22

nwCompare and AutoCompare Softwares for Proteomics and Transcriptomics Data Mining – Application to the Exploration of Gene Expression Profiles of Aggressive Lymphomas

Fréderic Pont[1,2,3], Marie Tosolini[1,2,3],
Bernard Ycart[4] and Jean-Jacques Fournié[1,2,3]
[1]*INSERM UMR1037-Cancer Research Center of Toulouse*
[2]*ERL 5294 CNRS, BP3028, CHU Purpan, Toulouse,*
[3]*Université Toulouse III Paul-Sabatier, Toulouse,*
[4]*Laboratoire Jean Kuntzmann, CNRS UMR 5224,*
Université Joseph Fourier, Grenoble
France

1. Introduction

The global protein and gene expression profiling technologies have revolutionized the study of normal and malignant cells. Transcriptomes permitted to delineate subtypes of B-cell lymphomas which were otherwise histologically and clinically undistinguishable. Although the data mining of proteomes or transcriptomes from these malignant cells can unveil new aspects of their biology, tools to simultaneously compare several samples are scarce. Here we depict nwCompare and Autocompare, two new freewares we developed with this aim, and examplify their use for the comparative data mining of transcriptomes from normal human B cells and B cell lymphomas such as follicular lymphomas (FL) and diffuse large B-cell lymphomas (DLBCL).

2. nwCompare software

Proteomics, transcriptomics and metabolomics implies the handling of a huge amount of data. Nano liquid chromatography combined with electrospray mass spectrometry enables the identification of hundreds of proteins in one complex sample whereas transcriptomics analyzes the expression level of about twenty thousand genes. It is very useful to be able to quickly compare lists of proteins, genes or molecules obtained from different patients, different pathological situations.

We designed nwCompare (Pont & Fournié, 2010), a software for n-way comparison of text files. nwCompare performs a line by line comparison of characters, thus, it can be quite useful to compare proteins names, gene names, molecules names, biological pathways names etc.

nwCompare has proven efficacy in proteomics to compare pathological situations (Pont & Fournié, 2010) or large-scale protein analysis (Pottiez & al., 2010).

The first versions of nwCompare were limited to analyse a maximum of 300 files, but, starting from version 3.20, this software is now only limited by the amount of memory of the computer. Moreover, a new feature has been introduced recently, by allowing the computation of a repartition table. It is thus possible to classify each file entry depending of its occurrence. nwCompare is light, very easy to use and enables users to run very complex comparisons just by selecting radio buttons, without learning any comparison syntax (Fig 1). nwCompare is a free software that can be download at:
https://sites.google.com/site/fredsoftwares/home

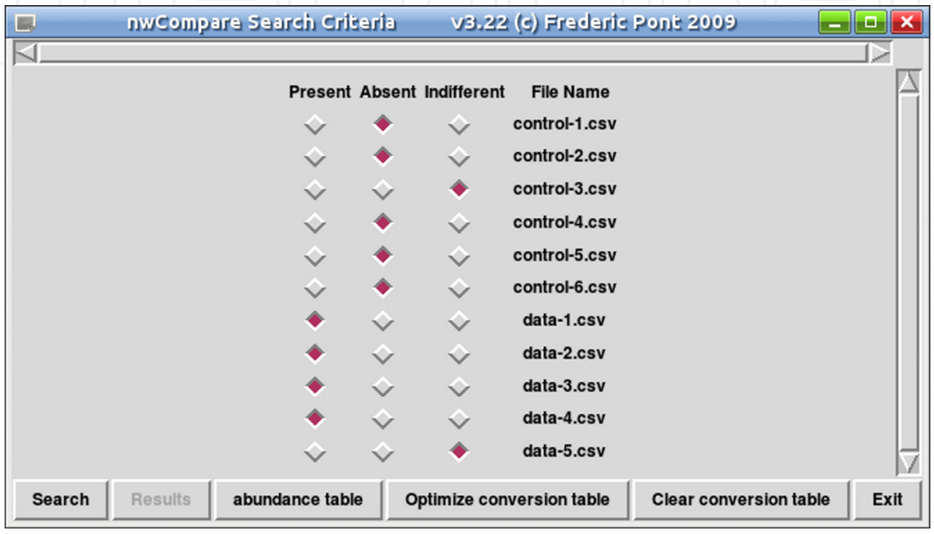


Fig. 1. Screenshot of nwCompare version 3.22 with the simultaneous comparison of eleven protein lists. This example shows the computation of the proteins present in four samples and absent in five controls with two files indifferents. The list of proteins matching those criteria is typically obtained in one second.

