

Supplementary Material

Integrated multiscale analysis reveals complex gender-specific changes in lymphocytes of smokers

Anne-Cécile Ribou, Florence Riera, Fabienne Durand, and Laurent Henry

Table S1: Raw hematological data.

Measurements of the number of white blood cells (WBC), lymphocytes (LYMP), monocytes (MONO), granulocytes (GR) and platelets (PLT), together with measures on erythron variables: number of red blood cells (RBC), hemoglobin (Hb), hematocrit (Hct), mean corpuscular volume (MCV), mean cell Hb (MCH), mean cell Hb concentration (MCHC), RBC distribution width (RDW), and WBC/PLT ratio (PRL) in 33 donors the day of the collection.

n°	Subgroups	WBC	LYMP	MONO	GR	RBC	Hb	Hct	MCV	MCH	MCHC	RDW	PLT	PLR
		$10^3/mm^3$	$10^2/mm^3$	$10^2/mm^3$	$10^2/mm^3$	$10^6/mm^3$	g/dl	%	fL	pg	g/dl	%	$10^3/mm^3$	%
don-1	4	10.6	22.5	7	70.5	4.66	16	44.5	95.6	34.3	35.9	12.5	362	160.9
don-2	3	6.6	35.6	5.9	58.5	5.19	16.1	44.2	85.2	31	36.4	12.9	160	44.9
don-3	4	5.4	19.7	4	76.3	5.06	16.7	45.3	89.5	33	36.9	12.1	192	97.5
don-4	3	5.1	29.8	8.8	61.4	4.5	14	38.9	86.5	31.1	36	12.5	226	75.8
don-5	1	7.2	40.2	8.1	51.7	4.06	12.8	36.3	89.5	31.5	35.2	11.9	186	46.3
don-6	1	5.2	19.9	7.6	72.5	3.97	13.5	37.6	94.8	34	35.9	12.2	263	132.2
don-7	4	7.6	33.3	7.2	59.5	4.49	14.8	41.1	91.6	33	36	13	273	82.0
don-8	4	4.9	37.8	5.6	56.6	4.28	12	34.3	80.2	28	35	15.4	231	61.1
don-9	3	3.9	35.5	7.2	57.3	4.4	13.5	37.3	84.7	30.7	36.2	12.3	195	54.9
don-10	4	6.1	31.4	5.7	62.9	4.94	16.3	45.6	92.4	33	35.7	12.6	154	49.0
don-11	3	5.8	37.4	6.9	55.7	4.36	14.7	41.2	94.5	33.7	35.7	12.5	237	63.4
don-12	3	4.2	38.2	6.1	55.7	4.12	13.6	37.5	91.1	33	36.2	12.5	182	47.6
don-13	3	4.1	27.9	6.6	65.5	4.54	13.7	37.9	83.4	30.2	36.2	12.4	184	65.9
don-14	2	6.5	25.4	5.5	69.1	5.05	12.6	48.2	95.4	25	26.2	12.9	279	109.8
don-15	4	6.4	32.3	5.7	62	4.83	14.7	42.4	87.7	30.4	34.7	13.4	211	65.3
don-16	1	5.3	28.1	5.7	66.2	4.21	12.7	36.2	86.1	30.2	35	14.2	233	82.9
don-17	4	4.7	33.9	6.4	59.7	4.34	14.1	38.8	89.5	32.5	36.3	12.5	222	65.5
don-18	1	5.1	24.5	7.4	68.1	3.98	12	33.7	84.7	30.2	35.6	12.6	227	92.7
don-19	1	5.8	31.4	4	64.6	4.19	12.5	35.4	84.5	29.8	35.3	13.6	175	55.7
don-20	1	5.6	29.3	5.3	65.4	4.34	13	36.7	84.5	30	35.4	12.5	204	69.6
don-21	4	10.9	23.5	5.6	70.9	4.7	15.5	43.1	91.7	33	36	13.2	234	99.6
don-22	4	9.2	32.4	3.6	64	4.37	14.6	39.8	91.1	33.4	36.7	13.1	201	62.0
don-23	4	7	32.1	7.3	60.6	4.33	14.2	39.9	92.1	32.8	35.6	13.3	247	76.9
don-24	3	4.3	39	5.8	55.2	4.61	15.2	43.8	95.1	33	34.7	12	221	56.7
don-25	4	13.3	19.3	2.47	78.3	4.74	15.9	44.5	93.9	33.5	35.7	14.3	221	114.5
don-26	1	5.7	21.5	6.4	72.1	3.98	12.5	37.2	93.4	31.4	33.6	12.3	205	95.3
don-27	3	3.8	37.9	8.7	53.4	4.94	13.9	46	93.2	28.1	30.2	12.5	240	63.3
don-28	3	4.7	42.6	6	51.4	4.81	15.2	45.9	95.4	31.6	33.1	12.2	208	48.8
don-29	2	8.6	24.8	5.6	69.6	4.38	12.9	39.3	89.8	29.5	32.8	13.9	162	65.3
don-30	1	5.3	43.8	5.2	51	4.33	12.9	38.5	89	29.8	33.5	11.6	276	63.0
don-31	3	4.6	37.2	2.1	60.7	4.42	14.1	41.2	93.3	31.9	34.2	11.5	197	53.0
don-32	2	9.2	26.7	4.9	68.4	4.19	13.5	39.9	95.3	32.2	38	12.8	218	81.6
don-33	2	11.3	23.2	5.9	70.9	4.63	13.5	40.8	88.2	29.2	33.1	13.9	262	112.9

Table S2: Proteasome parameters.

