

# A transcriptomic signature predicting septic outcome in patients undergoing autologous stem cell transplantation

**Running title: Predictive transcriptomic signature of sepsis**

*Labiad Yasmine*<sup>1\*</sup>, *Venton Geoffroy*<sup>1,2\*</sup>, *Farnault Laure*<sup>2</sup>, *Baier Céline*<sup>1</sup>, *Colle Julien*<sup>1,2</sup>, *Mercier Cedric*<sup>2</sup>, *Ivanov Vadim*<sup>2</sup>, *Nicolino Corinne*<sup>2</sup>, *Loriod Béatrice*<sup>1</sup>, *Fernandez-Nunez Nicolas*<sup>1</sup>, *Torres Magali*<sup>1</sup>, *Mattei Jean-Camille*<sup>3</sup>, *Rihet Pascal*<sup>2</sup>, *Nguyen Catherine*<sup>1</sup> and *Costello Régis*<sup>1,2 #</sup>

<sup>1</sup> Aix Marseille Univ, INSERM, TAGC, Marseille, France, <sup>2</sup> AP-HM, Service d'Hématologie et de Thérapie Cellulaire de La Conception, Marseille, France, <sup>3</sup> Aix Marseille Univ, INSERM, CRO2, Marseille, France

\* *Equal contribution*

**Paper's word count: 3174**

## **# Corresponding author :**

Prof. Costello Régis (MD., PhD.), AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone number: +33 4 91 38 41 50, Fax number: +33 4 91 36 41 54, E-mail address: [regis.costello@free.fr](mailto:regis.costello@free.fr)

***Complete contact informations***

**Labiad Yasmine PhD** : Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26 Mail : [labiad.yasmine@gmail.com](mailto:labiad.yasmine@gmail.com)

**Venton Geoffroy MD-PhD**: Campus de Luminy, 163 Avenue de Luminy, case 928 Marseille, 13288. AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : [geoffroy.venton@ap-hm.fr](mailto:geoffroy.venton@ap-hm.fr)

**Farnault Laure MD**: AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : [laure.farnault@ap-hm.fr](mailto:laure.farnault@ap-hm.fr)

**Baier Céline PhD**: Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26 Mail : [celine.baier@yahoo.fr](mailto:celine.baier@yahoo.fr)

**Colle Julien MD**: Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288. AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : [julien.colle@ap-hm.fr](mailto:julien.colle@ap-hm.fr)

**Mercier Cédric MD** : AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : [cedric.mercier@ap-hm.fr](mailto:cedric.mercier@ap-hm.fr)

**Ivanov Vadim MD**: AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50 Mail : [vadim.ivanov@ap-hm.fr](mailto:vadim.ivanov@ap-hm.fr)

**Nicolino Corinne MD:** AP-HM Centre Hospitalier Universitaire La Conception, Service d'Hématologie et Thérapie Cellulaire, 147 Boulevard Baille, 13005 Marseille, France. Phone : + 33 4 91 38 41 50  
Mail : [corinne.nicolino@ap-hm.fr](mailto:corinne.nicolino@ap-hm.fr)

**Loriod Béatrice :** Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26 Mail : [beatrice.loriod@inserm.fr](mailto:beatrice.loriod@inserm.fr)

**Fernandez-Nunez Nicolas :** Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26. Mail : [nicolas.fernandez-nunez@inserm.fr](mailto:nicolas.fernandez-nunez@inserm.fr)

**Torres Magali :** Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26. Mail : [magali.torres@univ-amu.fr](mailto:magali.torres@univ-amu.fr)

**Mattei Jean-Camille:** Aix-Marseille Université - INSERM, UMR 911, CRO2, Faculté de Médecine - 27, Boulevard Jean Moulin, Marseille, 13385, France. Phone: +33 4 91 38 00 00 Mail : [jeancamille.mattei@gmail.com](mailto:jeancamille.mattei@gmail.com)

**Rihet Pascal Pr:** Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26. Mail : [pascal.rihet@univ-amu.fr](mailto:pascal.rihet@univ-amu.fr)

**Nguyen Catherine PhD:** Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France Phone : +33 4 91 82 87 26. Mail : [catherine.nguyen@inserm.fr](mailto:catherine.nguyen@inserm.fr)

**Costello Régis Pr:** Aix-Marseille Université - INSERM, UMR1090, TAGC Campus de Luminy, 163 Avenue de Luminy, case 928, Marseille, 13288, France. Service d'hématologie et de thérapie cellulaire APHM, Hôpital de la Conception, Marseille, France. Phone: + 33 4 91 38 41 50 Mail: [regis.costello@free.fr](mailto:regis.costello@free.fr)

## Highlights :

- 1) A transcriptomic signature predicts infection in neutropenic patients
- 2) Expression of eleven genes can be used routinely for sepsis prediction
- 3) Early anti-infectious treatment may improve sepsis outcome
- 4) Identification of a predictive transcriptomic signature is ongoing in leukemia

1 **Abstract**

2 Autologous hematopoietic stem cell transplantation (auto-HSCT) is a standard treatment in multiple  
3 myeloma and relapsed or refractory lymphomas. After auto-HSCT, hematologic reconstitution and  
4 infectious complications are the main two critical issues. Though many patients develop infectious  
5 complications after therapeutic intensification, it remains impossible to predict infection for each  
6 individual. The goal of this work was to determine and identify a predictive transcriptomic signature  
7 of systemic inflammatory response syndrome (SIRS) and/or sepsis in patients receiving auto-HSCT.

8 High throughput transcriptomic and bioinformatics analysis were performed to analyze gene  
9 expression modulation in peripheral blood mononuclear cells (PBMCs) in 21 patients undergoing  
10 auto-HSCT for hematological malignancies (lymphoma or multiple myeloma [MM]).

11 Transcriptomic analysis of PBMCs samples collected just after conditioning regimen identified an  
12 eleven genes signature (*CHAT*, *CNN3*, *ANKRD42*, *LOC100505725*, *EDAR*, *GPAT2*, *ENST00000390425*,  
13 *MTRM8*, *C6orf192* and *LOC10289230* and *XLOC-005643*) that was able to early predict (at least 2 to 7  
14 days before its occurrence) the development of SIRS or sepsis.

15 The possibility of SIRS or sepsis occurrence early prediction (2–7 days before occurrence) opens up to  
16 new therapeutic strategies based on pre-emptive antibiotic and/or antifungal prophylaxis adapted to  
17 the specific risk profile of each patient.