3. AutoCompare software

Autocompare freeware was developed as an evolution of nwCompare program to understand the biological significance of large lists of genes or proteins. This software takes as input any data text file and performs string comparisons by line of this file, with a collection of reference files (Fig 2). Then, for each of them, it computes the p-value of the comparison test from hypergeometric distribution tails, then corrects the raw p-values to account for multiple testing, using Bonferroni and Benjamini-Yekutieli methods (Fig 3). We provide AutoCompare with a starting collection of about 5000 genes reference lists based on GSEA (http://www.broadinstitute.org/gsea/) version 3.0 pathways and 162 protein lists based on PANTHER pathways (http://www.pantherdb.org/pathway/) (Mi & al., 2005). Indeed users can also implement in a very straightforward fashion any additional reference list (as .txt format) of their choice.
Autocompare was developed using the Perl programming language (Perl v5.10.1, http://www.perl.org/) and the R statistical programming language under the Linux operating system (ubuntu 10.04, http://www.ubuntu.com/). Autocompare is available for Linux and Windows (https://sites.google.com/site/fredsoftwares/home), and runs on any operating system with Perl, either as a command line tool or with a graphical interface.

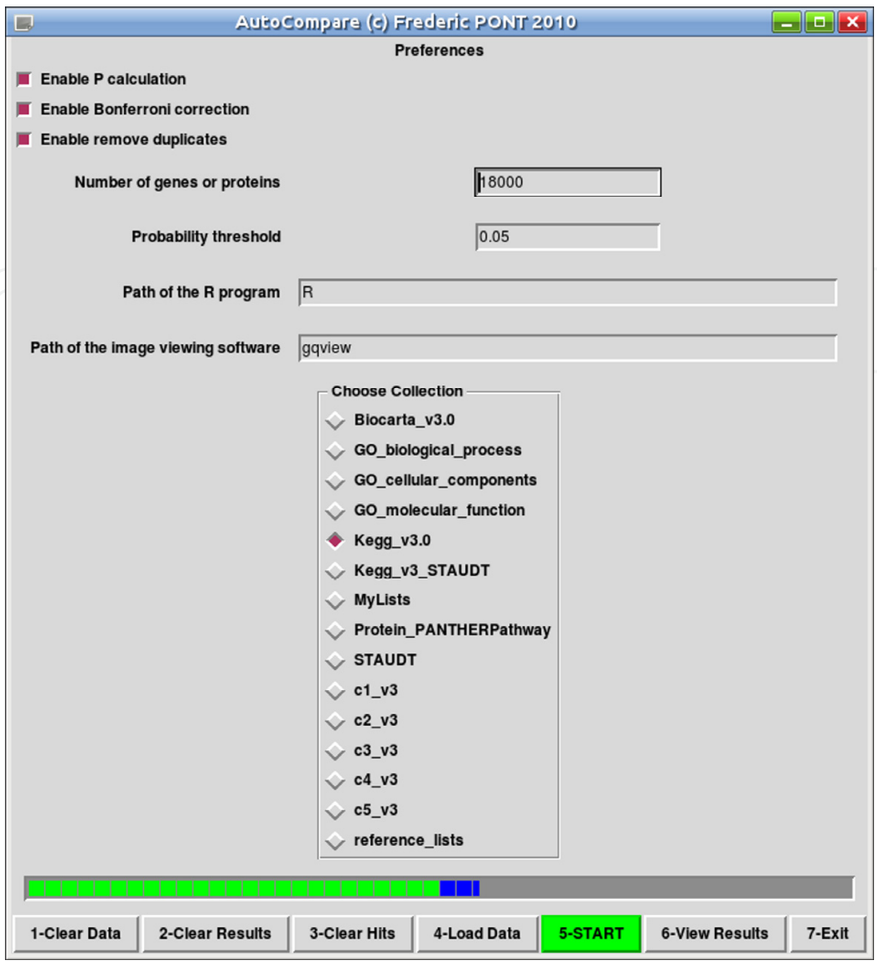
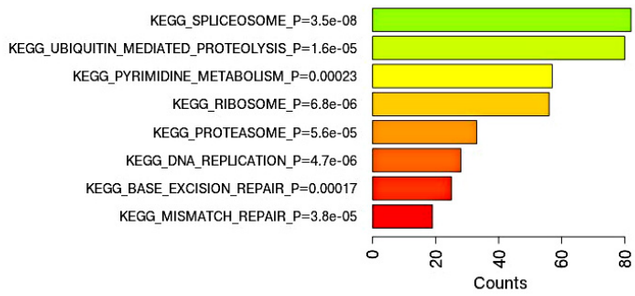


