Electronic Supplementary Material (ESI) for Sensors & Diagnostics.

A methodology study toward the diagnostic of a SARS-Cov-2 infection in human serum with a macrocyclic based sensor array

Monica Swetha Bosco[1,2], Zeki Topcu[4], Soumen Pradhan[6,7], Ariadne Sossah[1], Vassilis Tsatsaris[1,3], Christelle Vauloup-Fellous[5], Sarit S. Agasti[6,7], Yves Rozenholc[4*], Nathalie Gagey-Eilstein[1,2*]

[1]Université Paris Cité, INSERM UMR-S 1139, FHU PREMA, 4 avenue de l'observatoire, 75006 PARIS, France

[2]Université Paris Cité, CNRS, INSERM, UTCBS, 4 avenue de l'observatoire, 75006 PARIS, France

[3]Department of Obstetric, Cochin Hospital, AP-HP, Université Paris Cité, FHU PREMA, 123 Bd Port-Royal, 75014 Paris, France.

[4]UR 7537, BioSTM, Université Paris Cité, 4 avenue de l'Observatoire, 75006 Paris, France.

[5] Université Paris-Saclay, INSERM U1193, Virology Laboratory, Hôpital Paul-Brousse, AP-HP, Villejuif, France

[6]New Chemistry Unit, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore, Karnataka 560064, India.

[7]Chemistry & Physics of Materials Unit, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore, Karnataka 560064, India.

Correspondance e-mail: *nathalie.eilstein@u-paris.fr*

[*]These two authors contributed equally to the work.

TABLE OF CONTENTS

A. Clinical data

Table S1: Detailed clinical informations for each samples

N° sample	Covid status	Age	WG*	Days between symptoms and test	Breath symptoms	Hospitalization	Oxygen need	ICU
1	I	40	40	0	Y	Y	Y	Y
2	I	31	40	4	Y	N	N	N
3	I	37	35	5	Y	N	N	N
4	I	28	39	0	Y	N	N	N
5	I	31	40	1	Y	N	N	N
6	I	36	37	16	Y	N	N	N
7	I	31	29	16	N	N	N	N
8	I	32	28	11	Y	N	N	N
9	I	32	40	11	N	N	N	N
10	I	31	33	1	Y	Y	Y	Y
11	I	25	38	1	N	N	N	N
12	I	28	39	7	N	N	N	N
1	NI	39	40					
2	NI	27	40					
3	NI	19	40					
4	NI	30	38					
5	NI	37	38					
6	NI	31	38		N.A			
7	NI	30	39					
8	NI	42	38					
9	NI	38	39					
10	NI	31	41					

11	NI	44	36
12	NI	26	36
13	NI	23	28
14	NI	43	33

*WK = Week of Gestation

B. Sensing Covid in clinical specimen

Table S2: Excitation and emission wavelengths used for each CB[7]-FL.

FL in CB[7]-FL	$\lambda_{excitation}$ (nm)	$\lambda_{emission}$ (nm)
Bodipy	500	515
Coumarin	418	433
Cyanine 3	515	560
Cyanine 5	630	650
Fluorescein	500	515
Silicorhodamine	645	660
TAMRA	554	574

Figure S1: Effect of complex matrix on fluorescence of CB[7]-FL. Variation in the fluorescence emission spectra of a/ CB[7]-BDP b/ CB[7]-TMR after addition of proteins in serum matrix (1.6 mg/mL) and control (serum only). c/ Histogram of the fluorescence emission at $\lambda_{em\ max}$ of CB-BDP, CB-TMR, CB-Cy3, CB-Cy5, CB-Cou and CB-SiR, in presence of serum before or after addition of proteins (BSA, pepsin and lysozyme) at 1.6 mg/mL. ** indicate significant differences (p-value < 0.05)*

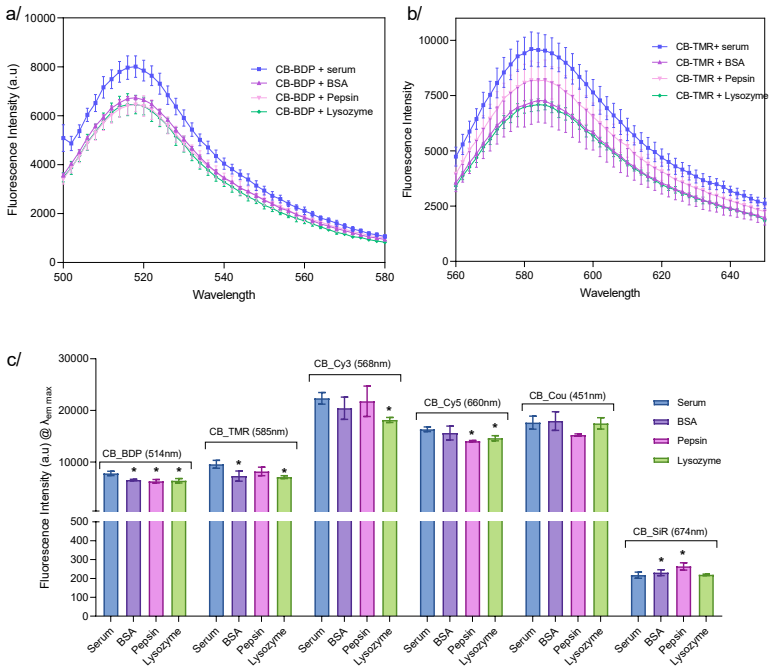


Figure S2: Stability of the fluorescence signals with different CB-[FL] during the 30 min of measurements