18

19 **Key words:** Systemic Inflammatory Response Syndrome (SIRS); Sepsis; Autologous Hematopoietic  
20 Stem Cell Transplantation (Auto-HSCT); Transcriptomic Analysis; Treatment related mortality (TRM)

21 **Abstract word count: 195**

## 22 **Introduction**

23 Auto-HSCT is based on the administration of myelosuppressive high-dose chemotherapy, followed by  
24 infusion of autologous hematopoietic stem cells to obtain hematologic reconstitution. Hematopoietic  
25 stem cells (HSCs) infusion reduces chemotherapy-induced myelosuppression period and procedure-  
26 related mortality rate below 3% [1–3]. With few exceptions (solid tumors, autoimmune diseases),  
27 auto-HSCT is essentially indicated for selected hematological malignancies and considered as a  
28 standard treatment in young patients with MM and for relapsed or refractory lymphoma.

29 Besides direct toxicity of conditioning regimens, deep (<0.5 G/L neutrophils) and prolonged (usually  
30 7–12 days) neutropenia exposes patients to significant risks of infection. The saprophytic gram-  
31 negative bacilli (such as *Escherichia coli*) are the most common cause of septic shock [4] and chronic  
32 immunosuppression exposes to the risk of fungal infection. An anti-fungal prophylaxis is usually  
33 administered [5] but antimicrobial prophylaxis is less often given because its effectiveness is not  
34 clearly established and increases the risk of *Clostridium difficile* diarrhea [6]. Unfortunately, it is not  
35 possible to foresee which patients will develop a SIRS and/or sepsis. Therefore, it remains impossible  
36 to adjust the antibiotic or antifungal prophylaxis to the specific risk profile of each patient.

37 The main objective of this work was to determine and identify a predictive transcriptomic signature  
38 of the SIRS and/or sepsis in patients receiving auto-HSCT, leading to the possibility of a pre-emptive  
39 anti-infectious treatment.

40

## 41 ***Patients, materials and methods***

### 42 ***Patients***

43 The **prospective** study was approved by the institutional review board of the Assistance Publique des  
44 Hôpitaux de Marseille (AP-HM – AORC2012 2012-08). Written informed consent was obtained from  
45 each patient. Patients were admitted in the hematology department of the Conception university  
46 hospital for undergoing auto-HSCT. All patients were under 65 years and were already followed in  
47 the hematology department for MM or high-grade lymphoma. Before auto-HSCT, patients were in  
48 complete remission (CR) or in partial remission (PR) after conventional chemotherapy. Inclusion  
49 criteria were the same as required for being eligible to auto-HSCT. Twenty four patients were  
50 included in this study protocol. Twenty one were analyzed. Blood samples were collected for each  
51 patient at three moments: before the conditioning regimen (T1), after the conditioning regimen and  
52 before the graft infusion (T2) and at the end of the neutropenic phase (T3).

53 **All patients (after written informed consent) hospitalized in the Hematology and Cellular Therapy**  
54 **Department of the university hospital of Marseille between August 2012 and December 2015 have**  
55 **been included. Only the patients (21 in all) in whom the RNA extraction was a success (after**  
56 **quantification and quality control) have been analyzed.**

57

### 58 ***SIRS and sepsis definitions***

59 SIRS and sepsis definition are based on the American College of Chest Physicians society of Critical  
60 Care Medicine consensus. SIRS is defined as the systemic inflammatory response to a variety of  
61 severe clinical symptoms with at least two of the following criteria: a) Temperature higher than 38 °C  
62 or lower than 36 °C b) Heart rate higher than 90 beats/min c) Respiratory rate higher than 20  
63 breaths/min or PaCO<sub>2</sub> lower than 32 mmHg d) White blood cell counts higher than 12,000 cells/mm<sup>3</sup>  
64 or lower than 4,000 cells/mm<sup>3</sup>, or the presence of more than 10% immature neutrophils. The last  
65 criterion cannot be considered in auto-HSCT context because of the aplasia phase following HSC

66 infusion [7, 8]. Sepsis is defined as SIRS secondary to documented or suspected infection. Patients  
67 with severe sepsis are patients with sepsis and at least one organ dysfunction. Septic choc is defined  
68 by severe sepsis associated with refractory hypotension [8].

69

#### 70 ***Conditioning regimens***

71 Conditioning regimen for therapeutic intensification was high dose melphalan (200 mg/m<sup>2</sup>) for  
72 patients with MM and BEAM for patients with lymphoma (carmustin 300 mg/m<sup>2</sup> at day -6,  
73 etoposide 150 mg/m<sup>2</sup> from day -5 to -2 twice daily, cytarabin 200mg/m<sup>2</sup> from day -5 to -2 twice daily  
74 and melphalan 140mg/m<sup>2</sup> day -1, with auto-HSCT on day 0).

75

#### 76 ***Hematopoietic stem cell collection***

77 HSCs mobilization required hematopoietic growth factors (G-CSF in most cases). Two procedures  
78 have been used to mobilize HSCs. For steady-state collection, G-CSF was injected at the dose of  
79 10µg/kg-a-day. Mobilization chemotherapy consisted in high doses cyclophosphamide (1.5 to 4g/m<sup>2</sup>)  
80 followed by 5µg/kg-a-day of G-CSF. When HSCs collection failed with these usual procedures,  
81 plerixafor (SDF-1/CXCR4 interaction inhibitor) was used at 240 µg/kg.

82

#### 83 ***PBMCs isolation***

84 Blood samples were centrifuged to separate plasma and other blood components. Concentrate blood  
85 was diluted and PBMCs were collected by Ficoll-Hystopaque density-gradient centrifugation [9].

86

#### 87 ***RNA extraction, quantification quality controls***

88 Depending on cell number, two kits were used for RNA extraction, the RNeasy Mini Kit Qiagen™  
89 (Qiagen, Valencia, California), which accepts 10 million of cells and the RNeasy Midi Kit Qiagen™,  
90 which accepts up to 100 million cells. RNAs were extracted according to the Qiagen protocol. RNAs



91 were quantified by NanoDrop 1000 (Nano Drop Technologies, San Diego, CA). Optical density was  
92 measured at 260 and 280 nm and the ratio 260/280 (> 1.8) indicates its purity. Extracted RNA quality  
93 was checked with Agilent 2100 Bioanalyzer™ (Agilent Technologies, Santa Clara, California). A score  
94 on a scale of 0 to 10 was automatically attributed to each sample and corresponded to RNA Integrity  
95 Number (RIN). Sample with a RIN under 7 was discarded.

96

### 97 ***Pangenomic gene expression assay***

98 One hundred nanograms of total RNA was labelled using One-Color Microarray-Based Gene  
99 Expression Analysis: Low Input Quick Amp (LIQA) labelling protocol. 0.6 µg of the purified Cy3  
100 labelled cRNA were hybridized for 17h at 65 °C, at 60 rpm, using the SurePrint G3 human GE 8x60K  
101 V2 chips (Agilent Technologies, Santa Clara, California). Microarrays were composed of 62 928  
102 features. Probes synthesized on chips had a size of 60 nucleotides. Microarrays were washed using  
103 Gene Expression Wash Buffer Kit (Agilent Technologies) and scanned through the standard Agilent  
104 protocol. Data were processed using Feature Extraction software.