Fig. 2. Screenshot of AutoCompare version 2.31. The software is very easy to use and provides users with biological significance of a large list of genes or proteins.

The main advantages of AutoCompare are that it is very easy to use, rapid and fully automated, it works off line, the number of data files and reference files is only limited by the available disk space, it is very easy to add personalized reference files. In addition, the memory consumption of AutoCompare is very low because only two files are analyzed simultaneously. As nwCompare, AutoCompare performs text file comparisons, so, any kind of data files can potentially be analyzed with it, provided that reference files of the same kind are used.

3.1 AutoCompare false discovery rate (FDR)

To calculate the FDR of AutoCompare, we randomized the genes comprised in the 186 genes lists from the Kegg library on the one hand and the genes comprised in the 3272 gene lists from the Broad institute's GSEA C2' curated library. We then compared the AutoCompare results obtained by querying the same experimental genes list with both the randomized and the correct libraries. With the C2 library, the first false positive was associated with a probability 44 times higher than the Bonferroni threshold. With the Kegg library, the first false positive was associated with a probability 212 times higher than the



Kegg_v3.0	Score	P value	Counts / Total (%)	Bonferroni	Benjamini/Yekutieli
KEGG_SPLICEOSOME	7	3.5e-08	82/128 (64 %)	6.51e-06	3.78e-05
KEGG_DNA_REPLICATION	5	4.7e-06	28/36 (78 %)	8.74e-04	2.45e-03
KEGG_RIBOSOME	5	6.8e-06	56/88 (64 %)	1.26e-03	2.45e-03
KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS	5	1.6e-05	80/138 (58 %)	2.98e-03	4.32e-03
KEGG_MISMATCH_REPAIR	4	3.8e-05	19/23 (83 %)	7.07e-03	8.21e-03
KEGG_PROTEASOME	4	5.6e-05	33/48 (69 %)	1.04e-02	1.01e-02
KEGG_BASE_EXCISION_REPAIR	4	0.00017	25/35 (71 %)	3.16e-02	2.62e-02
KEGG_PYRIMIDINE_METABOLISM	4	0.00023	57/98 (58 %)	4.28e-02	3.10e-02
KEGG_RNA_DEGRADATION	3	0.00037	37/59 (63 %)	6.88e-02	4.44e-02
KEGG_RNA_POLYMERASE	3	0.00043	21/29 (72 %)	8.00e-02	4.64e-02
KEGG_NUCLEOTIDE_EXCISION_REPAIR	3	0.00048	29/44 (66 %)	8.93e-02	4.71e-02
KEGG_CELL_CYCLE	3	0.0019	68/128 (53 %)	3.53e-01	1.71e-01
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	3	0.0022	26/41 (63 %)	4.09e-01	1.83e-01
KEGG_CHRONIC_MYELOID_LEUKEMIA	2	0.004	41/73 (56 %)	7.44e-01	3.09e-01
KEGG_APOPTOSIS	2	0.0043	48/88 (55 %)	8.00e-01	3.10e-01
KEGG_PURINE_METABOLISM	2	0.0057	80/159 (50 %)	1.00e+00	3.85e-01
KEGG_COLORECTAL_CANCER	2	0.013	34/62 (55 %)	1.00e+00	8.26e-01
KEGG_PANCREATIC_CANCER	2	0.021	37/70 (53 %)	1.00e+00	1.00e+00
KEGG_LYSOSOME	2	0.021	60/121 (50 %)	1.00e+00	1.00e+00
KEGG_HOMOLOGOUS_RECOMBINATION	2	0.022	17/28 (61 %)	1.00e+00	1.00e+00
KEGG_ENDOMETRIAL_CANCER	2	0.031	28/52 (54 %)	1.00e+00	1.00e+00
KEGG_N_GLYCAN_BIOSYNTHESIS	1	0.035	25/46 (54 %)	1.00e+00	1.00e+00
KEGG_ENDOCYTOSIS	1	0.047	85/183 (46 %)	1.00e+00	1.00e+00
KEGG_PROTEIN_EXPORT	1	0.055	14/24 (58 %)	1.00e+00	1.00e+00
KEGG_PROSTATE_CANCER	1	0.071	43/89 (48 %)	1.00e+00	1.00e+00
KEGG_P53_SIGNALING_PATHWAY	1	0.077	34/69 (49 %)	1.00e+00	1.00e+00