The amount of proteasomal antigens, directed against the proteasomal proteins subunits p21K, p23K, p25K, p27K, p29K, p30-33K and p31K, are calculated by flow cytometry. Mean fluorescence intensity by cells (MFI) and percent positive cell (% cell+) of at least three experiments.

n°	Subgroups	Number of cells	MFlp21	MFlp23	MFlp25	MFlp27	MFlp29	MFlp3033	MFlp31	%cell+p21	%cell+p23	%cell+p25	%cell+p27	%cell+p29	%cell+p3033	%cell+31
don-1	4	3560	112	802	162	74	86	133	90	83.3	89.6	92.3	82.2	85.9	90.4	86.9
don-2	3	2250	107	505	63	48	70	54	45	62.6	58.4	36.4	60.1	54.1	52.4	57.2
don-3	4	2980	86	379	73	48	92	74	46	77.8	72.8	46.1	69.8	73.9	57	63.4
don-4	3	1970	84	467	64	37	61	36	51	79.1	87.8	72.8	75.3	79	61.5	72.5
don-5	1	4020	66	215	54	50	44	53	50	56.4	60.8	35.5	47.5	43.3	40.1	31.2
don-6	1	1990	114	471	62	83	78	44	62	64.2	59.2	40.2	71.1	51.1	52.4	55.1
don-7	4	3330	41	166	39	41	39	34	29	32.2	45.3	31.9	50	40.3	37.6	38.2
don-8	4	3780	76	228	49	44	44	38	57	34.5	41.1	32.9	39.7	29.8	41.5	38.6
don-9	3	3550	74	375	45	55	51	44	60	75.3	70.3	62.5	67.8	69.7	56.2	74.6
don-10	4	3140	96	683	62	61	68	42	77	64.5	69.8	41.7	64.4	54.2	52.1	73.7
don-11	3	3740	117	743	116	95	93	86	67	51.8	56.5	35.5	47.6	40.4	34.8	33.1
don-12	3	3820	32	68	27	27	28	27	29	56.4	75.2	78.0	72.5	74.7	71	71.9
don-13	3	2790	47	210	32	37	39	39	59	41.8	48.7	24.5	50.0	39.5	29.2	37.2
don-14	2	2540	50	89	25	39	26	19	32	51.8	72.9	37.5	64	43	29.9	50.7
don-15	4	3230	45	103	38	38	42	36	50	36.8	35.5	25.8	50.4	33.8	22.7	54.3
don-16	1	2810	61	175	37	49	55	54	94	41.2	40.8	40.6	41.1	43.8	41.2	53.4
don-17	4	3390	44	156	42	67	36	58	29	27.6	39.9	23.7	28.3	26.6	30.9	31.1
don-18	1	2450	96	245	47	71	57	41	64	46.6	45.5	45.2	36.5	36.1	37.9	35.7
don-19	1	3140	61	596	87	96	101	89	126	48.9	48.3	32.3	39.9	13.1	42.1	37.0
don-20	1	2930	147	298	112	116	117	119	133	41.8	47.8	23.5	39.5	46.0	50.6	40.7
don-21	4	2350	86	374	38	65	74	79	62	45.2	48.5	37.9	42.3	60.3	40.7	44.4
don-22	4	3240	133	200	75	181	131	100	107	48.6	45.8	32.7	66.5	40.5	45.7	42.5
don-23	4	3210	44	67	43	27	48	32	49	38.9	68.2	33.4	20.6	42.2	47.2	27.6
don-24	3	3900	43	50	28	29	41	39	45	26.7	30.1	23.8	21.6	28.7	38.2	25.8
don-25	4	1930	33	41	41	94	17	31	63	33.5	61.7	25.6	24.7	25.1	47.5	25.9
don-26	1	2150	395	789	65	147	179	181	154	54.5	65.9	56.0	68.8	64.8	67.7	60.4
don-27	3	3790	425	1517	207	217	228	325	269	36.4	44.2	36.9	31.9	40.0	48.5	43.8
don-28	3	4260	255	498	144	153	120	142	151	32.7	40.5	37.3	30.4	33.2	27.8	37.8
don-29	2	2480	59	389	105	65	115	92	110	79.8	45.0	31.4	41.5	40.8	52.9	32.9
don-30	1	4380	147	141	107	137	87	92	94	33.8	46.0	32.7	45.0	32.3	36.2	32.1
don-31	3	3720	64	129	41	118	112	135	99	46.3	44.7	39.7	37.9	48.3	42.6	43.0
don-32	2	2670	140	270	48	136	82	85	77	34.3	51.5	33.7	45.2	43.5	30.9	36.0
don-33	2	2320	94	117	53	131	96	52	96	51.6	67.6	39.8	66.2	58.9	68.2	48.0

Table S3: Cells imaging of lymphocytes.