105

### 106 ***Differential gene expression analysis***

107 The library AgiND is implemented in R software in order to analyze and visualize data. AgiND was  
108 developed on Bioconductor library model ([tagc.univ-mrs.fr/ComputationalBiology/AgiND/](http://tagc.univ-mrs.fr/ComputationalBiology/AgiND/)) and is  
109 used to diagnose data quality and data-microarray normalization. Quantile method was used to  
110 normalize data; the objective was to homogenize distribution of microarray intensity [10]. A filter  
111 was applied on row data to delete controls, then a second filter was applied to delete genes which  
112 were expressed under the background in at least 80% of samples in each group (SIRS–, SIRS+,  
113 Sepsis+).

114 To test co-factors effects (gender, treatment, infection) on gene expression, GeneANOVA software  
115 was used to perform ANalysis Of Variance (ANOVA) on normalized data to determine an estimation

116 of the contribution of each factor (genes, gender, treatment, infection) in gene expression variation.  
117 Global ANOVA model is given in the following formula:  $Y = \mu + \beta_1 G + \beta_2 T + \beta_3 I + \epsilon$ , where Y is  
118 explained variable,  $\mu$  is global mean,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ , are model coefficients, and G, T, I are the quantitative  
119 variables;  $\beta_1 G$ , is the gene effects,  $\beta_2 T$ , is the treatment effects,  $\beta_3 I$ , the infection effects, and  $\epsilon$  is  
120 the error term [11]. Differential gene expression analysis was performed using linear Models for  
121 Microarray Data (Limma). Limma is a multivariate analysis and takes into account co-factors-effect  
122 tested by ANOVA analysis (treatment and gender) (**Table 2**). In order to determine infection impact  
123 on gene expression and on the heat map, new gene expression was calculated after subtracting  
124 gender and treatment co-factors effects. Unsupervised hierarchical clustering was applied on  
125 adjusted gene expression median adjusted data to group genes and samples, according to their  
126 expression using 'TMeV' (Tigr MultiExperiment Viewer) MeV: MultiExperiment Viewer | Part of the  
127 TM4 Microarray Software Suite [<http://www.tm4.org/mev/>]. Pearson correlation was used, and  
128 clusters were grouped on the basis of average linkage method.

129

### 130 ***Support Vector Machines (SVMs)***

131 SVMs were applied to predict classification of patients according to the predictive signature. SVMs  
132 use training set in which genes known to be related to each other by function or samples related to a  
133 group are described as positive examples and genes or samples known not to be members of that  
134 class are labelled as negative examples. In this study, samples were attributed to two groups: SIRS  
135 and sepsis patients in the first group and patients with no temperature in the second. They were  
136 combined into a set of training examples used by SVM to distinguish class members from non-  
137 members on the basis of expression data. After learning the class expression features, the SVM can  
138 be used to recognize and classify each sample on the basis of their expression [11, 12]. SVMs tool is  
139 implemented in TMeV software MeV: MultiExperiment Viewer | Part of the TM4 Microarray  
140 Software Suite. [<http://www.tm4.org/mev/>].

141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156

**High-throughput quantitative PCR**

RT-PCR was performed using SuperScript™ VILO™ MasterMix protocol described by Invitrogen. Sixty nanogram of RNA of each sample was retrotranscribed, using 4µL of SuperScript™ VILO™ MasterMix, and RNase DNase Free water for a final volume of 20µL. Mix was incubated 10 min at 25 °C, 60min at 42 °C and 5min at 85°C. Quantitative PCR was performed with FlexSix array on BioMark™ HD (Fluidigm). A pre-amplification of each sample was applied according to the protocol provided by Fluidigm. The pre-amplification step was followed by an Exonuclease I treatment (BioLegend) to remove unincorporated primers. The final product was diluted 5-fold using 18µl of TE Buffer (10 mM Tris-HCl, 1mM EDTA). In a second part, 12 x 12 (samples x primers) qRT-PCR reactions are performed for each primer pair on each sample on the 12 x 12 array (FlexSix). We used the EvaGreen detection assay for following standard Fluidigm protocols. Ct values were calculated from the system software Biomark Real-Time PCR Analysis (Fluidigm). The primers used were:

| Gene           | R                             | F                          |
|----------------|-------------------------------|----------------------------|
| <b>SLC18B1</b> | 5'-GCAGGACAGCTTTTCAGTATCAC-3' | 5'-CCTGGCCTTAAACATCACCG-3' |
| <b>CNN3</b>    | 5'-GGCTGGCTCCTTTATTAGTGC-3'   | 5'-AAGCTGGCCAAAGTGAATTG-3' |
| <b>MTMR8</b>   | 5'-ACTTAAGTGGGGAGAGGGGT-3'    | 5'-GAGGCAGACTACTCCAAGCA-3' |
| <b>GPAT2</b>   | 5'-GCCCAGAGAAGCCTACATCA-3'    | 5'-CATCAGAAGCTCCTGGGGGA-3' |
| <b>PLCG1</b>   | 5'-ACAGGAATCTTGGTGCTTCAGT-3'  | 5'-CAGGGAGGTACATGGCCAAT-3' |
| <b>ANKRD42</b> | 5'-CGCCCATGAAAAGCTGCATA-3'    | 5'-GGAGCAAATCTGACAGCCCA-3' |
| <b>GAPDH</b>   | 5'-CCACCACCCTGTTGCTGTGA-3'    | 5'-CCCCTCCTCCACCTTTGAC-3'  |
| <b>EDAR</b>    | 5'-ACATGAGCTGACTGGCTG-3'      | 5'-TGAGTGTGCCATTCCAGGAT-3' |
| <b>TRAV3 -</b> | 5'-GCCTCGGTTGGGGTATTGAA-3'    | 5'-CTCAGCCGGAAGATCAGGTC-3' |

|                     |                            |                            |
|---------------------|----------------------------|----------------------------|
| <b>CHAT</b>         | 5'-CAGCAGAACATCTCCGTCGT-3' | 5'-ATGGCCATTGACAACCACCT-3' |
| <b>LOC100289230</b> | 5'-GTGCTCCGGAAAACGTAAA-3'  | 5'-TCCTGTGCCCGTAATTTCTC-3' |

157

## 158 **Results**

### 159 ***Patient's analysis***

160 Twenty-four patients were included in the study, among these patients; 21 validated the molecular  
161 criteria for transcriptomic analysis. Among these 21 patients, 6 patients did not develop a fever or  
162 SIRS (28.6%), nine patients developed SIRS (42.6%), 5 a sepsis (23.8%) and 1 a severe sepsis (5%).  
163 Patients' clinical data are summarized in the **Table 1**.