Fig. 3. Example of results obtained with AutoCompare version 2.31. Top : histogram of the significant biological functions of a gene list data file. Counts indicate the number of genes found in the corresponding reference gene sets. Bottom: Table of the biological functions identified in a data file, sorted by statistical significance.

Bonferroni threshold. We thus implemented the FDR method of Benjamini and Yekutieli (Benjamini, Y. & Yekutieli, 2005, 2001) to adjust the P values in AutoCompare. This method controls the false discovery rate, the expected proportion of false discoveries among the rejected hypotheses. With the C2 library, the first false positive was associated with a probability 2.3 times higher than the probability of the first false positive. With the Kegg library by contrast, the first false positive was associated with a 119 times higher probability than for the first false positive. Hence, AutoCompare hits that are above the Bonferroni threshold are highly significant and without any false positive result. Furthermore, a good estimate of the FDR is thus given by the correction of Benjamini and Yekutieli.

4. Application of nwCompare and AutoCompare to explore the functional significance of gene expression profiles from normal B-cell subsets and of aggressive lymphomas

Normal differentiation of mature B lymphocytes comprises successive stages of maturation, in which naïve B cells reach germinal centers (GC) in lymph nodes and are activated by antigen to form centroblasts. These highly dividing GC centroblasts may further differentiate into centrocytes which, in turn, mature into either quiescent memory B cells or Ig-secreting plasmablasts which leave lymph nodes to home in bone arrow. Hence the

normal B cell maturation in lymph nodes comprises the following sequence: Naïve > GC centroblasts > GC centrocytes > Memory cells, so we searched for the functions associated with the corresponding switches of gene expression signatures.

The transcriptome datasets (Affymetrix CEL files) GSE12195 (Compagno et al., 2009) and GSE15271 (Caron et al., 2009) produced with HG U133-Plus 2.0 platform were downloaded from the NCBI repository GEO database. Together, these comprised 27 normal B cell samples, including 4 naïve B cells, 9 tonsillar germinal center-derived centroblastic cells, 9 tonsillar germinal center-derived centrocytic and 5 memory B cells, 5 lymphoblastoid B-cell lines (B-LCL), 39 follicular lymphomas (FL) and 73 diffuse large B-cell lymphomas (DLBCL) (Caron et al., 2009; Compagno et al., 2009). The raw data from these 144 samples were log (base 2) transformed, normalized in batch by the RMA software and the 54676 probe sets were then reduced to a total of 20606 genes (HUGO symbols) by using the GSEA collapse function set on maximal probe mode (GSEA, http://www.broadinstitute.org/gsea), 18236 of which were fully annotated and thus kept for further study. The genes differentially expressed between two sample groups were defined using two-way Student's tests and $P<0.05$. These gene lists with one gene name per line were converted to text files and then uploaded in Autocompare. More than 4609 genes reference lists based on GSEA (http://www.broadinstitute.org/gsea/) version 3.0 pathways and 162 protein lists based on PANTHER pathways (http://www.pantherdb.org/pathway/) were collected. The differentially-expressed gene subsets were analyzed for enrichment in functionally-related genes among lists downloaded from the gene sets collection. Selective enrichment analysis was then computed with Autocompare using one-sided hypergeometric comparison tests, and False Discovery Rate corrections.

