increases the statistical confidence of the analysis substantially. This is in accordance with recently published data from Friedman (29), who showed that the DIGE approach allows detection of very small quantitative changes. In addition, in our study the ratio between median total variation and median technical variation is higher than in the other studies (2.5 in comparison with 2 and lower). This further corroborates that the DIGE technology is very sensitive for quantitative variations. However, a few alkaline spots show a technical cv that is higher than the total cv (see Fig. 5B). Six spots show a technical cv above 0.6, whereas their total variation is below 0.5 (spot numbers 1478, 774, 421, 488, 803, and 783). A closer evaluation of these spots revealed that four of them (774, 488, 782, and 803) are spots that showed considerable overlap with other spots in some of the gels. This resulted in a false high spot volume in those gels. The reason for the high technical cv of the other two spots, however, remained unclear. Very recently a proteomics study was published that investigated the total variation in the human liver proteome using 2D DIGE (30). The authors investigated samples from 10 individuals and found a median cv value of 19% (range, 6.4–108.5%). This is in good accordance with our findings and supports our conclusion that the biological variation is rather low in human tissues from healthy subjects. Similarly low biological variation can be observed in the levels of human plasma proteins. The variation of these proteins in the healthy population is very well documented and is generally expressed as reference values (95% reference interval). This corresponds nearly to the median $\pm 2\sigma$ (95.4%) and therefore allows an approximation of the cv. On the basis of the reference values used in our hospital we calculated the cv of the 10 major abundant plasma proteins. This resulted in a median cv of 20% (range, 9–41%: albumin, 9%; IgG, 20%; transferrin, 14%; IgA, 38%; haptoglobin, 41%; complement C3c, 17%; IgM, 38%; α_2 -macroglobulin, 20%; α_1 -antitrypsin, 19%; α_1 -acid glycoprotein, 21%). However, it has to be taken into account that these proteins have been quantified by other methods than 2DE. Highly abundant proteins are supposed to fulfill housekeeping rather than regulatory functions, and this may be the reason for the low variability. However, this situation may change in case of illness. It was not within the scope of this study to analyze the variation experienced between samples that have been consecutively taken from the same individual. Due to standardized sampling conditions and sample preparation procedures, this type of variation is considered to be low and of constant height. This assumption is supported by findings of Nelsestuen *et al.* (31). They describe that the protein expression profile of each individual changes very little over time (weeks) and that the observed differences between individuals are significantly higher at any time point.

A number of spots were selected for identification by MALDI mass spectrometry. Vinculin (115 kDa, spot 447) is known to be involved in the attachment of the actin-based

microfilaments to the plasma membrane. Although it is typically expressed in skeletal muscle it was also found to be a part of the platelet proteome by others (4). Vinculin showed a cv (0.36) that is twice as high as the median value (0.18). A similar cv was found for a gelsolin isoform (acidic spot 872). Gelsolin is also an actin-binding protein. Surprisingly the other gelsolin isoforms (acidic spots 854, 871, and 879) showed much lower cv values (0.16, 0.13, and 0.12, respectively). This indicates that for gelsolin the biological variation is mainly determined on the level of posttranslational modification rather than on the level of the primary transcript. In contrast, the eight isoforms of fibrinogen α chain showed very similar variations (alkaline spots 518, 527, 528, 533, 541, 549, 551, and 575). Fibrinogen is a cofactor in platelet aggregation and plays an important role in clot formation. Fibrinogen α chain is known to be highly modified in platelets. The data indicate that the variation of the different fibrinogen α chain isoforms depends on a common precursor (*e.g.* mRNA) rather than on the isoform itself. GRP 75 (acidic spot 1078), nucleosome assembly protein 1-like 1 (acidic spot 1367), and triose-phosphate isomerase (alkaline spot 1488) all showed a very low variation, although they are structurally and functionally completely unrelated. In contrast, acidic spot 2118 varied extraordinarily between individuals. It corresponded to the apolipoprotein A-I (apoA-I). ApoA-I is primarily synthesized in the liver and secreted in plasma where it is present in huge amounts. In the platelet preparation protocol used here the cellular particles are separated from plasma proteins by gel filtration. This removes plasma proteins nearly completely as demonstrated by the fact that serum albumin can be found only in traces in our proteome. Therefore we suppose that apoA-I either binds specifically to platelets or is actively imported into platelets (especially megakaryocytes, their precursor cells, are known to import plasma proteins) or *de novo* synthesized in platelets. The last possibility is supported by the fact that a recent study found mRNAs coding for apoA-I in the platelet transcriptome (32). A further support comes from the fact that apoA-I was also found in the platelet proteome in two other studies (4, 33). However, we cannot exclude that platelet-associated apoA-I originates at least partially from plasma. The plasma level of apoA-I is known to be reduced during chronic inflammation. This might contribute to the high variability of this protein. The identified proteins are grouped in Table II according to their function. However, there was no apparent difference in the biological variation between these groups.

