SUPPLEMENTARY INFORMATION

1- Selection of Affymetrix Probesets associated with Progression Free Survival

Affymetrix robust multi-array average (1) data were filtered to retain only the 10% (1979) probesets displaying the highest inter-sample variation coefficient.

A) Selection of probesets associated with PFS in all patients

We first selected the probesets that showed an association with PFS in Cox models adjusted for IPI and treatment effects, with a p value <0.1. We decided to use a low stringency *p* value in this first screening step to avoid the too early elimination of potential biomarkers.This selection yielded a list of 204 candidate probesets. This list of probesets was filtered to exclude the probesets showing less than 0.5 (log_2) differential mean expression in patients with *versus* without progression. We also excluded 5 probesets associated with sex origin (XIST probeset 214218_s_at and 4 probesets corresponding to genes located on chromosome Y). A list of 42 probesets, corresponding to 32 distinct known genes, was selected with this procedure. It is to note that the Cox models for PFS adjusted on IPI and treatment effects indicated a *p* value between 0.05 and 0.1 for ten of these 32 genes (COPA, VSIG4, BCL7A, CCL18, BHLHB3, SPRED2, FLJ20647, BCL6, BHLHB2, ANKRD15).

B) Selection of probesets which may indicate a specific effect in R-CHOP treated patients

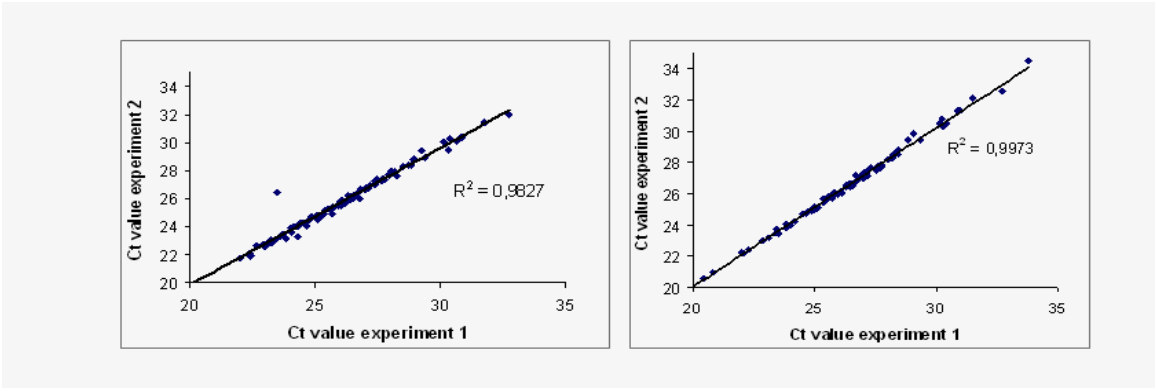
We tested the interaction between gene expression and treatment groups and selected probesets showing a significantly different ($p < 0.05$) association with PFS in CHOP versus R-CHOP patients. This selection yielded a list of 111 probesets. This list was filtered to exclude probesets showing less than 0.5 (log_2) differential mean expression in R-CHOP treated patients with *versus* without progression. 39 probesets, corresponding to 34 distinct genes, were selected with this procedure.

2- TLDA design

This TLDA comprised 96 assays (Supplementary Table). These assays were selected to evaluate transcripts that showed a differential effect on the PFS in CHOP and R-CHOP patients (PFS interaction, n=33), or that showed an association with PFS in all patients of the screening set (PFS global, n=33). The TLDA included assays corresponding to housekeeping genes (n=6) and additional genes (n=26) previously reported to be associated with OS or to discriminate GC and ABC samples (2-8). Three of the housekeeping genes (18S, TBP and PGK1) were selected because of their usual use in quantitative RT-PCR experiments, and three of them (BRF2, FBXO7, MGC15396) were selected because of their very low inter-sample variation coefficient and low, medium or high expression levels in the Affymetrix screening data set.

3- Reproducibility of the data obtained with TLDA

Overall, we observed a high reproducibility of the raw Ct values obtained in two different TLDA runs for the 11 samples tested in duplicates.