By using this approach, the genes that appeared differentially expressed (*P*-value <0.05) between respectively, naïve-and GC centroblast, GC centroblasts and GC centrocytes and between GC centrocytes and memory B cells were thus analyzed for functional significance by Autocompare using the KEGG library (V 3.0) of functional genesets in *H. sapiens*. In this example, Autocompare performed the corresponding 1970 comparisons within 529 seconds. The GEP of the naïve-to-GC centroblast transition, the so-called "GC GEP signature", comprised 5516 differentially expressed genes. These latter witnessed of a significant increase of cell cycle ($P<10^{-20}$), DNA replication ($P<10^{-11}$), DNA damage and mismatch repair response, STAT3 signaling pathways together with reduced expression of genes for Krebs cycle metabolism and of IRF4-dependent plasmacytic differentiation genes (all with $P<10^{-5}$). Overall, this pattern reflected the unique differentiation program of B cells at the germinal center stage: a strong proliferation and high mutational activities which are both controlled by the Bcl-6 repressor, a program necessary for the clonal expansion of B-cells expressing mutated Ig. The profile of the centroblast-to-centrocyte GC transition comprised fewer differences (1966 differentially expressed genes) which corresponded to up-regulation of genes normally repressed by Bcl-6 ($P<10^{-8}$), hence reflecting the progressive disappearance of this transcriptional repressor. Finally, the GC centrocyte-to-memory B cell transition (5602 differentially expressed genes) showed a significant up-regulation of genes usually repressed by BLIMP-1, together with down-regulation of both cell cycle ($P<10^{-15}$), DNA replication ($P<10^{-11}$), DNA damage and mismatch repair response. This maturation profile, almost reverse to that of the N-to-GC centroblastic transition genes, indicated not only termination of the Bcl-6 dependent GC reaction but also a switch-off of the Blimp-1-dependent plasmacytic differentiation which, together, characterize quiescent memory cells. Hence the main physiological significance emerging from these comparisons is a signature