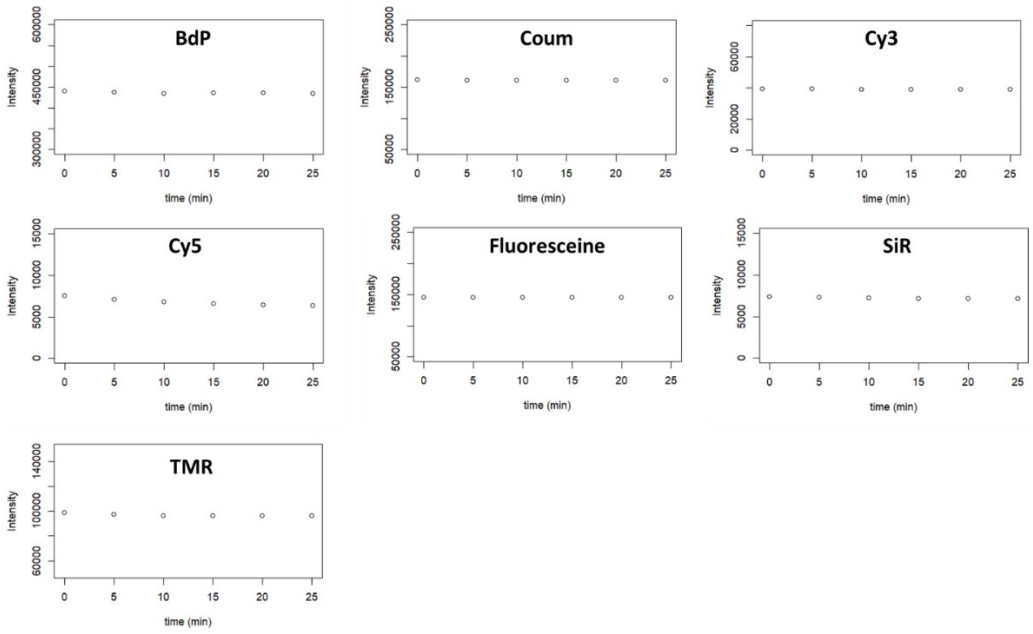


Figure S3: Microplates organization – 1+, 2+, 3+, 4+, 5+, 6+, 7+, 8+, 9+, 10+, 11+, 12+ are Covid+ samples and 1-, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11-, 12-, 13-, 14- are control samples. Black positions in B11 to E11 are CB[7]-FL with PBS only instead of serum.

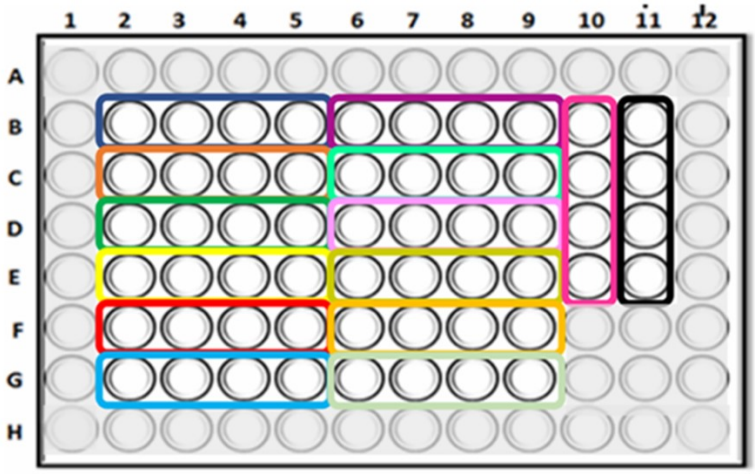


Plate 1

1+ : B2, B3, B4, B5
3+ : C2, C3, C4, C5
5+ : D2, D3, D4, D5
7+ : E2, E3, E4, E5
9+ : F2, F3, F4, F5
11+ : G2, G3, G4, G5
1- : B6, B7, B8, B9
3- : C6, C7, C8, C9
5- : D6, D7, D8, D9
7- : E6, E7, E8, E9
9- : F6, F7, F8, F9
11- : G6, G7, G8, G9
13- : B10, C10, D10, E10
Contrôle : B11, C11, D11, E11

Plate 2

2+ : B2, B3, B4, B5
4+ : C2, C3, C4, C5
6+ : D2, D3, D4, D5
8+ : E2, E3, E4, E5
10+ : F2, F3, F4, F5
12+ : G2, G3, G4, G5
2- : B6, B7, B8, B9
4- : C6, C7, C8, C9
6- : D6, D7, D8, D9
8- : E6, E7, E8, E9
10- : F6, F7, F8, F9
12- : G6, G7, G8, G9
14- : B10, C10, D10, E10
Contrôle : B11, C11, D11, E11

Table S3: Data of Fluorescence modulation, $\bar{\Delta I}$, of CB-FLs against serum samples (**1**: Sars-Cov2 infected serum samples, **2**: control serum samples)