Disease-related expression changes observed in biomarker studies are mainly moderate. A recent meta-analysis of 752 genetic studies on different diseases including cardiovascular diseases, various cancers, schizophrenia, dementia, and diabetes revealed a median gene-disease association of $\theta = 1.43$ (range, 1.16–2.58) (23). Considering the rather small median cv of 18% measured here in platelets this would correspond to only a 1.26-fold change. This is far below the changes

reported in proteomics studies with cell culture systems. Thus, to find disease-related or indicative biomarkers in humans it is essential to have a highly sensitive detection system for quantitative measurements. In addition, it is necessary to include a sufficiently high number of samples in each group to get a representative overview of the disease-related protein changes. The sample size increases with the number of proteins that are investigated. But the proteins have to be present in all gels because proteins that are present only in a few gels do not reach the necessary sample size. Based on the results presented here (total variation of 18%, number of repeatedly found proteins of about 500 per gel) and on the assumption of a minimal effect size of about 1.25 (based on the literature) it can be estimated that a biomarker study in human platelets should include at least 20–25 samples per group, and the quantification method used should have a very low technical variation (presumably below 10%). Considering the indications that the total variation in other human tissues is similar to that in platelets, we propose that the conclusions raised here might be a good general guess for clinical biomarker studies.

Acknowledgment—We thank Ellen Umlauf of Randox Laboratories (Crumlin, UK) for technical support and helpful discussion.

* Parts of this work were supported by Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF) Fund P18698-n15. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

§ Both authors contributed equally to this work.

§§ To whom correspondence should be addressed: Dept. of Surgery, Medical University of Vienna, Waehringer-Guertel 18-20, A-1090 Vienna, Austria. Tel.: 43-1-40400-6979; Fax: 43-1-40400-6782; E-mail: rudolf.oehler@meduniwien.ac.at.