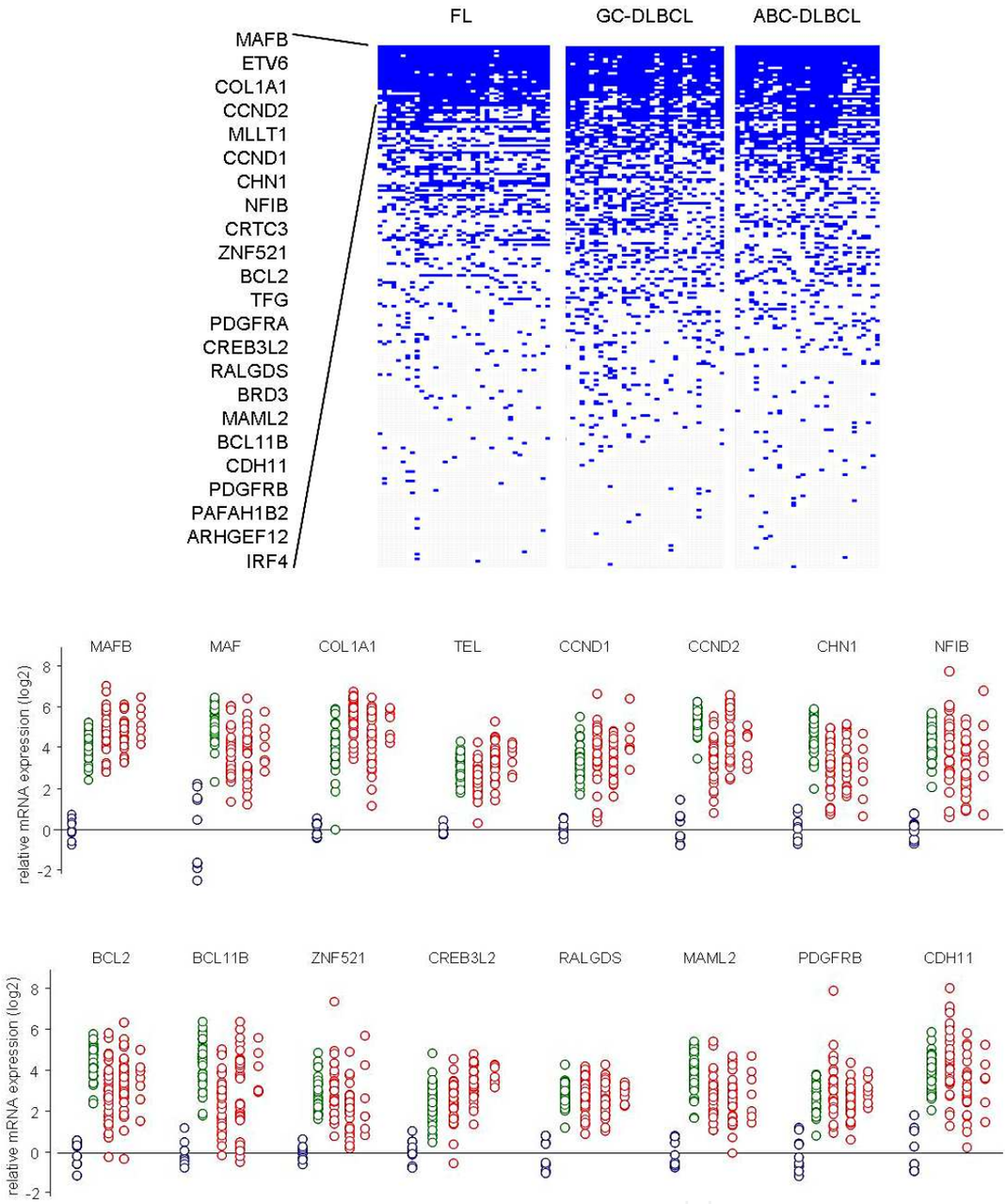
of the germinal center reaction occurring in centroblasts: a unique combination of rapid proliferation and DNA remodeling (somatic hypermutations) without cell death.

4.1 Significance of gene signatures of non-Hodgkin's B cell lymphoma

Most non-Hodgkin's B cell lymphomas emerge from B cells in the germinal center (GC) stage however, by juxtaposing on their normal development the additional programs triggered by their genetic alterations. Accordingly, follicular lymphoma and diffuse large B-cell lymphoma are known to arise in normal GC B cells through genome alterations and mutations targeting genes controlling apoptotic cell death (BCL2, NFKB), differentiation (BCL6, MYC, BLIMP1, IRF4, CREBBP, EP300) or proliferation (BCR, CARD11, MYD88, NFKB, A20, STAT3) (for review, see Lenz & Staudt, 2010). The spectrum of oncogenes over-expressed and tumor suppressor genes under-expressed in each lymphoma translates to a corresponding profile which now defines the clinical subtype and contributes to predict outcome. We asked which of the 457 known human cancer genes (downloaded from the human cancer gene census (http:// www.sanger.ac.uk/ genetics/CGP/Census) were significantly deregulated in terms of either over-expressed oncogenes or down-regulated tumor suppressor genes in each lymphoma sample. Autocompare yielded such a list for each sample, and we then asked which were present in only one, in just two, in several, or in all of the these samples. Using corresponding requests, the 112 individual cancer gene lists were thus compared using nwCompare. This approach revealed that a total of 221 cancer genes were significantly deregulated in FL and DLBCL, among which 23 oncogenes were consistently upregulated in most (>75%) of the samples, like MAFB, ETV6 and COL1A1 which are strongly up-regulated in all (100%) of the samples (Figure 4). On the other hand 49 cancer genes were deregulated in only one or two patients, indicating these cancer genes are probably not driver cancer genes in the B-cell lymphomagenesis.