First peak of the distribution (FI-1_{max}), mean of the total fluorescence intensity (MFI), and the standard deviation (SD) of fluorescent markers for the plasma membranes (Nil Red, NR) and the mitochondria (rhodamine 123, R123) give functional information. The illuminated pixel, related to the morphological parameters, of fluorescent markers staining the cell membranes (NR) and staining the nucleus (Hoechst 33342, Ho342) allows to calculate the FI-1_{max}, mean and SD of cell and nucleus surface, respectively. We also calculated these three parameters (FI-1_{max}, mean and SD) for the ratio cell area/nucleus area, Cell/Nuc. To consider the heterogeneity of the population, two parameters were calculated: the percentage of large cells (% Cell > 350 pixels), the percentage of cell with a small nucleus relative the cell surface (% Cell/Nuc > 3). The analyses were performed one day after blood collection.

n°	Subgroups	Number of cells	FI-1 _{max} NR	MFI NR	SD NR	FI-1 _{max} R123	MFI R123	SD R123	Cell-1 _{max}	Mean Cell	SD Cell	Cell/Nuc-1 _{max}	Mean Cell/Nuc	SD Cell/Nuc	% Cell>350	% Cell/Nuc>3
don-1	4	65	12500	16101	8677	1125	1952	1518	282.5	325	121	1.5	2.36	1.14	0.369	0.277
don-2	3	148	12750	17638	13797	1003	1298	1237	237.5	295	141	1.625	1.82	0.99	0.176	0.122
don-3	4	225	11250	17530	9189	1125	1293	648	225	243	69	1.25	1.61	0.67	0.107	0.071
don-4	3	113	11250	17635	13381	1080	1426	1038	225	263	89	1.5	1.375	0.92	0.195	0.053
don-5	1	141	11250	12736	5647	750	699	413	237.5	228	57	1.375	1.47	0.43	0.05	0.021
don-6	1	178	12500	23763	15373	875	1513	1125	187.5	233	82	1.25	1.45	0.47	0.202	0.017
don-7	4	112	8750	11952	6980	750	1056	939	182.5	205	55	1.25	1.17	0.3	0.08	0.009
don-8	4	209	11250	12600	6288	875	967	430	200	203	58	1.25	1.34	0.45	0.019	0.01
don-9	3	84	12500	14462	5010	875	1045	370	225	236	48	1.25	1.29	0.32	0.024	0.000
don-10	4	180	13750	15030	6137	625	871	663	250	249	65	1.375	1.49	0.47	0.044	0.033
don-11	3	81	15000	21456	12788	1375	1720	1041	225	273	77	1.25	1.45	0.45	0.198	0.012
don-12	3	198	13750	14989	6041	625	629	352	212.5	230	48	1.25	1.39	0.35	0.015	0.000
don-13	3	262	18750	20642	7173	2125	3933	3342	237.5	250	64	1.5	1.58	0.47	0.092	0.019
don-14	2	239	15000	17489	6744	1000	1108	602	237.5	248	57	1.375	1.56	0.41	0.079	0.008
don-15	4	137	21250	24963	9487	3125	3317	1128	237.5	251	58	1.25	1.38	0.41	0.051	0.022
don-16	1	151	16250	21081	10956	1250	1645	1076	275	264	72	1.375	1.5	0.38	0.073	0.007
don-17	4	134	13750	16873	8439	1175	1406	995	212.5	227	54	1.25	1.29	0.32	0.045	0.007
don-18	1	92	16250	19355	9470	1125	1795	1359	225	233	47	1.25	1.26	0.28	0.054	0.000
don-19	1	184	11250	22856	16539	1375	1712	1154	212.5	255	82	1.375	1.76	0.81	0.185	0.152
don-20	1	172	16250	19827	10494	1875	2620	2805	225	279	201	1.25	1.55	1.00	0.076	0.047
don-21	4	95	37500	35049	12562	2250	2667	1145	312.5	304	65	1.375	1.98	0.81	0.347	0.179
don-22	4	73	12500	15349	9682	1500	2206	3180	225	257	191	1.25	1.43	1.41	0.027	0.027
don-23	4	227	20000	24647	11523	2125	2541	1575	225	241	64	1.375	1.46	0.42	0.066	0.009
don-24	3	158	21250	22940	7939	2625	2834	1338	250	263	61	1.375	1.45	0.45	0.101	0.013
don-25	4	154	21250	30045	14749	1750	2494	1322	225	266	74	1.375	1.46	0.39	0.136	0.013
don-26	1	180	26260	41755	25383	1125	4048	3272	225	288	106	1.5	1.72	0.59	0.217	0.083
don-27	3	262	13750	17557	8892	1125	1518	899	225	223	61.6	1.375	1.59	0.59	0.042	0.042
don-28	3	278	16250	19025	6061	2125	3246	3080	225	226	39	1.375	1.47	0.32	0.007	0.004
don-29	2	276	21250	24654	8802	2875	3298	1299	225	223	43	1.375	1.52	0.38	0.007	0.007
don-30	1	210	12500	13663	5302	1500	1523	612	237.5	236	45	1.5	1.53	0.34	0.019	0.01
don-31	3	213	17500	20079	5456	1875	2142	785	212.5	221	39	1.125	1.22	0.26	0.019	0.000
don-32	2	262	16250	17700	5828	1750	2230	932	237.5	241	45	1.375	1.44	0.38	0.038	0.011
don-33	2	334	11250	14466	6668	1125	1409	784	187.5	221	51	1.375	1.45	0.38	0.027	0.009

Cell imaging: the triple labelling protocols

To record parameters related to lymphocytes morphology (cell and nuclear sizes) and functional information (level of mitochondria energetic state and amount, and properties of the plasma membrane), videomicrofluorometry associated to multiple labelling has been used.