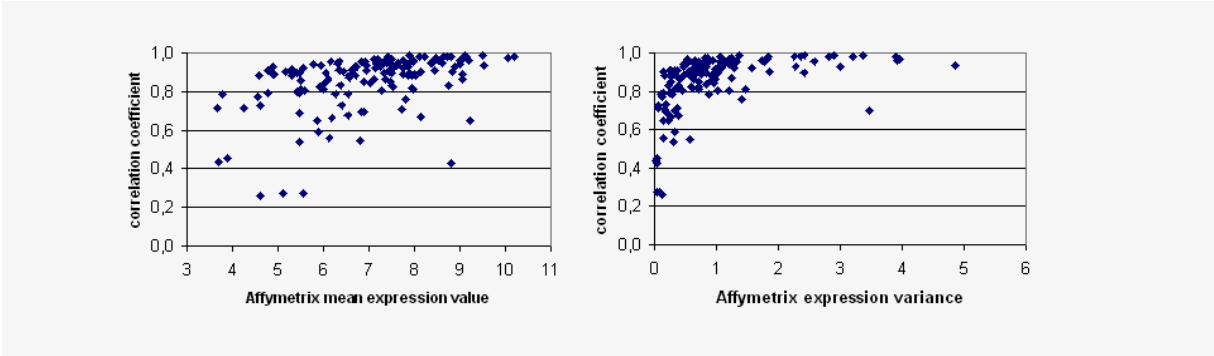
The above Figure shows the raw Ct values obtained for the same sample run in two different TLDA experiments. The left plot shows the sample with the lowest correlation coefficient, the right plot shows the sample with the highest correlation coefficient. Ct values for 18S RNA are excluded of the plot for clarity of the figure.

4- Normalization to housekeeping genes

The TLDA comprised 6 house keeping genes. To select the best normalization procedure, we used the 11 samples which had been run in two different TLDA experiments. We computed the intra-class correlation coefficient for each duplicate after normalization (delta Ct = Ct gene – mean Ct housekeeping), where mean Ct housekeeping was the arithmetic mean or geometric mean of the 63 possible combinations of the 6 housekeeping genes (18S, TBP, PGK1, BRF2, FBXO7, MGC15396) and averaged this coefficient for the 11 duplicates. The mean intra-class coefficient ranged from 0.989 to 0.993, and was at its highest value and lowest SD when the arithmetic mean of the 6 housekeeping genes was used to normalize the data. The data obtained with the geometric mean were similar, but more sensitive to the combination of housekeeping genes retained in the combination.

5- Correlation of Affymetrix and Taqman Low Density Array data

Gene expression data obtained with Affymetrix hybridization microarrays and real-time PCR TLDA in the 23 screening samples were overall highly correlated The mean correlation coefficient calculated for 87 genes and 142 probesets was 0.87, although –as could be expected- the correlation coefficient dropped for genes weakly expressed or with a low inter-sample variability.



In the above Figure, Pearson linear correlation coefficient of the gene expression data obtained in TLDA versus Affymetrix experiments are plotted according to the Affymetrix probeset mean expression level (left panel) or Affymetrix probeset mean variance (right panel) across the 23 R-CHOP screening samples. Coefficients were calculated for 87 genes and 142 probesets, excluding housekeeping genes and assays which did not target the same reference transcript.

6 – Cox model with L1 penalty (lasso) and path algorithm procedure

We used a Cox model with L1 penalty known as lasso (least absolute shrinkage and selection operator) algorithm (9, 10) which allows simultaneously to select variables and shrink regression coefficients to control overfitting. This method was shown to have much better performance than other approaches to predict survival in a recent microarray comparative study (11) and was used to reanalyse different Diffuse Large B-Cell Lymphoma published datasets (12).

The lasso was computed using the sixteen genes associated with survival in the 67 R-CHOP patients in univariate analysis, starting from an initial model that included the International Prognostic Index. The Lambda parameter controls the penalty applied to the regression coefficients and the gene entry in the lasso. When Lambda is maximum, no gene is selected. When Lambda is null, all genes are entered and the lasso is equivalent to the corresponding Cox model. To select the optimal step, we computed the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) using either the number of events (13) (n=21) or the number of patients (n=67). These different criteria did not allow us to identify an unambiguous optimal step. Therefore we computed at each step the lasso C index and the non penalized Cox C index. To evaluate the potential overfitting of these models, we estimated the bias of the Cox C index scores by a five-fold cross-validation procedure (14). As shown in the table below, the lasso and Cox cross-validated C indexes exhibited similar values until the step 9, suggesting that the lasso was probably not overfitted. However after that step, the lasso and the non penalized Cox gave similar C index scores, which were much higher than the cross-validated ones indicating an increased overfitting bias. We therefore considered that the 6 genes model (APOBEC3G, MME, LMO2, RAB33A, LPP and FOXP1) associated with IPI had the best predictive efficiency in this dataset without overfitting bias.