Sample	Covid status	BdP	Coum	Cy3	Cy5	Fluor	SiR	TMR
1+	1	-100977,38	-37362,04	34368,25	-1088,58	27325,33	-6286,17	-26557,79
3+	1	-335107,08	-159768,08	29186,13	-1722,71	-106612,13	-6319,00	-88688,42
5+	1	-68893,33	-92973,00	45036,42	-1596,04	87689,71	-6640,63	-23434,75
7+	1	-133037,46	-101560,67	17072,88	-1464,50	-60315,75	-6412,75	-29726,63
9+	1	-203599,00	-104554,63	8662,88	-1840,08	-72260,67	-5843,04	-26606,33
11+	1	-185874,92	-113050,88	9921,58	-1672,29	-70588,13	-5735,38	-30772,83
1-	-1	-196079,15	-99345,79	16444,75	-2039,04	-49643,17	-7106,19	-27855,54
3-	-1	-128953,67	-92031,04	18264,88	-2251,75	-45384,92	-7304,50	-28776,29
5-	-1	-140275,71	-96240,96	23421,25	-1820,58	-14528,46	-6886,79	-24953,93
7-	-1	-184890,71	-90600,38	13964,49	-2043,92	-69207,21	-6803,79	-20248,42
9-	-1	-247232,75	-122156,88	7224,63	-2107,33	-77408,54	-6136,54	-27620,29
11-	-1	-145756,29	-89371,08	14840,13	-2010,71	-54077,79	-6547,71	-21355,75
13-	-1	-168824,88	-89789,21	14415,50	-1868,29	-48213,00	-6823,21	-25259,67
2+	1	-96936,25	-110928,13	14218,38	-2052,71	3529,75	-10645,21	-33003,96
4+	1	-75001,71	-88250,88	16230,75	-2226,08	-11588,71	-10648,71	-30618,96
6+	1	-135989,33	-140846,13	2282,37	-2508,29	-46848,88	-11085,13	-51040,96
8+	1	-104084,04	-90749,29	-148,21	-2690,08	-48809,88	-10669,42	-33807,46
10+	1	-173953,54	-150597,96	3774,75	-2693,25	-82865,58	-10146,21	-60229,29
12+	1	-134602,83	-115489,79	4471,17	-2173,54	-55977,00	-10245,50	-31619,71
2-	-1	-187206,58	-165789,67	-13957,38	-3575,56	-62294,42	-11550,54	-81514,96
4-	-1	-142263,46	-116576,04	-674,58	-3144,54	-59988,67	-11617,00	-36797,79
6-	-1	-103875,36	-112677,75	19119,21	-3051,79	-10778,67	-11148,29	-33746,67
8-	-1	-121953,79	-110793,17	7781,37	-3065,58	-36462,13	-11212,58	-33059,08
10-	-1	-54591,83	-92565,33	21856,67	-3088,04	36676,96	-10803,67	-28357,79
12-	-1	-140930,46	-112764,13	-678,58	-3033,33	-63730,25	-10807,67	-35957,21
14-	-1	-139832,83	-131591,54	7531,50	-2963,92	-43370,96	-11263,08	-32212,75

Figure S4: Fluorescence modulation after addition of serum in each well. Covid+ samples are on the left, Control sample on the middle and control well on the right.

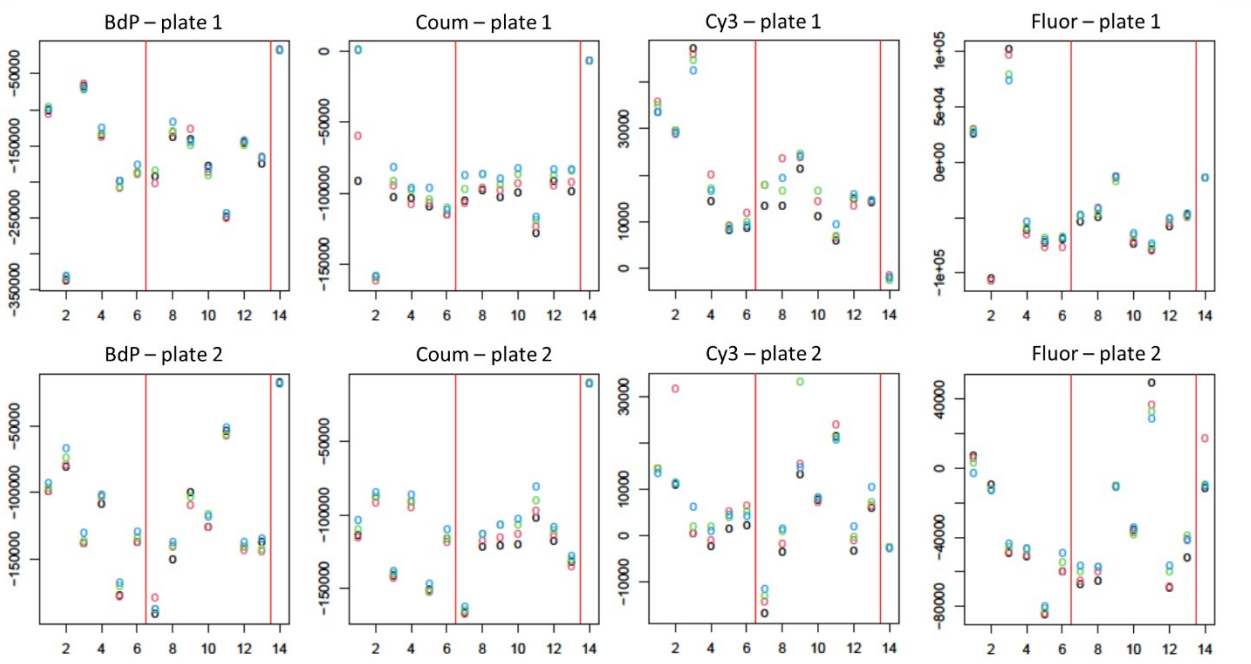


Figure S5: Stabilization of the error rates for the random forest on all the variables (a) and only on the important variables, Cy5 and SiR (b)