1. Apparatus

The fluorescence digital imaging microscopy system (Salmon, 1992; Vigo, 1991) consists of an inverted fluorescence microscope (IMT2, Olympus, Rungis, France) equipped with an epi-illuminator, a 40X objective (NA 0.75) (Leitz, Rueil-Malmaison) and a Silicon Intensified Target (SIT) camera from Lhesa (Saint Quentin, France) coupled to a Matrox MVP-AT digitizing card (Matrox, Dorval, Canada). The mercury lines of a high-pressure mercury lamp (100 W) at 365 nm and 435 nm were selected by dichroic mirrors to selectively excite Ho342 and R123 or NR. A computer-controlled filter holder automatically stepped a filter depending on the specific fluorescence of the dyes: no filter for Ho342;

an interference filter $\lambda_{\text{max}} = 525 \text{ nm}$, half bandwidth = 10 nm for R123; and an interference filter $\lambda_{\text{max}} = 610 \text{ nm}$, half bandwidth = 20 nm for NR. A specific data acquisition program, which included the background subtraction, the correction of non-linearity of the digitizer and the pixel-to-pixel heterogeneity of the camera gain, allowed quantitative measurements of the fluorescence intensities of the dyes.

2. Cell labelling

A stock solution of Ho342 (1 mM) from Aldrich (Saint Quentin Fallavier, France) and R123 (0,1 mM) from Sigma (Saint Quentin Fallavier, France) were prepared in phosphate buffer saline (PBS) from ICN (Orsay, France) and kept at -20°C while NR (1 mM) from Sigma (Saint Quentin Fallavier, France) was prepared in ethanol from Merck (Nogent-sur-Marne, France) and stored at 4°C . Ten μl of Ho342 were added to 1 ml (1 000 lymphocytes) of cells suspended and incubated for 2 h at 37°C . After the first hour of incubation, 115 μl of R123 were added to the cell suspension. The cells were then incubated for 57 mn simultaneously with the two probes before 15 μl of NR were added. After the remaining 3 mn, the cell suspension was centrifuged and rinsed three times with 1 ml of Hank's balanced salt solution (HBSS) from ICN (Orsay, France). The cells were again suspended in 1 ml of HBSS and from this 300 μl were plated in Sykes Moore chambers for observation and numerical image multiparametric analysis.

3. Numerical image analysis

A protocol was developed to analyze the fluorescence of the three probes: Ho342-stained nuclei, NR-stained plasma membrane and R123-stained mitochondria. The NR fluorescence automatically provided the cell outline from an image segmentation using thresholding, followed by a cell contour smoothing from opening and closing sequences. NR fluorescence image of the cells was required as it allows the determination of the cell size from plasma membrane staining as well as determination of parameters related to the shape of the cells. Such parameters were necessary to exclude the data related to subcellular debris or cell clumps and gave an estimation of cell integrity. Because R123 fluorescence was heterogeneous as it is distributed in clusters, no segmentation was possible. The mask of each cell was then generated from NR fluorescence image and superimposed on the R123 fluorescence image of the corresponding cell, in order to delineate the cell contour in which the fluorescence of R123 was quantified after subtraction of the background corresponding to the fluorescence of the medium. Thresholding and contour smoothing by closing and opening sequences isolated the corresponding Ho342-stained nucleus. The fluorescence intensity parameters were normalized, using a reference standard, to compensate for mercury lamp fluctuations.

4. Parameters recording

The numerical image multiparametric analysis allows to evaluate cell and nucleus area and perimeter, shape factors, and fluorescence intensity parameters (total, mean and standard error) for Ho342, R123 and NR on about 200 cells. Then the properties of the lymphocyte populations were analyzed through cell population distribution histograms for each of the selected parameter. For each histogram we took the value of the maximum of the distribution ($FI-1_{\text{max}}$), the mean value (MTI) and the associated standard deviation (SD) in order to account for the distribution pattern as the lymphocyte population is heterogeneous. Due to their symmetry coefficient, not showing any significant variation, the nuclear labeling (Ho342 intensity and surface) were removed from this study.

Figure S1. Box plots of the 41 parameters. Donors were separated by gender and smoking status and compared with the total population. SM: smoking male, SF: smoking female, nSM: non-smoking male, nSF: non-smoking female. Cross (+): mean, ●: minimum/maximum. Marker abbreviations as shown in Table 2.