Lasso model								Cox models			
Step	Lambda	Entered variable	AIC	BIC (n=21)	BIC (n=67)	C index	C index (sd)	C index	C index (sd)	cross-validation Bias	cross-validated C index
0	15.51	IPI	160.03	161.08	162.24	0.62	0.12	0.62	0.12	0.02	0.60
1	11.99	APOBEC3G	156.86	158.95	161.27	0.68	0.11	0.71	0.11	0.03	0.68
2	11.99		156.86	158.95	161.27	0.68	0.11				
3	10.57	MME	156.17	159.30	162.78	0.71	0.11	0.77	0.11	0.01	0.76
4	8.17	LMO2	152.48	156.66	161.30	0.77	0.10	0.79	0.09	0.04	0.76
5	8.11		152.34	156.52	161.16	0.77	0.10				
6	8.11		152.34	156.52	161.16	0.77	0.10				
7	7.04	RAB33A	151.69	156.91	162.71	0.77	0.10	0.80	0.09	0.05	0.74
8	6.88	LPP	153.23	159.49	166.46	0.77	0.10	0.83	0.07	0.06	0.77
9	6.56	FOXP1	154.12	161.43	169.56	0.78	0.09	0.84	0.08	0.05	0.79
10	3.96	VNN2	147.49	155.84	165.12	0.83	0.08	0.84	0.08	0.07	0.78
11	3.92	JAK2	149.35	158.75	169.19	0.83	0.08	0.85	0.10	0.07	0.77
12	3.92		149.35	158.75	169.19	0.83	0.08				
13	2.93	PLAU	147.95	158.40	170.00	0.84	0.08	0.84	0.10	0.10	0.74
14	2.20	ANKRD15	147.82	159.31	172.07	0.84	0.09	0.83	0.10	0.09	0.75
15	1.45	MYBL1	146.57	159.10	173.03	0.84	0.09	0.85	0.08	0.13	0.72
16	1.43		146.50	159.03	172.95	0.84	0.09				
17	1.43		146.50	159.03	172.95	0.84	0.09				
18	1.02	BCL7A	147.01	160.59	175.67	0.84	0.09	0.85	0.09	0.14	0.71
19	1.02		147.01	160.59	175.67	0.84	0.09				
20	0.36	SYPL	147.12	161.74	177.99	0.85	0.09	0.85	0.09	0.16	0.68
21	0.31	CSTA	149.01	164.68	182.08	0.85	0.09	0.84	0.10	0.20	0.64
22	0.31		149.01	164.68	182.08	0.85	0.09				
23	0.31		149.01	164.68	182.08	0.85	0.09				
24	0.28	PDE4B	150.92	167.64	186.20	0.84	0.09	0.85	0.09	0.21	0.64
25	0.00	RAFTLIN	152.51	170.27	189.99	0.84	0.10	0.84	0.10	0.19	0.65

References

1. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U*, et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England)* 2003 Apr; 4(2): 249-264.
2. Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D*, et al.* Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *The New England journal of medicine* 2004 Apr 29; 350(18): 1828-1837.
3. Ramuz O, Bouabdallah R, Devilard E, Borie N, Groulet-Martinec A, Bardou VJ*, et al.* Identification of TCL1A as an immunohistochemical marker of adverse outcome in diffuse large B-cell lymphomas. *International journal of oncology* 2005 Jan; 26(1): 151-157.
4. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI*, et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *The New England journal of medicine* 2002 Jun 20; 346(25): 1937-1947.
5. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC*, et al.* Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature medicine* 2002 Jan; 8(1): 68-74.
6. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* 2003 Aug 19; 100(17): 9991-9996.
7. Banham AH, Connors JM, Brown PJ, Cordell JL, Ott G, Sreenivasan G*, et al.* Expression of the FOXP1 transcription factor is strongly associated with inferior survival in patients with diffuse large B-cell lymphoma. *Clin Cancer Res* 2005 Feb 1; 11(3): 1065-1072.
8. Barrans SL, Fenton JA, Banham A, Owen RG, Jack AS. Strong expression of FOXP1 identifies a distinct subset of diffuse large B-cell lymphoma (DLBCL) patients with poor outcome. *Blood* 2004 Nov 1; 104(9): 2933-2935.
9. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J R Statist Soc B* 1996; 58(1): 267-288.
10. Park MY, Hastie T. L1-regularization path algorithm for generalized linear models. *J R Statist Soc B* 2007; 69(4): 659-677.
11. Bovelstad HM, Nygard S, Storvold HL, Aldrin M, Borgan O, Frigessi A*, et al.* Predicting survival from microarray data--a comparative study. *Bioinformatics* 2007 Aug 15; 23(16): 2080-2087.
12. Segal MR. Microarray gene expression data with linked survival phenotypes: diffuse large-B-cell lymphoma revisited. *Biostatistics (Oxford, England)* 2006 Apr; 7(2): 268-285.
13. Volinsky CT, Raftery AE. Bayesian information criterion for censored survival models. *Biometrics* 2000 Mar; 56(1): 256-262.
14. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005 Aug 1; 21(15): 3301-3307.