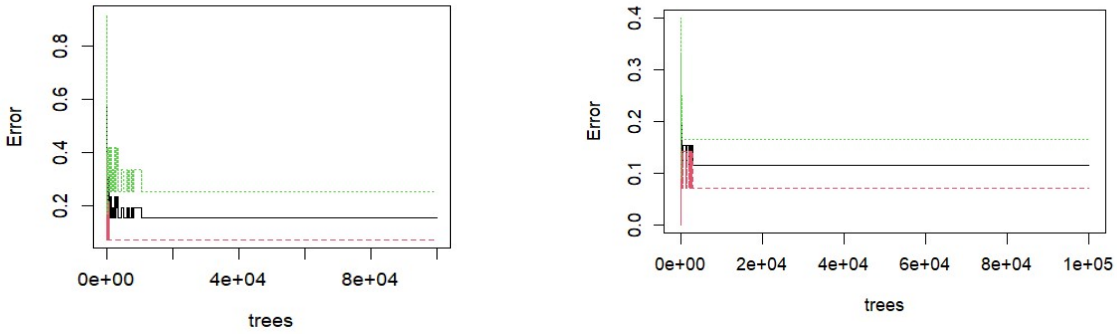


Table S4: Data of Fluorescence modulation, ΔI, of CB-FLs against serum samples (P: Pregnant (2 combined datasets), NP: Non pregnant, Covid: Sars-Cov2 infected serum samples)

	CB_Cy5	*CB_Bdp*	*CB_SiR*	*CB_TMR*	*CB_Cy3*	*CB_Coum*
P	-3446,389	-187206,58	-11550,542	-81514,96	-13957,375	-165789,67
P	-3144,542	-142263,46	-11617	-36797,79	-674,5833	-116576,04
P	-3051,792	-104629,06	-11148,292	-33746,67	19119,2083	-112677,75
P	-3065,583	-121953,79	-11212,583	-33059,08	7781,375	-110793,17
P	-3088,042	-54591,83	-10803,667	-28357,79	21856,6667	-92565,33
P	-3033,333	-140930,46	-10807,667	-35957,21	-678,5833	-112764,12
P	-2963,917	-139832,83	-11263,083	-32212,75	7531,5	-131591,54
P	-2039,042	-193913,33	-7026,778	-27855,54	16352,1111	-99345,79
P	-2251,75	-128953,67	-7304,5	-28776,29	18264,875	-92031,04
P	-1820,583	-140275,71	-6886,792	-24953,92	23421,25	-96240,96
P	-2043,917	-184890,71	-6803,792	-20248,42	14094,1667	-90600,38
P	-2107,333	-247232,75	-6136,542	-27620,29	7224,625	122156,88
P	-2010,708	-145756,29	-6547,708	-21355,75	14840,125	-89371,08
P	-1868,292	-168824,88	-6823,208	-25259,67	14415,5	-89789,21
P	1690,06	-88540,52	-5319,65	-8902,73	32807,21	-52670,48
P	1525,73	-85174,19	-5827,48	-10058,06	34345,21	-59142,98
P	1543,4	-77320,19	-6016,65	-8264,9	34334,38	-57700,98
P	1678,23	-73041,52	-5907,81	-5588,56	35764,04	-53545,81
P	1516,06	-74916,19	-6016,81	-8062,4	34382,38	-56137,65
P	1664,23	-90913,69	-5774,48	-7884,73	32524,71	-52595,65
P	1626,06	-67777,19	-5828,31	-9456,06	37104,38	-46513,98
P	1694,23	-77103,02	-6039,98	-9610,9	32856,21	-45789,65
P	1638,56	-68564,52	-6225,81	-10059,73	40317,88	-46861,98
P	1821,73	-63000,19	-6324,65	-9297,23	36709,38	-46239,48
P	1653,4	-68675,19	-6608,65	-9320,06	34823,38	-45587,65
P	1820,23	-72750,02	-6123,15	-9089,73	36033,04	-44187,31
P	1685,06	-95091,69	-6108,65	-7551,4	39311,04	-50310,81
P	1719,4	-92602,35	-6180,98	-8938,23	38436,38	-53123,31
P	1695,9	-91944,85	-6410,81	-7931,23	39311,38	-52402,98
P	1603,23	-94356,19	-6442,98	-7990,9	39081,54	-52295,81
P	1719,23	-96028,35	-6719,31	-7742,23	39747,21	-50034,31
P	1824,56	-88318,52	-6539,98	-7524,23	39579,38	-51266,81
P	1638,06	-93054,85	-6579,31	-8051,4	38059,21	-52160,81
P	1735,9	-98695,02	-6543,48	-8264,9	40313,88	-54326,31
P	1725,06	-96420,52	-6796,15	-7756,4	41745,54	-54301,65
P	1839,9	-92472,35	-6877,65	-7556,73	41340,38	-53055,15
P	1926,73	-95547,85	-6739,81	-7609,06	40231,54	-52496,81
P	1809,73	-102659,02	-6559,48	-7050,9	40872,88	-50680,15
P	1883,73	-102128,02	-6475,65	-8007,9	36731,38	-65553,48
P	1967,4	-105015,69	-6945,48	-8148,9	37092,21	-69410,81
P	1952,9	-98213,02	-6901,48	-8637,06	39244,21	-64575,15
P	1865,06	-98316,69	-6790,65	-10028,73	36665,38	-66733,98
P	1940,9	-104259,85	-6792,65	-8346,06	38603,21	-66893,31