A. hematological data



B. Proteasome parameters.

C. Cells imaging of lymphocytes.

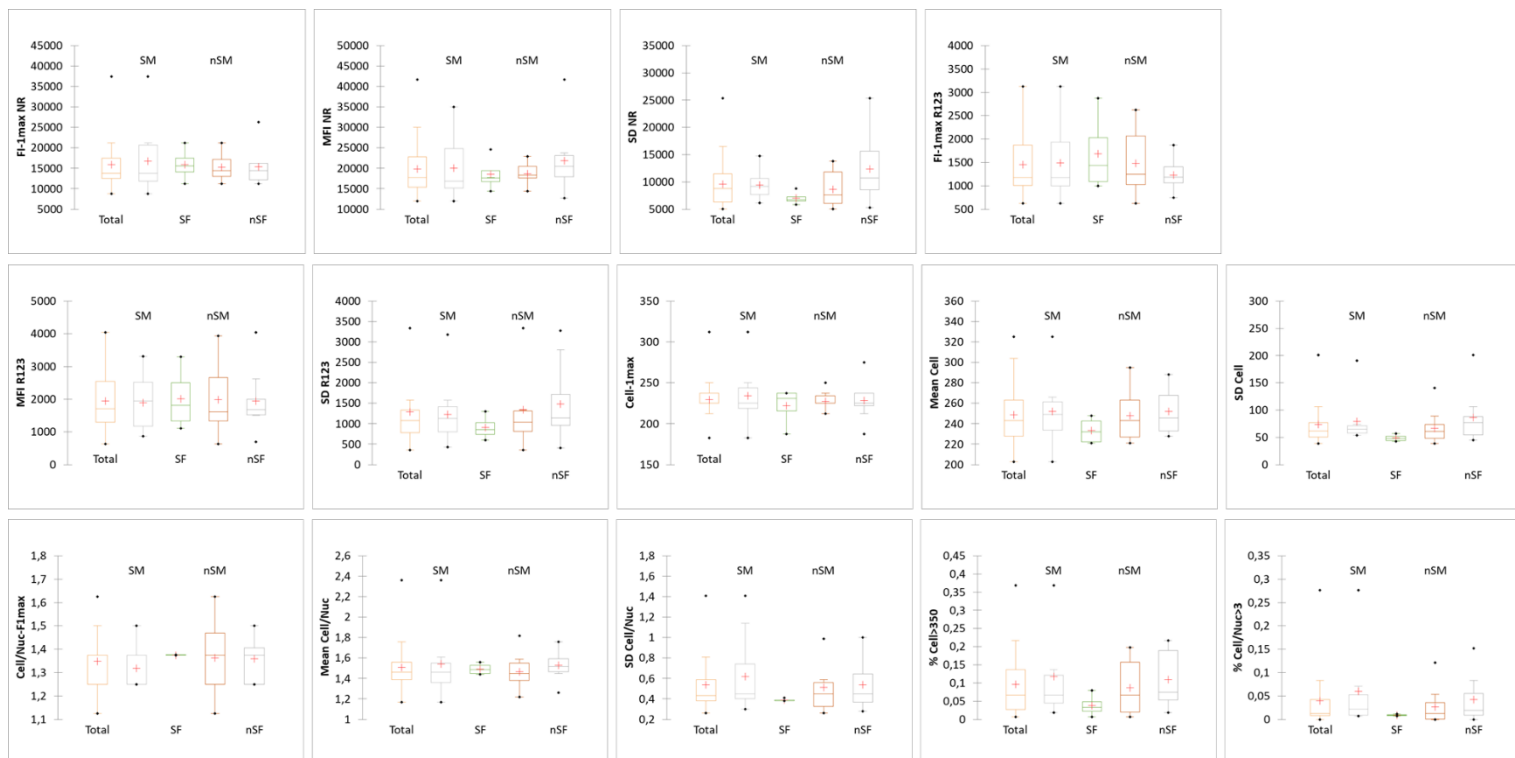


Table S4

Inter-population variations for the 41 parameters among the six sub-populations of donors, analyzed in pairs. Donors were grouped by smoking status (smokers vs. non-smokers separated or not (total population) according to sex), and gender (men vs. women according to their smoking habit). The mean and standard deviation (SD) for each sub-population are provided in Table 1. The comparison between two sub-populations is presented by (i) the variation, calculated using the equation $\% \text{ var} = (x-y)/\text{mean}(x,y)$, where x and y the parameters for smokers or males, and nonsmokers or females, respectively; and (ii) the adjusted p-value, calculated using the Benjamini & Hochberg false discovery correction procedure. Boldface indicates variation greater than 12%, and adjusted p-value less than 0.05.

A. Information on hematogram data

	SMOKERS / NON SMOKERS						MEN / WOMEN			
	TOTAL POPULATION		MEN		WOMEN		SMOKERS		NON SMOKERS	
	% var(s-ns)/mean	p value	% var(s-ns)/mean	p value	% var(s-ns)/mean	p value	%var (m-w)/mean	p value	%var (m-w)/mean	p value
WBC	45	<0.001	50	0.001	45	0.012	-13	0.45	-18	0.44
LYMP	-18	0.068	-22	0.059	-18	0.36	14	0.40	19	0.10
MONO	-14	0.60	-15	0.6	-13	0.72	1	0.99	3	0.99
GR	10	0.031	13	0.031	8	0.32	-6	0.42	-11	0.09
RBC	5	0.11	1	0.93	10	0.080	1	0.93	10	0.010
Hb	6	0.082	4	0.32	3	0.63	13	0.018	12	0.011
Hct	7	0.088	1	0.95	14	0.033	-1	0.95	13	0.016
MCV	2	0.76	0	0.89	4	0.76	-2	0.76	2	0.76
MCH	1	0.77	3	0.40	-6	0.26	11	0.051	2	0.74
MCHC	0	0.98	3	0.55	-7	0.20	10	0.12	0	0.98
RDW	6	0.025	7	0.036	6	0.16	-1	0.83	-2	0.63
PLT	9	0.85	12	0.85	4	0.92	1	0.90	-8	0.98
PLR	25	0.11	39	0.11	15	0.62	-8	0.75	-32	0.16