164

### 165 ***Pangenomic gene expression***

166 The 21 samples were analyzed on Agilent microarrays 'SurePrint G3 GE 8x60K human.' After 17 hours  
167 of hybridization, the chips were washed and scanned. Results passed microarrays quality controls.  
168 Raw data were transformed into *log2* and normalized with quantile method. 24 046 probes  
169 expressing a higher signal than background (in at least one group) have been selected. Two methods  
170 were used for statistical analysis: ANOVA analysis to measure the impact of each factor on gene  
171 expression variation (infection, gender, treatment) and Limma analysis on three groups (SIRS-,  
172 SIRS+, sepsis +) with the following co-factors: gender and treatment, to define differential gene  
173 expression (**Figure 1**). ANOVA analysis was performed on the 24 046 probes to estimate impact of  
174 each factor (infection, gender, treatment) in gene expression variation. P-value was calculated for  
175 each factor. For all factors (genes, infections, gender, and treatment) p-values were  $<10^{-4}$  (**Table 2**).  
176 Infection F score was the highest, suggesting that 'infection' had the greatest impact on the gene  
177 expression. Genes F score was of 115.45 (p-value  $<10^{-4}$ ) and explains genes fluctuations. Treatment  
178 and gender had an impact on the variation of gene expression with a F score of respectively 176.21  
179 and 156.07 (p-value  $<10^{-4}$ ). Unsupervised hierarchical clustering method was used to classify the  
180 differently expressed genes identified by Limma analysis. Expression similarity profiles of the genes  
181 were grouped on the horizontal axis and samples on the vertical axis. Gene expression profiles are

182 shown in heat map (**Figure 2**). With a FDR (False Discovery Rate) fixed to 5%, 11 genes differentially  
183 expressed were identified between the patients who did not have temperature (SIRS-) and the  
184 patients who developed SIRS and sepsis (taking account effect of gender and treatment). The eleven  
185 differentially expressed genes were: *CHAT*, *CNN3*, *ANKRD42*, *LOC100505725*, *EDAR*, *GPAT2*,  
186 *ENST00000390425*, *MTRM8*, *LOC10289230* and *XLOC-005643* (**Table 3**). Gene expression profiling  
187 distinguished two groups: patients who did not develop temperature, on the one hand, and patients  
188 with SIRS or sepsis, on the other hand. *CHAT*, *CNN3*, *C6orf192*, *ANKRD42*, *LOC100505725*, *EDAR*,  
189 *GPATZ*, *ENST00000390425*, *MTRM8*, *LOC10289230* were over expressed and *XLOC-005643* was  
190 under-expressed in patient samples with SIRS or Sepsis in contrast to patients who did not develop  
191 fever and SIRS.

192

### 193 **SVM**

194 SVM and leave-one-out cross-validation were used to classify patients according to their gene  
195 expression. 21 samples were separated into two groups, the first group was composed of 15 SIRS and  
196 septic patients and the second of the 6 patients who did not have temperature. SIRS and sepsis  
197 patients were considered as positive experiments, and the other as negative. All 15 patients of the  
198 positive experiments were classified as positive, staying in positive class (true positives=15) and none  
199 was transferred from negative class to positive (False negatives=0). On the 6 patients of the negative  
200 experiments, all were classified as negative, all retained in negative class (true negatives=15) and  
201 none was recruited into negative class from positives (false positives=0). **We performed a Fisher  
202 exact test on our results; the p-value = 1.84 E-5.**

203

### 204 **RT-qPCR**

205 In order to confirm the microarray gene expression results, RT-qPCR was performed on 9 genes that  
206 composed the transcriptomic signature. The *LOC 100289230* gene was differently expressed (p-value

207 = 0.003) in patients who developed SIRS or sepsis compared with patients who did not. *TRAV3*, *EDAR*,  
208 *PLCG1-AS1-001*, *GPAT2*, *MTRM8*, *CNN3*, and *SLC18B1* were also differentially expressed with p-values  
209 of 0.004, 0.01, 0.01, 0.005, 0.009, 0.09 and 0.01, respectively (Figure 3).

210

211

## 212 ***Discussion and conclusion***

213 This clinical study identified eleven genes significantly and differentially expressed in patients who  
214 developed SIRS or sepsis after the conditioning regimen for auto-HSCT. Ten of them were up-  
215 regulated (*CHAT*, *CNN3*, *ANKRD42*, *LOC100505725*, *EDAR*, *GPATZ*, *ENST00000390425*, *MTRM8*,  
216 *C6orf192* and *LOC10289230*) while only one was down-regulated (*XLOC-005643*). All patients with  
217 this specific transcriptomic signature developed a SIRS or sepsis within 48 hours (range 48 hours-7  
218 days) following conditioning regimen. All patients were classified in the right group according to their  
219 gene expression and based on SVMs analysis. After a wide scientific literature review, genes  
220 composing our predictive signature are not directly involved in sepsis or infection pathways. In this  
221 cohort of 21 patients, 9 developed a SIRS and 5 a sepsis. Only one patient developed a severe sepsis,  
222 and no patient had septic shock, thus impeding the possibility to identify a specific transcriptomic  
223 signature predictive for these life-threatening conditions. In addition, since a SIRS/sepsis predictive  
224 signature before the conditioning regimen was not identified, the transcriptomic signature was not  
225 linked to the patient pre-auto HSCT status but depended on the conditioning regimen patient's  
226 response.

227 We wondered about the definition of sepsis and SIRS and the clustering of sirs and sepsis;  
228 first, white blood cells count criteria is not relevant in patients in deep aplasia. Furthermore, in the  
229 same way, patients in deep aplasia present grade II-III anemia especially after BEAM conditioning  
230 regimen. In patients with no cardiovascular and/or pulmonary comorbidity, hemoglobin until 8g/dl is

231 tolerated. However, that often results in an increase of the heart rate higher than 90 beats/min  
232 and/or respiratory rate higher than 20 breaths/min, especially in patients with fever.  
233 Fever is neither sensitive or specific in conventional patients. But, in our severely  
234 immunocompromised patients fever is more sensitive and specific than in immunocompetent  
235 patients. Nevertheless, it is not unusual to start an antibiotic therapy in patients with no fever but  
236 with a microbiological documentation and/or a major increase of the C reactive protein only.  
237 At last, we have been very drastic on the definition of our sepsis patients group. Only the patients  
238 with fever and a microbiological documentation have been considered in sepsis. Anyway, our  
239 immunocompromised patients are paucisymptomatic and very few infections were clinically  
240 probable.