P	1986,4	-96352,35	-6371,15	-8262,23	38085,04	-65476,15
P	2852,6	-86423,44	-4708,81	-13964,25	24509,75	-54130,4
P	2806,44	-86354,94	-4810,15	-13986,75	28210,75	-54654,4
P	2707,77	-86210,94	-4803,65	-13551,25	22747,92	-53321,4
P	2738,6	-86888,6	-4756,31	-13707,58	22496,08	-53053,56
P	2780,1	-86822,44	-4764,65	-13038,92	23144,08	-52723,23
P	2808,44	-84545,6	-4764,98	-13831,25	22237,25	-53559,23
P	2775,94	-75756,6	-5163,98	-5608,75	42138,75	-31130,9
P	3218,6	-75496,44	-5320,48	-6188,92	46812,42	-32529,23
P	3286,6	-77256,6	-5553,98	-5756,92	38543,92	-30847,23
P	4042,77	-78086,1	-5198,98	-5715,75	39119,25	-31317,4
P	4090,77	-79526,77	-5503,65	-6332,58	41220,75	-32222,56
P	3306,27	-78002,27	-5033,81	-5949,08	36943,92	-30601,73
NP	1863,06	-132815,69	-6410,15	-7433,56	38167,54	-59593,81
NP	1711,23	-125201,02	-6701,65	-8707,4	38356,54	-59085,31
NP	1886,06	-130831,02	-6763,98	-8596,73	38317,88	-59245,98
NP	1747,4	-125199,52	-6847,81	-7313,4	36398,71	-60088,48
NP	1989,73	-131704,69	-6839,81	-7156,23	38962,38	-57306,81
NP	1885,73	-131921,02	-6710,65	-6876,73	41057,88	-58338,98
NP	1964,4	-120134,85	-6626,65	-15123,4	24781,71	-76338,31
NP	1792,56	-120444,02	-6770,81	-15315,4	25153,21	-78879,65
NP	1949,4	-125060,52	-7287,65	-14868,06	24917,71	-80286,65
NP	2031,23	-119649,02	-6635,31	-15919,06	30390,54	-80713,65
NP	2047,4	-122046,19	-6998,48	-15043,23	28651,38	-77786,48
NP	1910,73	-118153,52	-7327,31	-14571,73	30095,21	-78445,48
NP	1931,4	-162399,19	-4454,65	-6509,06	40340,38	-73221,48
NP	1881,56	-178452,35	-4645,65	-7782,56	41036,21	-75121,15
NP	1827,73	-164926,52	-4803,15	-7622,73	41667,54	-76412,98
NP	1855,73	-173629,69	-4776,15	-7657,9	41996,88	-74895,15
NP	1889,4	-166619,19	-4669,15	-6716,73	54340,71	-72779,98
NP	1917,73	-149115,35	-4863,65	-6902,9	42758,21	-75225,31
NP	2244,4	-132602,02	-4441,48	-10666,9	54362,54	-65376,81
NP	2128,23	-124045,35	-4683,31	-12516,06	61515,54	-67507,31
NP	2312,56	-122827,02	-4781,98	-11967,9	48499,71	-66178,98
NP	2175,9	-123014,69	-4877,98	-11959,4	41833,04	-66063,98
NP	2090,73	-123506,69	-4784,65	-12788,73	40323,04	-64876,48
NP	2347,73	-120813,02	-4733,31	-11226,73	38655,88	-63284,31
NP	2128,73	-157387,85	-4472,31	-10704,23	45957,04	-67993,98
NP	2132,73	-155890,52	-4859,65	-10197,9	44088,38	-66203,15
NP	2069,23	-147190,19	-4761,98	-10349,9	40534,71	-64789,15
NP	2166,73	-145353,19	-4892,31	-11381,4	39208,71	-64387,81
NP	1998,56	-131797,69	-5149,15	-11420,06	38016,71	-65861,81
NP	2161,56	-120526,02	-4667,65	-11264,73	36441,21	-62606,81
NP	2498,44	-105088,77	-8875,98	-7632,92	46298,42	-46085,9
NP	3000,77	-102449,1	-8994,81	-7468,92	43131,25	-46287,23
NP	2133,77	-101896,94	-8792,15	-8336,92	36052,08	-48313,23
NP	2073,44	-105028,44	-8905,81	-7039,75	43603,25	-47071,4
NP	2884,94	-102947,6	-8882,31	-7446,42	38290,75	-44894,06
NP	3259,44	-103547,6	-8788,15	-6952,75	36692,75	-45425,73
NP	2947,27	-119070,27	-4643,65	-6228,25	51104,58	-62738,23
NP	3719,77	-121121,1	-4696,65	-6441,42	47888,08	-61390,56
NP	3003,27	-119983,77	-4811,48	-6591,08	45047,42	-58682,73
NP	2703,44	-119158,44	-4588,65	-6180,42	48608,25	-58685,9
NP	3009,27	-119665,94	-4431,48	-5429,75	39946,58	-56622,9
NP	3855,77	-117609,44	-4790,31	-6017,92	40124,08	-66290,9
NP	4967,1	-129962,1	-4687,31	-11029,08	43143,58	-51114,9
NP	5111,6	-127439,1	-4442,31	-10623,08	42798,75	-50427,4
NP	4487,44	-129477,6	-4635,15	-10932,92	42302,42	-51303,73
NP	4198,94	-129285,27	-4813,31	-11291,42	35822,42	-50653,9
NP	5128,6	-128218,44	-4603,15	-9688,92	39779,42	-51592,23
NP	4369,94	-125100,77	-4354,48	-9077,25	48471,58	-50071,9
Covid	-1088,583	-100977,38	-6286,167	-24584,42	34368,25	-37362,04
Covid	-1722,708	-335107,08	-6319	-88688,42	34368,25	-159768,08
Covid	-1596,042	-68893,33	-6640,625	-23434,75	45036,4167	-92973
Covid	-1464,5	-133037,46	-6412,75	-29726,62	17072,875	-101560,67
Covid	-1840,083	-203599	-5843,042	-26606,33	8662,875	-104554,62
Covid	-1672,292	-185874,92	-5735,375	-30772,83	9921,5833	-113050,88
Covid	-2052,708	-96936,25	-10645,208	-33003,96	14218,375	-110928,12
Covid	-2226,083	-75001,71	-10648,708	-30618,96	16230,75	-88250,88
Covid	-2508,292	-135989,33	-11085,125	-51040,96	2282,375	-140846,12
Covid	-2690,083	-104084,04	-10669,417	-33807,46	-148,2083	-90749,29