B. Lymphocyte proteasome analysis by flow cytometry

	SMOKERS / NON SMOKERS						MEN / WOMEN			
	TOTAL POPULATION		MEN		WOMEN		SMOKERS		NON SMOKERS	
	% var(s-ns)/mean	p value	% var(s-ns)/mean	p value	% var(s-ns)/mean	p value	%var (m-w)/mean	p value	%var (m-w)/mean	p value
MF1p21	-52	0.46	-53	0.46	-45	0.71	-17	0.92	-8	0.92
MF1p23	-42	0.72	-44	0.72	-52	0.84	29	0.80	22	0.80
MF1p25	-22	0.97	-24	0.97	-21	0.97	4	0.97	7	0.97
MF1p27	-16	0.89	-19	0.89	-1	0.98	-32	0.89	-14	0.89
MF1p29	-27	0.75	-31	0.75	-12	0.89	-26	0.89	-6	0.89
MF1p3033	-38	0.71	-43	0.78	-30	0.91	-4	0.98	10	0.98
MF1p31	-34	0.47	-37	0.56	-21	0.82	-27	0.82	-10	0.82
% cell+p21	-1	0.94	-7	0.94	12	0.94	-13	0.94	5	0.94
% cell+p23	6	0.93	1	0.93	13	0.93	-5	0.93	7	0.93
% cell+p25	-10	0.96	-15	0.96	-7	0.96	8	0.96	16	0.96
% cell+p27	3	0.98	-1	0.98	11	0.98	-10	0.98	2	0.98
% cell+p29	0	0.99	-9	0.99	12	0.99	0	0.99	21	0.99
% cell+p30-33	0	0.99	1	0.99	-1	0.99	3	0.99	0	0.99
% cell+p31	-1	0.93	-4	0.93	-3	0.93	13	0.93	14	0.93

C. Cell imaging of lymphocyte

	SMOKERS / NON SMOKERS						MEN / WOMEN			
	TOTAL POPULATION		MEN		WOMEN		SMOKERS		NON SMOKERS	
	% var(s-ns)/mean	p value	% var(s-ns)/mean	p value	% var(s-ns)/mean	p value	%var (m-w)/mp value	%var (m-w)/mear	p value	
FI-1_{max} NR	8	0.94	9	0.94	4	0.94	5	0.99	0	0.94
MFI NR	-2	0.98	7	0.98	-16	0.98	7	0.95	-16	0.98
SD NR	-16	0.54	9	0.75	-55	0.32	29	0.1	-36	0.32
FI-1_{max} R123	12	0.88	1	0.97	31	0.88	-12	0.77	18	0.88
MFI R123	-2	0.97	-5	0.97	3	0.97	-6	0.75	2	0.97
SD R123	-21	0.87	-9	0.87	-48	0.87	31	0.34	-9	0.87
Cell-1_{max}	1	0.97	3	0.97	-3	0.97	5	0.93	0	0.97
Mean Cell	-1	0.99	2	0.99	-8	0.99	8	0.24	-2	0.99
SD Cell	-6	0.79	17	0.79	-55	0.79	47	0.005	-26	0.79
Cell/Nuc-1_{max}	-2	0.97	-3	0.97	-4	0.97	-4	0.23	0	0.97
Mean Cell/Nuc	2	0.99	0	0.99	0	0.99	3	0.64	-4	0.99
SD Cell/Nuc	6	0.91	19	0.91	-32	0.91	46	0.11	-5	0.91
% Cell>350	-1	0.91	30	0.91	-97	0.91	102	0.1	-23	0.91
%c Cell/Nuc>3	32	0.79	77	0.79	-129	0.79	148	0.08	-46	0.79

Figure S2. DFA correlation circle of DFA performed with four groups for univariate and bivariate analyses.

DFA was performed to classified in four groups healthy donors ($n = 33$). DFA was run using “fixed class weight” option. The graphs present the correlation circle, parameters projection on the factor space (F1 and F2). (A) Proteasome and lymphocyte imaging data ($p = 28$), (D) Proteasome and CBC data ($p = 27$), (C) Lymphocyte imaging and CBC data ($p = 27$), (D) Only CBC data ($p = 13$). In these four DFAs, all the donors are correctly separated (training accuracy was 100%). We zoomed in to distinguished between parameters.

This representation can sometimes be used to identify groups characteristics according to parameters. No obvious information is gathered from these graphs, but all the CBC parameters, with the exception of PLT and MONO, are, for DFAs performed with hematogram data, the parameters closest to the circle, as expected according to the adjusted p-values presented in Table S4.