241 Our transcriptomic signature predicts SIRS/sepsis profiles and is more robust than the main  
242 confounding factors, such as conditioning regimen, type or gender. BEAM conditioning regimen was  
243 more myelosuppressive than melphalan alone and men have had a more significant risk to develop  
244 infectious complications than women [13]. Auto-HSCT patients affected by MM are conditioned by  
245 high dose of melphalan while patients with lymphoma by high dose BEAM chemotherapy, meaning  
246 that treatment and pathology are confounding factors. SAM and ANOVA analyses were performed  
247 on the data of patient samples before chemotherapy, and no differentially expressed genes were  
248 found between patients affected by lymphoma or MM.

249 This work proposes a transcriptomic approach of the sepsis issue during auto-HSCT  
250 neutropenic phase. Vanska et al. have shown that high pentraxin 3 level predicted septic shock and  
251 bacteremia at the onset of febrile neutropenia after intensive chemotherapy of hematologic patients  
252 [14]. Nonetheless, high pentraxin 3 level had only a predictive value for septic shock in patients who  
253 already had a febrile neutropenia. In contrast, our predictive transcriptomic signature identifies  
254 patients who have a major risk to develop SIRS and/or sepsis at least 48 hours (range 48 hours–7  
255 days) before onset of fever. In order to confirm our results, our transcriptomic signature has been



256 validated on a 10 patient's prospective validation cohort (data not shown). Since total gene  
257 expression analysis by microarrays is not possible in clinical routine for time and cost-effective  
258 reasons, RT-qPCR analysis on a blood sample of the eleven deregulated genes will have to be  
259 developed in order to assess the possibility of a routine use.

260

261           Prophylactic anti-fungal and/or anti-microbial treatments have not proven a high efficiency  
262 regarding the outcome in patients undergoing highly myeloablative chemotherapy. For example, the  
263 use of ciprofloxacin induces a 20% reduction of infection rates but lead to 70% increase in the  
264 development of ciprofloxacin resistance [15]. Early identification of patients who will develop SIRS  
265 and/or sepsis could perhaps contribute to a better use of anti-infectious agents as preemptive  
266 treatment instead of a prophylactic treatment, although this is only an hypothesis by should be  
267 tested in a prospective clinical trial. A limitation of our study, directly related to the small number of  
268 patients and of infectious events, is the impossibility to identify a specific profile for microbial or  
269 fungal or viral infections, in order to precisely orientate the preemptive treatment. Finally,  
270 identification of high vs low infectious risk patients after myeloablative chemotherapy, could help to  
271 select patients with a safe septic profile and allows an early hospital discharge or even outpatient  
272 based blood stem cell transplantation, in order to develop a more rational utilization of hospital  
273 resources but still maintaining optimal safety conditions [16-19]. A validation of this transcriptomic  
274 signature in a prospective and larger cohort of patients is necessary to have a stronger clinical  
275 impact. Another strategy (work in progress) is to validate in RT-qPCR only the eleven genes of  
276 interest in a prospective and larger cohort.

277

278 ***Conflict of interest***

279 All authors declare no financial conflict of interest

280

281 ***Funding***

282 This work was supported by grants from AP-HM, Assistance Publique – Hôpitaux de Marseille (AORC

283 APHM Junior), and Aix-Marseille Université, Advanced BioDesign. TGML platform supported by the

284 France Génomique

285

286 ***Acknowledgments***

287 We would like to thank INSERM, Aix-Marseille Université and Advanced BioDesign (Parc

288 Technologique de Lyon Bâtiment Cèdre 1, 97 Allée Alexandre Borodine, 69 800 Saint-Priest – Lyon)

289 for grants.

290 We are grateful to the patients who gave their informed consent to the use of their samples for

291 research. We are grateful to the IBiSA Transcriptomics and Genomics Marseille-Luminy (TGML)

292 platform was supported by the France Génomique National infrastructure. We thank Laurence Borge

293 for assistance and the use of the cell culture platform facilities (CRCM U1068, Marseille) Plateforme

294 de Culture Cellulaire de Marseille-Luminy, bâtiment TPR2 163 Avenue de Luminy, 13009 Marseille.

295 We also thank Geneviève Victorero and Noushine Mossadegh-Keller for technical supports.

296

297

- 299 [1] Kurnick N. Autologous and Isologous Bone Marrow Storage and Infusion in the Treatment of  
300 Myelo-Suppression. *Transfusion*. 1962;2:178-187.
- 301 [2] MCFARLAND W, GRANVILLE NB, DAMESHEK W. Autologous bone marrow infusion as an  
302 adjunct in therapy of malignant disease. *Blood*. 1959;14:503-521.
- 303 [3] Clifford P, Clift R, Duff J. Nitrogen-mustard therapy combined with autologous marrow  
304 infusion. *The Lancet*. 1961;277:687-690.
- 305 [4] Veeraputhiran M, Jain T, Deol A, et al. BEAM Conditioning Regimen Has Higher Toxicity  
306 Compared With High-Dose Melphalan for Salvage Autologous Hematopoietic Stem Cell  
307 Transplantation in Multiple Myeloma. *Clinical lymphoma, myeloma & leukemia*. 2015;15:531-535.
- 308 [5] Maertens J, Marchetti O, Herbrecht R, et al. European guidelines for antifungal management  
309 in leukemia and hematopoietic stem cell transplant recipients: summary of the ECIL 3—2009 update.  
310 *Bone marrow transplantation*. 2011;46:709-718.
- 311 [6] Sohn BS, Yoon DH, Kim S, et al. The role of prophylactic antimicrobials during autologous  
312 stem cell transplantation: a single-center experience. *European journal of clinical microbiology &  
313 infectious diseases* : official publication of the European Society of Clinical Microbiology.  
314 2012;31:1653-1661.
- 315 [7] Physicians ACoC. Definitions for sepsis and organ failure and guidelines for the use of  
316 innovative therapies in sepsis. *Critical care medicine*. 1992;20:864-874.
- 317 [8] Robertson CM, Coopersmith CM. The systemic inflammatory response syndrome. *Microbes  
318 and Infection*. 2006;8:1382-1389.
- 319 [9] English D, Andersen BR. Single-step separation of red blood cells, granulocytes and  
320 mononuclear leukocytes on discontinuous density gradients of Ficoll-Hypaque. *Journal of  
321 immunological methods*. 1974;5:249-252.
- 322 [10] Bolstad BM, Irizarry RA, Åstrand M, Speed TP. A comparison of normalization methods for  
323 high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19:185-193.
- 324 [11] Brown MP, Grundy WN, Lin D, et al. Knowledge-based analysis of microarray gene expression  
325 data by using support vector machines. *Proceedings of the National Academy of Sciences*.  
326 2000;97:262-267.
- 327 [12] Quackenbush J. Computational analysis of microarray data. *Nat Rev Genet*. 2001;2:418-427.
- 328 [13] Adrie C, Azoulay E, Francois A, et al. Influence of gender on the outcome of severe sepsis: a  
329 reappraisal. *CHEST Journal*. 2007;132:1786-1793.
- 330 [14] Vänskä M, Koivula I, Hämäläinen S, et al. High pentraxin 3 level predicts septic shock and  
331 bacteremia at the onset of febrile neutropenia after intensive chemotherapy of hematologic  
332 patients. *Haematologica*. 2011;96:1385-1389.
- 333 [15] Pohlen M, Marx J, Mellmann A, et al. Ciprofloxacin versus colistin prophylaxis during  
334 neutropenia in acute myeloid leukemia: two parallel patient cohorts treated in a single center.  
335 *Haematologica*. 2016;haematol. 2016.147934.
- 336 [16] Ferrara F, Palmieri S, Viola A, et al. Outpatient-based peripheral blood stem cell  
337 transplantation for patients with multiple myeloma. *The hematology journal : the official journal of  
338 the European Haematology Association / EHA*. 2004;5:222-226.
- 339 [17] Mank A, van der Lelie J, de Vos R, Kersten MJ. Safe early discharge for patients undergoing  
340 high dose chemotherapy with or without stem cell transplantation: a prospective analysis of clinical  
341 variables predictive for complications after treatment. *Journal of clinical nursing*. 2011;20:388-395.
- 342 [18] Faucher C, Soriano ALC, Esterni B, et al. Randomized study of early hospital discharge  
343 following autologous blood SCT: medical outcomes and hospital costs. *Bone marrow transplantation*.  
344 2012;47:549-555.
- 345 [19] Martino M, Morabito F. Autologous stem cell transplantation in multiple myeloma is not  
346 dead but alive and well. *Expert opinion on biological therapy*. 2015;15:149-154.