Covid	-2693,25	-173953,54	-10146,208	-60229,29	3774,75	-150597,96
Covid	-2173,542	-134602,83	-10245,5	-31619,71	4471,1667	-115489,79

Figure S6: Discrimination of Pregnant (P) Non-pregnant (NP) and Sars-COV-2 pregnant (Sars-COV-2 P) clinical samples from two sets of experiments. a/ Explanation of the different cohorts and combination b/ Confusion matrix for discrimination of P and NP with the random forest analysis using the whole variables. c/ Confusion matrix for discrimination of P, NP and Sars-COV-2 P with the random forest analysis using the whole variables and the most important variables.

a/

Cohort 1	Cohort 2	Combined cohort
48 NP		42 NP
42 P	14 P	56 P+P
	12 Sars-Cov-2 P	12 Sars-Cov-2 P

b/

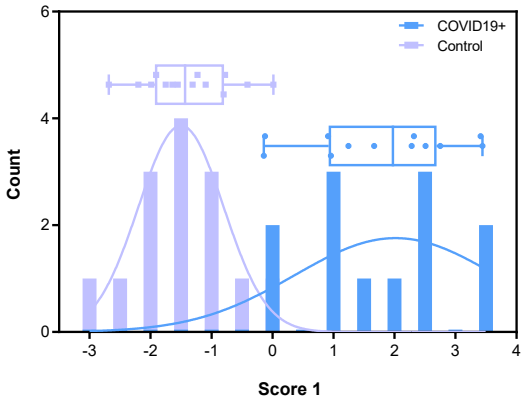
OOB estimate of error rate : 2.9 %		
	P	NP
P	54	2
NP	1	47

c/

	ALL - OOB estimate of error rate : 12.1 %			Cy5, Bdp, SiR, TMR, Cy3 OOB estimate of error rate : 6.9 %		
	P	NP	P Sars-COV-2	P	NP	P Sars-COV-2
P	51	2	3	52	2	2
NP	1	47	0	1	47	0
P Sars-COV-2	4	0	8	3	0	9

Figure S7: LDA models for the 2-group “control versus covid” model. The histogram is marked with normal distributions fitted to the full data consisting of average of replicates for each serum sample. The box-plot depicts the max/min of canonical scores obtained from the LDA.

a.

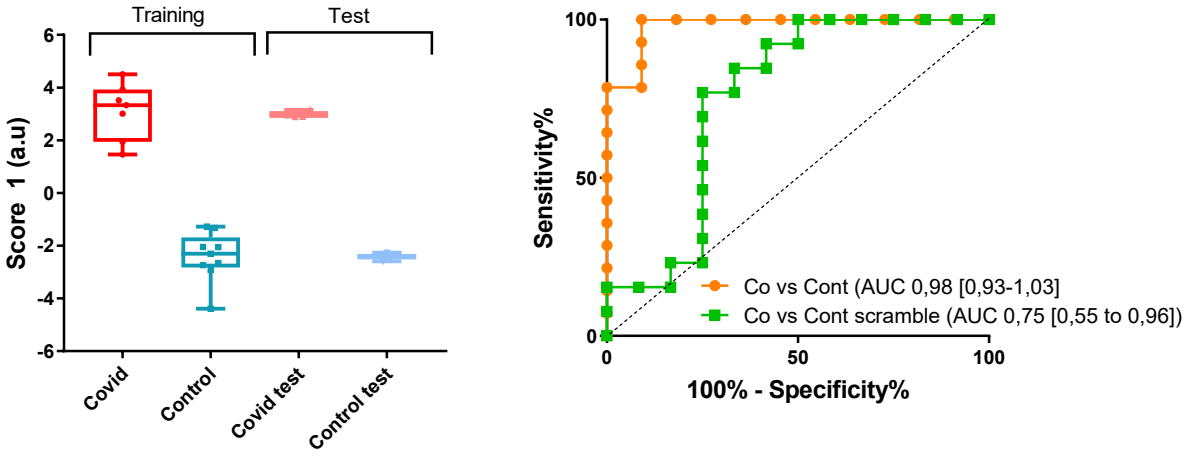


b.