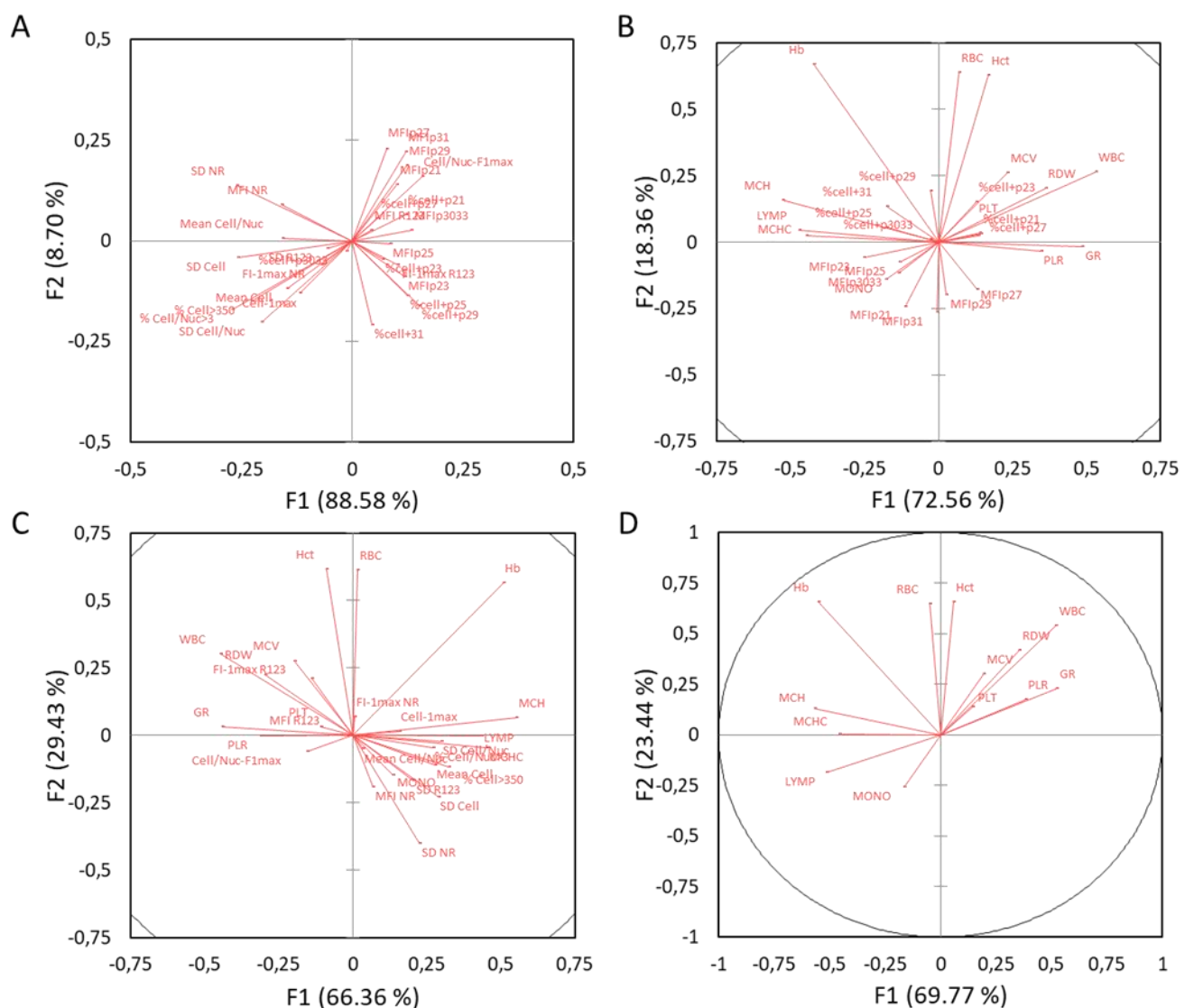


Figure S3. DFA classification into three sub-groups for univariate and bivariate analyses.

DFA was performed to classify healthy donors (n = 29) into three groups, excluding the four female smokers. Projection of donors on the two axes (F1 and F2), with confidence circles. SM: smoking male, nSM: non-smoking male, nSF: non-smoking female. (A) Lymphocyte-based approaches (imaging and cytometry) (p = 28), (D) Proteasome and CBC data (p = 27), (C) Lymphocyte imaging and CBC data (p = 27), (D) Only CBC data (p = 13).

For bivariable analyses, the training population showed 100% training accuracy and low validation accuracy (17%, 31% and 41%, for A, B and C, respectively). In contrast to the DFA performed on four groups, there are only two factorial axes, allowing the man to be separated according to their smoking habits. Hematological data alone (Figure D) give a training accuracy of 93% (validation accuracy 63%).