348 **Table legends:**

349 **Table 1: Patients, pathologies and auto-HSCT characteristics**

350 **Table 2: co-factors effect on gene expression.**

351 ANOVA analysis was performed to test the co-factors impact (infection, treatment and gender) on  
352 gene expression for samples taken at T2 (after chemotherapy). With p-values <0, 05 infection,  
353 treatment and gender have a significant effect on the gene expression. DF: Degrees of Freedom

354 **Table 3: Differentially expressed gene characteristics**

355 Ch: Chromosome, +1: Forward strand, from 3' to 5', -1: Reverse strand, from 5' to 3'

356

357

358 **Figure legends:**

359 **Figure 1: Differential gene expression flow chart**

360 Pangenomic array for the transcriptomic analysis have been used, data were filtered and the controls  
361 were suppressed, cofactors effects were tested based on their implication on variation of gene  
362 expression. A multivariate analysis – LIMMA – using the cofactors already tested was performed to  
363 define the gene differentially expressed using Bioconductor library. The LIMMA model is given in the  
364 following formula:  $Y = \alpha + \beta_1 \cdot T + \beta_2 \cdot G + \beta_3 \cdot I + \epsilon$ , where Y is the explained variable,  $\alpha$  is the global  
365 mean,  $\beta_1 \cdot T$ ,  $\beta_2 \cdot G$ ,  $\beta_3 \cdot I$ , are the model coefficients, T, G, I, are the quantitative variables,  $\beta_1 \cdot T$  is the  
366 treatment effect,  $\beta_2 \cdot G$  is the gene effect,  $\beta_3 \cdot I$  is the infection effect and  $\epsilon$  is the error term. Then, we  
367 performed multi-testing correction by fixing the threshold to 5%. Finally, we adjusted the expression  
368 data based on the co-factors before performing the hierarchical clustering \*: Linear Models for  
369 Microarray data.

370

371 **Figure 2: Gene expression profile adjusted on the effects of the cofactors**

372 Microarray analysis revealed 11 genes differentially expressed. This genes cluster distinguished SIRS  
373 or sepsis patients from patients who didn't develop any fever or SIRS.

374

375 **Figure 3: Transcriptomic validation signature by RT-qPCR**

376 Relative gene expression composed the transcriptomic signature of patients who developed SIRS or  
377 sepsis compared to patients who did not develop fever and SIRS

378 \* = 0.05; \*\* = 0.01; \*\*\* ≤0.01, NS = not significant

379

Table 1 : Patients, pathologies and auto-HSCT characteristics

|   |                              |
|---|------------------------------|
| <b>Baseline and Demographic Characteristics</b>           | n = 21                       |
| <b>Patients age (median)</b>                              | 58 years<br>Range [32-68]    |
| <b>Patient gender</b>                                     |                              |
| Male  | 14 (67%)                     |
| Female  | 7 (33%)                      |
| <b>Hematologic malignancy</b>                             |                              |
| Lymphoma  | 2                            |
| Multiple Myeloma  | 19                           |
| <b>Conditioning Regimen</b>                               |                              |
| BEAM  | 2                            |
| Melphalan high dose                                       | 19                           |
| <b>Injected CD 34<sup>+</sup> (x 10<sup>6</sup> / kg)</b> | 2.82<br>Range [0.751-3.64]   |
| <b>Mean duration of fever</b>                             | 3 days<br>SEM [-4.30 - 4.30] |
| <b>Mean duration of neutropenia</b>                       | 9 days<br>SEM [-2.90 – 2.90] |
| <b>Documented bacterial infections</b>                    | 6 (30%)                      |
| Gram-positive bacteria                                    | 75%                          |
| <i>Staphylococcus hominis</i>                             | 1                            |
| <i>Staphylococcus epidermidis</i>                         | 2                            |
| <i>Staphylococcus aureus</i>                              | 1                            |
| <i>Streptococcus mitis</i>                                | 1                            |
| <i>Clostridium difficile</i>                              | 1                            |
| Gram-negative bacteria                                    | 25%                          |
| <i>Escherichia coli</i>                                   | 1                            |
| <i>Klebsiella pneumonia</i>                               | 1                            |
| <b>Documented fungal infections</b>                       | 0 (0%)                       |
| <b>Patient classification</b>                             |                              |
| SIRS-   | 6 (29%)                      |
| SIRS+   | 9 (42%)                      |
| Sepsis+   | 5 (24%)                      |
| Severe sepsis+  | 1 (5%)                       |
| Septic shock+   | 0 (0%)                       |

**Table 2: Co-factors effect on gene expression.**

ANOVA analysis was performed to test the co-factors impact (infection, treatment and gender) on genes expression for samples taken at T2 (after chemotherapy). With p-values < 0, 05 infection, treatment and gender have a significant effect on gene expression. DF: Degrees of Freedom

| Factors   | Sum of squares | DF     | Variance | F score | P-value            |
|-----------|----------------|--------|----------|---------|--------------------|
| Genes     | 3476905.49     | 28114  | 123.67   | 115.45  | < 10 <sup>-4</sup> |
| Infection | 954.39         | 1      | 954.39   | 890.96  | < 10 <sup>-4</sup> |
| Treatment | 188.75         | 1      | 188.75   | 176.21  | < 10 <sup>-4</sup> |
| Gender    | 167.18         | 1      | 167.18   | 156.07  | < 10 <sup>-4</sup> |
| Residual  | 602328.16      | 562297 | 1.07     | -       | -                  |
| Total     | 4080543.97     | 590414 | 6.91     | -       | -                  |