REFERENCES

- Jurk, K., and Kehrel, B. E. (2005) Platelets: physiology and biochemistry. *Semin. Thromb. Hemost.* **31**, 381–392
- Gnatenko, D. V., Dunn, J. J., McCorkle, S. R., Weissmann, D., Perrotta, P. L., and Bahou, W. F. (2003) Transcript profiling of human platelets using microarray and serial analysis of gene expression. *Blood* **101**, 2285–2293
- Marcus, K., Immler, D., Sternberger, J., and Meyer, H. E. (2000) Identification of platelet proteins separated by two-dimensional gel electrophoresis and analyzed by matrix assisted laser desorption/ionization-time of flight-mass spectrometry and detection of tyrosine-phosphorylated proteins. *Electrophoresis* **21**, 2622–2636
- Garcia, A., Prabhakar, S., Brock, C. J., Pearce, A. C., Dwek, R. A., Watson, S. P., Hebestreit, H. F., and Zitzmann, N. (2004) Extensive analysis of the human platelet proteome by two-dimensional gel electrophoresis and mass spectrometry. *Proteomics* **4**, 656–668
- Coppinger, J. A., Cagney, G., Toomey, S., Kislinger, T., Belton, O., McRedmond, J. P., Cahill, D. J., Emili, A., Fitzgerald, D. J., and Maguire, P. B. (2004) Characterization of the proteins released from activated platelets leads to localization of novel platelet proteins in human atherosclerotic lesions. *Blood* **103**, 2096–2104
- Maguire, P. B., Wynne, K. J., Harney, D. F., O'Donoghue, N. M., Stephens, G., and Fitzgerald, D. J. (2002) Identification of the phosphotyrosine proteome from thrombin activated platelets. *Proteomics* **2**, 642–648
- Marcus, K., Moebius, J., and Meyer, H. E. (2003) Differential analysis of phosphorylated proteins in resting and thrombin-stimulated human platelets. *Anal. Bioanal. Chem.* **376**, 973–993
- Unlu, M., Morgan, M. E., and Minden, J. S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071–2077
- Zellner, M., Winkler, W., Hayden, H., Diestinger, M., Eliassen, M., Gesslbauer, B., Miller, I., Chang, M., Kungl, A., Roth, E., and Oehler, R. (2005) Quantitative validation of different protein precipitation methods in proteome analysis of blood platelets. *Electrophoresis* **26**, 2481–2489
- Jensen, O. N., Wilm, M., Shevchenko, A., and Mann, M. (1999) Sample preparation methods for mass spectrometric peptide mapping directly from 2-DE gels. *Methods Mol. Biol.* **112**, 513–530
- Belgacem, O., Bowdler, A., Brookhouse, I., Brancia, F. L., and Raptakis, E. (2006) Dissociation of biomolecules using an ultraviolet matrix-assisted laser desorption/ionisation time-of-flight/curved field reflectron tandem mass spectrometer equipped with a differential-pumped collision cell. *Rapid Commun. Mass Spectrom.* **20**, 1653–1660
- Pittenauer, E., Zehl, M., Belgacem, O., Raptakis, E., Mistrik, R., and Allmaier, E. (2006) Comparison of CID spectra of singly charged polypeptide antibiotic precursor ions obtained by positive-ion vacuum MALDI IT/TOF and TOF/TOF, AP-MALDI-IT and ESI-IT mass spectrometry. *J. Mass Spectrom.* **41**, 421–447
- Vorm, O., Roepstorff, P., and Mann, M. (1994) Improved resolution and very high sensitivity in MALDI TOF of matrix surfaces made by fast evaporation. *Anal. Chem.* **66**, 3281–3287
- Mattow, J., Schmidt, F., Hohenwarter, W., Siejak, F., Schaible, U. E., and Kaufmann, S. H. (2004) Protein identification and tracking in two-dimensional electrophoretic gels by minimal protein identifiers. *Proteomics* **4**, 2927–2941
- Smirnov, I. P., Zhu, X., Taylor, T., Huang, Y., Ross, P., Papayanopoulos, I. A., Martin, S. A., and Pappin, D. J. (2004) Suppression of α -cyano-4-hydroxycinnamic acid matrix clusters and reduction of chemical noise in MALDI-TOF mass spectrometry. *Anal. Chem.* **76**, 2958–2965
- Zhang, W., and Chait, B. T. (2000) ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal. Chem.* **72**, 2482–2489
- Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567
- Bairoch, A., and Apweiler, R. (1997) The SWISS-PROT protein sequence database: its relevance to human molecular medical research. *J. Mol. Med.* **75**, 312–316
- Savitzky, A., and Golay, M. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680
- Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300
- Storey, J., Taylor, J., and Siegmund, D. (2004) Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. B* **66**, 187–205
- Ioannidis, J. P., Trikalinos, T. A., and Khoury, M. J. (2006) Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am. J. Epidemiol.* **164**, 609–614
- Wilkerson, W. R., and Sane, D. C. (2002) Aging and thrombosis. *Semin. Thromb. Hemost.* **28**, 555–568
- Molloy, M. P., Brzezinski, E. E., Hang, J., McDowell, M. T., and VanBogelen, R. A. (2003) Overcoming technical variation and biological variation in quantitative proteomics. *Proteomics* **3**, 1912–1919
- Asirvatham, V. S., Watson, B. S., and Sumner, L. W. (2002) Analytical and biological variances associated with proteomic studies of Medicago truncatula by two-dimensional polyacrylamide gel electrophoresis. *Proteomics* **2**, 960–968
- Jorge, I., Navarro, R. M., Lenz, C., Ariza, D., Porrás, C., and Jorrián, J. (2005) The holm oak leaf proteome: analytical and biological variability in the protein expression level assessed by 2-DE and protein identification

- tandem mass spectrometry de novo sequencing and sequence similarity searching. *Proteomics* **5**, 222–234
28. Ramirez-Boo, M., Garrido, J. J., Ogueta, S., Calvete, J. J., Gomez-Diaz, C., and Moreno, A. (2006) Analysis of porcine peripheral blood mononuclear cells proteome by 2-DE and MS: analytical and biological variability in the protein expression level and protein identification. *Proteomics* **6**, Suppl. 1, S215–S225
 29. Friedman, D. B. (2006) Quantitative proteomics for two-dimensional gels using difference gel electrophoresis. *Methods Mol. Biol.* **367**, 219–240
 30. Zhang, X., Guo, Y., Song, Y., Sun, W., Yu, C., Zhao, X., Wang, H., Jiang, H., Li, Y., Qian, X., Jiang, Y., and He, F. (2006) Proteomic analysis of individual variation in normal livers of human beings using difference gel electrophoresis. *Proteomics* **6**, 5260–5268
 31. Nelsestuen, G. L., Zhang, Y., Martinez, M. B., Key, N. S., Jilma, B., Verneris, M., Sinaiko, A., and Kasthuri, R. S. (2005) Plasma protein profiling: unique and stable features of individuals. *Proteomics* **5**, 4012–4024
 32. Bugert, P., and Kluter, H. (2006) Profiling of gene transcripts in human platelets: an update of the platelet transcriptome. *Platelets* **17**, 503–504
 33. Martens, L., Van Damme, P., Van Damme, J., Staes, A., Timmerman, E., Ghesquiere, B., Thomas, G. R., Vandekerckhove, J., and Gevaert, K. (2005) The human platelet proteome mapped by peptide-centric proteomics: a functional protein profile. *Proteomics* **5**, 3193–3204