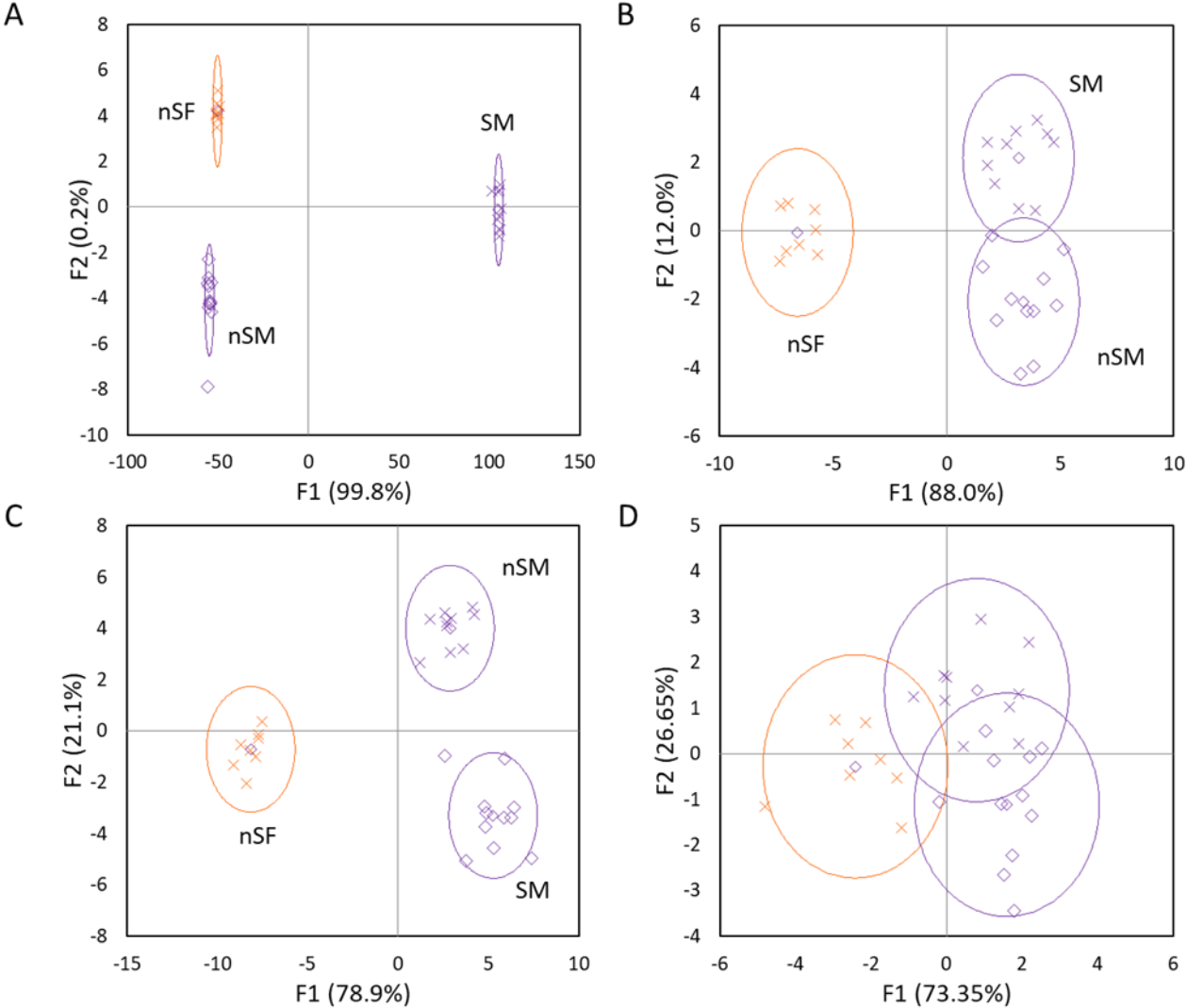


Table S5. Informative parameters selected from the three datasets by regression. Column 1: Ridge regression, Column 2: Lasso regression, Column 3: Elastic net regression with equal weight for Lasso and Ridge. The selected parameters were used in DFA with the donors divided in four and three groups (without female smokers). The results are presented as training accuracy (validation accuracy) in percentage.

Regressions					
Ridge		Lasso		Elastic Net	
4 groups	3 groups	4 groups	3 groups	4 groups	3 groups
88% (57%)	97% (66%)	82% (71%)	86% (69%)	92% (71%)	93% (62%)
Cell/Nuc-F1_{max} Mean Cell/Nuc SD Cell/Nuc % Cell>350 % Cell/Nuc>3		SD NR Cell/Nuc-F1_{max}		SD NR Cell/Nuc-F1_{max}	
MFip31		MFip31		MFip31	
RBC Hb Hct MCH MCHC RDW		Hb RDW		GR RBC Hb Hct RDW	

Ridge, Lasso and Elastic net regressions were tested by varying the partition until R^2 (coefficient of determination) was closest to 1 and MSE (mean square error) was minimized. R^2 indicated the goodness of fit of the model, while MSE reflected the model's predictive accuracy. The analysis involved 33 donors (four groups) and 29 donors (three groups). A compromise was chosen for Lasso regression with R^2 equal to 0.60 and MSE equal to 0.55 with a small lambda (0.02), resulting in the selection in of 5 parameters. Ridge regression, which was unaffected by partitioning, produced ten more adequate parameters with a lambda around 4 ($R^2 < 0.60$ and $MSE \approx 0.6$). Using Elastic Net regression, with equal weight for Lasso and Ridge, 7 parameters were identified, including one new parameter (GR), with a lambda around 0.2 ($R^2 = 0.62$ and $MSE = 0.5$). The differences observed when the four female smokers were removed were minimal. One donor, donor 16, was repeatedly misclassified. All parameters identified by these methods were already included in the final model, except for Cell/Nuc-F1_{max}.