**Table 3: Differentially expressed gene characteristics.**

Ch : Chromosome, +1 : Forward strand, from 3' to 5', -1 : Reverse strand, from 5' to 3'

| Gene   | Complete name  | Ensembl ID      | Location                       | Sens |
|--|--|-----------------|--------------------------------|------|
| <b>SLC18B1 / C6ORF192</b>                              | solute carrier family 18 member B1                       | ENSG00000146409 | Chr 6: 132,769,370-132,798,553 | -1   |
| <b>CNN3</b>  | calponin 3   | ENSG00000117519 | Chr 1: 94,896,949-94,927,278   | -1   |
| <b>MTMR8</b>   | myotubularin related protein 8                           | ENSG00000102043 | Chr X: 64,268,081-64,395,431   | -1   |
| <b>LOC100505725<br/>PLCG1-AS1-001 /<br/>TOP1-AS1 /</b> | PLCG1 antisense RNA 1                                    | ENSG00000226648 | Chr 20: 41,098,329-41,138,003  | -1   |
| <b>CHAT</b>  | choline O-acetyltransferase                              | ENSG00000070748 | Chr 10: 49,609,095-49,665,104  | 1    |
| <b>EDAR</b>  | ectodysplasin A receptor                                 | ENSG00000135960 | Chr 2: 108,894,471-108,989,372 | -1   |
| <b>ANKRD42</b>   | ankyrin repeat domain 42                                 | ENSG00000137494 | Chr 11: 83,193,739-83,260,694  | 1    |
| <b>GPAT2</b>   | glycerol-3-phosphate<br>acyltransferase 2, mitochondrial | ENSG00000186281 | Chr 2: 96,021,946-96,039,451   | -1   |
| <b>ENST00000390425/<br/>TRAV3</b>                      | T cell receptor alpha variable 3<br>(gene/pseudogene)    | ENSG00000211777 | Chr 14: 21,723,713-21,724,321  | 1    |
| <b>LOC100289230</b>                                    | NS   | NS              | Chr 5: 98,929,134-98,931,009   | 1    |
| <b>XLOC_005643</b>                                     | linc-CMAHP-1:1 / linc-FAM65B-<br>1/RP3-425P12.2          | ENSG00000230372 | Chr 6: 25,061,853-25,063,735   | -1   |