85%	Control	Covid	%error
Control	13	1	7
Covid	2	10	16.6

Random attribution			
58%	Control	Covid	%correct
Control	5	9	64
Covid	5	7	42

Figure S8*:* Diagnosis of Covid-19 samples based on serum response toward the CB[7]-FL sensor array. A/ Box plot of the first canonical score of the training (70% of the samples) and test set (30% of the samples). The horizontal line in the boxes represents the median and the bottom and top of the boxes represent the 25th and 75th percentiles, respectively. B/ ROC curves for the array-based sensor in patients with Covid-19 (Co) compared with control (Cont). (~~~) is for normal data and (~~~) is for scrambled data.



Annex 1 : R-code for the data extraction

❖ Lecture-raw-data

```
rows.letters = LETTERS[2:7] # B,C,D,...,G
cols.info = 3:12
n.rows = length(rows.letters)
n.cols = 10
n.cycles = 6

raw.folder = '../Raw Data Covid'
out.folder = '../Table Data Covid'
ref.folder = '../Ref Data Covid'

if (!dir.exists(out.folder)) dir.create(out.folder)
if (!dir.exists(ref.folder)) dir.create(ref.folder)

list.dir = dir(raw.folder,all.files=F,recursive=F,full.names=T)


for (name in list.dir) { # read directory content
  out.name = gsub(raw.folder,out.folder,name) # directory for output
  if (!dir.exists(out.name)) dir.create(out.name)

  list.file = list.files(path=name,pattern="*.csv",full.names=T)

  for (file in list.file) { # read each file
    print(paste('processing of',file))
    X = read.csv(file, header=F, blank.lines.skip=FALSE, stringsAsFactors=FALSE)
    if (dim(X)[2]==1) X = read.csv(file, header=F, blank.lines.skip=FALSE, stringsAsFactors=FALSE, sep=';')
    IntensityArray = array(0,dim=c(n.rows,n.cols,n.cycles),dimnames=list(rows.letters,1+(1:n.cols),1:n.cycles))
    i = tps = 1
    for (line in 1:dim(X)[1]) { # read line by line the file
      champs = X[line,] # get the contents of the line
      # print(champs[1])
      if (!(champs[1] %in% rows.letters)) next() # forget non intensity line
      # here we are on intensity line
      intensities = as.numeric(champs[cols.info])
      # print(intensities)
      IntensityArray[i,,tps] = intensities
      i = i+1

      if (champs[1]==rows.letters[n.rows]) {i=1; tps=tps+1}
      if (tps==n.cycles+1) break() # too many cycles read // stop
    }

    if (grepl('Coum',file)) print(IntensityArray)

    save(IntensityArray,file=gsub(raw.folder,out.folder,gsub('.csv','.rda',file)))

    if (grepl('Cy3',file)) { # we deal with NA's on experiment 2569 Cy3 (1) I1
      print(file)
      ref.name = gsub(raw.folder,ref.folder,name) # directory for output
      if (!dir.exists(ref.name)) dir.create(ref.name)
      founded = FALSE
      while (!founded) {
        line = line+1
        old.champs = champs
        champs = X[line,]
        if (length(old.champs)<=9 | length(champs)<=9) next()
        founded = ((old.champs[9]=='Wavelength') & (as.numeric(champs[9])==560))
      }

      line = line+6
      Ref560Array = array(0,dim=c(n.rows,n.cols),dimnames=list(rows.letters,1+(1:n.cols)))
      i = 1
      for (l in line:(line+5)) {
        champs = X[l,]
        intensities = as.numeric(champs[cols.info])
        Ref560Array[i,] = intensities
        i = i+1
      }

      save(Ref560Array,file=gsub(raw.folder,ref.folder,gsub('.csv',' ref560.rda',file)))
    }  }  }
```