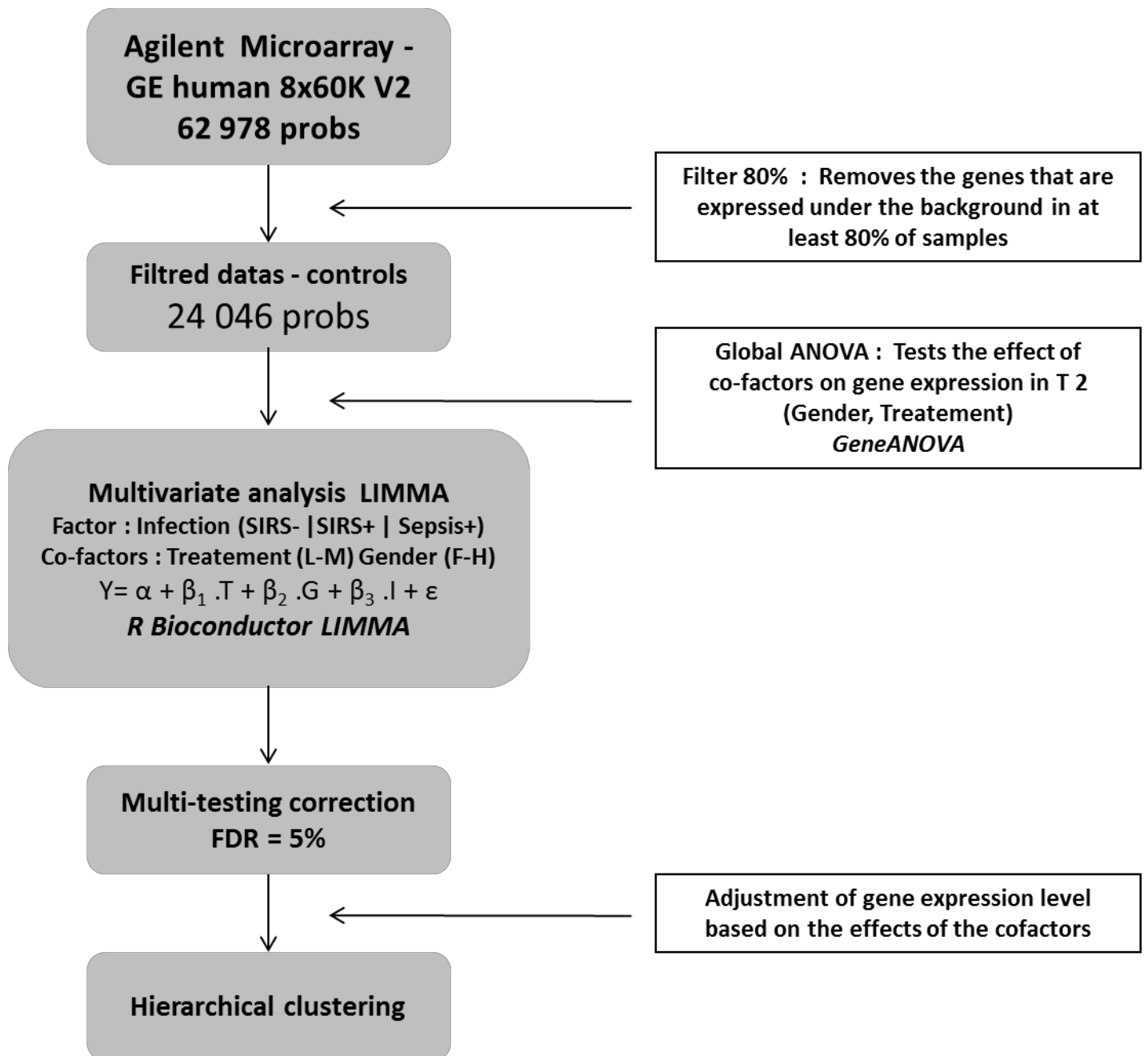


Figure 1: Differential genes expression flow-chart

Figure2

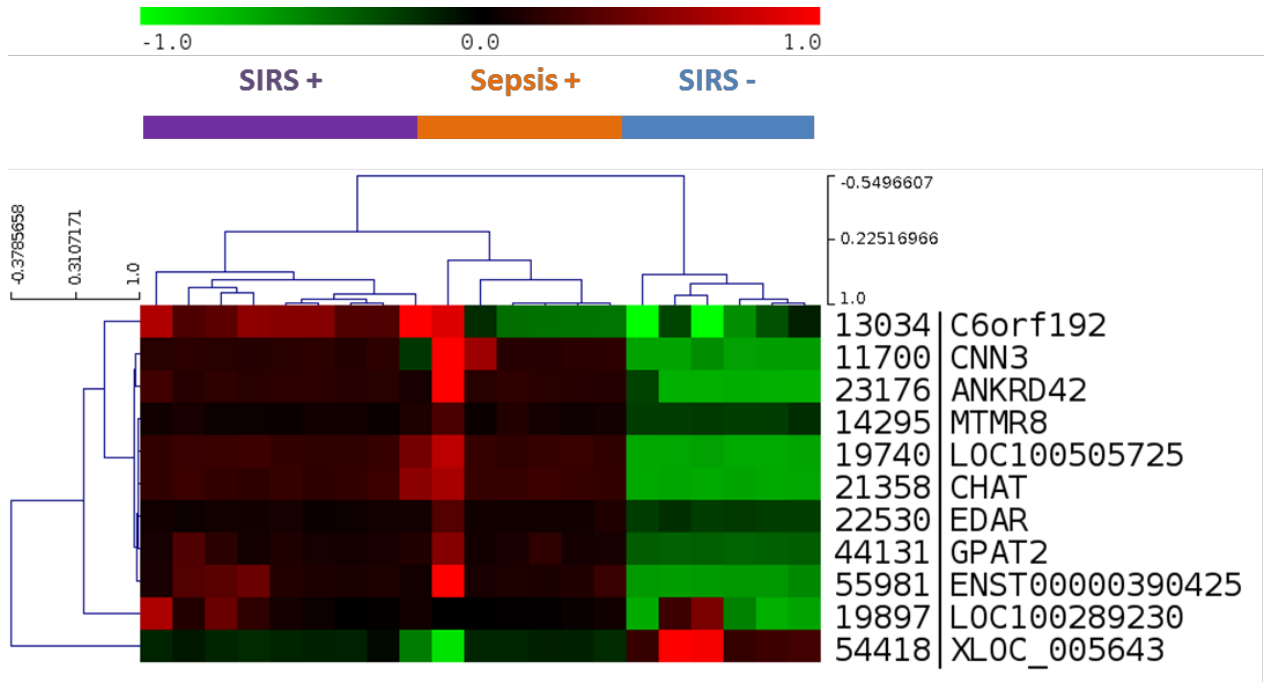


Figure 2: Gene expression profile adjusted on the effects of the cofactors

Color should be used

Figure3

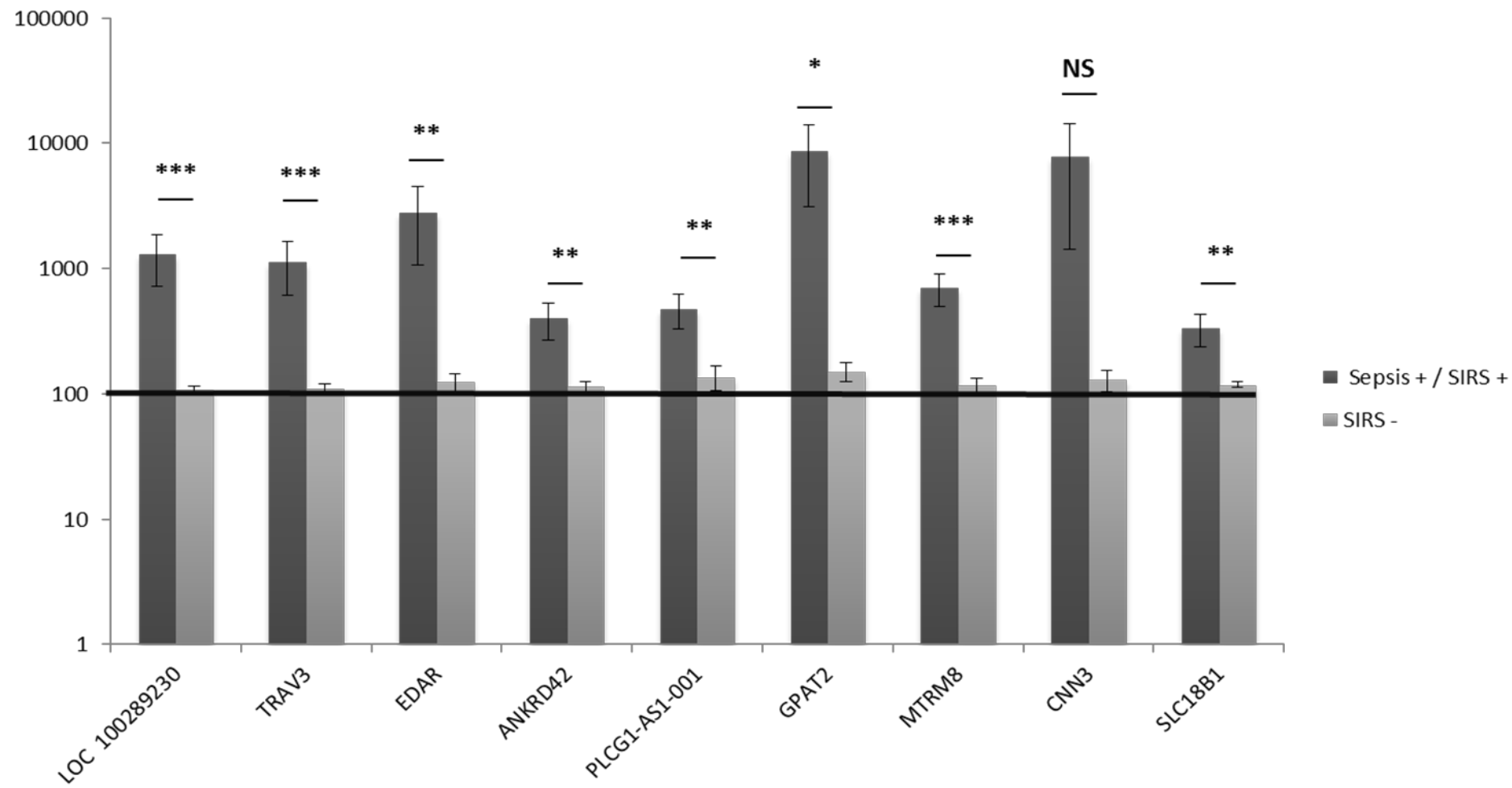


Figure 3 : Transcriptomic validation signature by RT-qPCR