❖ read-plate-data.R

```
the.method = 'Fatio'

plate2table = function(plate) {
  # print(plate)
```

```
  table  = rbind(plate[1:6,1:4], plate[1:6,4+(1:4)], plate[1:4,9], plate[1:4,10])
  rownames(table) = c(LETTERS[1:13],0)
  colnames(table) = 1:4
  # print(table)
  table
}

# directory des fichiers .rda
in.folder = '../Table Data Covid'

# directory des fichiers resultats
out.folder = sub('Table','Treated',in.folder)

if (!dir.exists(out.folder)) dir.create(out.folder)

list.dir = dir(in.folder,all.files=F,recursive=T,full.names=T)
list.I0 = grep('I0.rda',dir(in.folder,all.files=F,recursive=T,full.names=T), value=T)
list.I1 = grep('I1.rda',dir(in.folder,all.files=F,recursive=T,full.names=T), value=T)

for (name in list.I1)
{ # read directory content

  load(name)

  # get name of output sub-folder and create it if not existing
  in.sub.folder = substr(name,gregexpr('/',name)[[1]][2],gregexpr('/',name)[[1]][3])
  out.sub.folder = paste0(out.folder,in.sub.folder)
  fluo = gsub('/','',in.sub.folder); fluo = gsub(' ','',fluo)
  print(fluo)

  if (!dir.exists(out.sub.folder)) dir.create(out.sub.folder)

  MeanIntensityArray = apply(IntensityArray,1:2,mean,na.rm=TRUE,trim=0.1)

   IntensityTable = array(dim=c(14,4,dim(IntensityArray)[3]))
  for (i in 1:dim(IntensityArray)[3]) IntensityTable[,,i] = plate2table(IntensityArray[,,i])

  # Reorganize in a table instead of a plate // 4 values per line // control is the last line
  I = plate2table(MeanIntensityArray)

  bool = apply(I,1,function(l) all(is.nan(l))) # check for line made only of NaN
  if (any(bool)) {
    ref = sub('.rda',' ref560.rda',sub('Table','Ref',name))
    print(paste('use',ref))
    load(ref)
    I[bool,] = plate2table(Ref560Array)[bool,] # get line from ref
  }

  I.mean = rowMeans(I,na.rm=T)  # mean over replicates

  # we deal with changes in the plate numbering from I1 to I0 in order to find I0 name
  last.barre = rev(gregexec('/',name)[[1]][1,])[1]
  end.name = substring(name,last.barre)
  next.white = gregexec(' ',end.name)[[1]][1,][1]
  end.name = substring(end.name,next.white+1)
  new.end = sub('I1','I0',end.name)

  # Do same for the associated control plate
  name.I0 = grep(new.end,list.I0,fixed=T,value=T)
  load(name.I0)
  MeanIntensityArray = apply(IntensityArray,1:2,mean,na.rm=TRUE,trim=0.1) # variance 1/6 - If only NA's create NaN
  IntensityTable0 = array(dim=c(14,4,dim(IntensityArray)[3]))
  for (i in 1:dim(IntensityArray)[3]) IntensityTable0[,,i] = plate2table(IntensityArray[,,i])
  I0 = plate2table(MeanIntensityArray)
  I0.mean = rowMeans(I0,na.rm=T)  # mean over replicates

  out.name = paste0(out.sub.folder,substring(name,gregexpr('/',name)[[1]][3]+1))
  out.name = gsub('.rda',paste0('-',the.method,'.rda'),out.name)
  if (the.method=='Eatio') {
    Eatio.all = I - I0.mean
    Eatio.mean = I.mean - I0.mean
    Eatio.sd = apply(Eatio.all,1,sd,na.rm=T)
  }
  use.all = eval(parse(text=paste(the.method,'all',sep='.')))
  use.mean = eval(parse(text=paste(the.method,'mean',sep='.')))
  use.sd = eval(parse(text=paste(the.method,'sd',sep='.')))
  covid.status = as.factor(c(rep(1,l=6),rep(-1,l=7))) # 6 first are covid+

  save(fluo, I, I0, use.all,use.mean,use.sd, covid.status, file=out.name)

  }
```

- **build-data.R**

```
the.method = 'Eatio'

# directory des fichiers .rda
in.folder = '../Treated Data Covid'

# directory des fichiers resultats
out.folder = sub('Treated','Tree',in.folder)
if (!dir.exists(out.folder)) dir.create(out.folder)

list.dir = dir(in.folder,all.files=F,recursive=T,full.names=T,pattern = paste0('*-',the.method,'.rda'))
```

```
sample1 = sample2 = fluo1 = fluo2 = NULL
for (name in list.dir) { # read directory content
  print(name)
  tmp = load(name)
  if (grepl('(1)',name,fixed=T)) {
    sample1 = cbind(sample1,use.mean[names(use.mean)!='0'])
    fluo1 = c(fluo1,fluo)
    covid1 = covid.status
  } else {
    sample2 = cbind(sample2,use.mean[names(use.mean)!='0'])
    fluo2 = c(fluo2,fluo)
    covid2 = covid.status
  }
}
samples = rbind(sample1[,order(fluo1)],sample2[,order(fluo2)])
colnames(samples) = fluo1[order(fluo1)]
covid.status = c(covid1,covid2)
rownames(samples) = covid.status

save(samples,covid.status,file=paste0(out.folder,'/',the.method,'-covid-19-fluo-data.rda'))
```

Annex 2: R-code for the data analysis

```
library('randomForest')
library('FactoMineR')
library('plotly')
library('MASS')
library('plot3D')
library('ggplot2')
library('e1071')
library('caTools')
library('caret')
library('factoextra')

the.method = 'Eatio'

#Preparing the data
load(paste0('../Tree Data Covid/',the.method,'-covid-19-fluo-data.rda'))
n.data = dim(samples)[1]
the.data <- data.frame(covid.status, samples)

#### RANDOM FORESTS OOB ####
tree.full = randomForest(covid.status~., data=samples, ntree=50000, mtry=2, importance=T, proximity=T)
tree.full

importance(tree.full)
plot(tree.full)
legend('right',col=c(3,2,1),pch=16,legend = c('OOB','Covid -', 'Covid +'))
varImpPlot(tree.full)


tree.ok = randomForest(covid.status~Cy5+SiR, data=samples, ntree=50000, importance=T)
tree.ok
importance(tree.ok)
plot(tree.ok)
legend('right',col=c(3,2,1),pch=16,legend = c('OOB','Covid -', 'Covid +'))
varImpPlot(tree.ok)


##### LDA using lOCV  #####
the.data <- data.frame(samples)
mLDA <- lda(covid.status~., data=the.data, CV=F)
mLDA
plot(mLDA)
mean(mLDA!=the.data$covid.status)
```