We then determined the complete set of genes which were differentially expressed by each individual lymphoma relative to the normal GC B cells (P<0.05). On average, 6735 genes were differentially expressed by each follicular lymphoma, 601 of which were shared by all FL. Although these comprised the hallmark over-expression (on average 20-fold) of the anti-apoptotic BCL2 gene, this 601 FL gene set also comprised other deregulated pathways. Using Autocompare, we found that these FL-deregulated pathways were significantly enriched for cytokine-cytokine receptor interactions (37/267 genes, P= 6.7e-12), complement and coagulation cascades (18/69 genes, P=5.1e-11), chemokine signalling (23/190 genes, P= 7.9e-07), ECM receptor interactions (14/84 genes, P=2.7e-06), focal adhesion (22/201 genes, P=7.2e-06), cell adhesion molecules (15/134 genes, P=0.0001), targets of BCL6 (7/19 genes, P= 3.5 e-06), targets of HIF-α (17/164 genes, P=0.0001). In addition, the FL GEP was significantly enriched in the previously depicted FL-type 1 (favourable outcome -associated) and type 2 (poor outcome-associated) immune response genes (respectively 10/40 genes, P=1.0e-06 and 6/23 genes, P=0.0001).

Within GC-type DLBCLs on average, 7365 genes were differentially expressed by each lymphoma, 376 of which were shared by all GC-type DLBCL. With ABC type DLBCL on average 7184 genes were differentially expressed by each ABC type DLBCL, 618 of which were shared by all ABC-type DLBCL.This suggested that DLBCL are more heterogeneous than FL, and that ABC-type DLBCL harbour the most genetically diversified profiles. The functional significance of both GC-type and ABC-type DLBCL gene sets comprised the same pathways as for FL plus the lysosome pathway (12/121 genes, P=5.1e-5).



Top: Patterns of cancer genes differentially expressed (oncogenes over-expressed and tumor suppressor genes under-expressed) in follicular lymphomas and diffuse large B cell lymphomas shows that Follicular lymphomas are more homogeneous than DLBCL. Each column corresponds to a patient sample and genes are lines. A blue dot means that the expression of the gene was deregulated for the corresponding patient, a white dot means that the expression of the gene was similar to normal individuals. MAFB, for example, is represented by an horizontal blue line, which mean that this gene was deregulated in all patients. Bottom: Most significantly up-regulated oncogenes in aggressive lymphomas. mRNA expression was normalized to the mean of normal samples (blue), compared to patient's samples: follicular lymphomas (green), DLBCL (red circles), successively grouped as GC, ABC and unclassified DLBCL subtypes, respectively).

Fig. 4. Pattern of oncogenes overexpressed in aggressive lymphomas.

5. Using AutoCompare with proteomics datasets

Proteomic scientists have two options to take benefit of AutoCompare with their proteomic datasets. The first option is straightforward: it is to use directly the starting collection of PANTHER protein pathways provided with AutoCompare. PANTHER protein pathways are built with Uniprot (http://www.uniprot.org/) protein accession numbers. If another protein database is to be used, the Protein Identifier Cross-Reference Service (PICR, http://www.ebi.ac.uk/Tools/picr/) can be applied to convert the data. Further, we present below two examples illustrating how AutoCompare can help data mining proteomes

Example 1: A virtual follicular lymphoma's proteome was created by converting genes up-regulated in follicular lymphoma (as depicted in §4) into protein accession numbers with PANTHER protein pathways. By using AutoCompare, this virtual proteome was then conveniently compared to a series of other proteomes, namely the whole PANTHER pathways proteome collection. Table 1 shows that the top ranking matches concerned proteins of apoptotic cell death, differentiation or proliferation (apoptosis, p53, p38 MAPK and Wnt pathways), focal adhesion (integrin and cadherin pathways), coagulation pathways, cytokines and chemokines signaling and immune response were differentially expressed in FL. In addition, angiogenesis (118/1231 proteins) and various growth factor signaling pathways (PDGF 65/938 proteins; VEGF 39/416 proteins; EGF 61/1071 proteins; IGF 32/276 proteins; FGF 53/978 proteins) were also enriched. Indeed in this example, these proteome comparisons matched with the results from transcriptome comparisons depicted in §4. Of note, the reverse strategy: converting protein accession numbers into gene names is also possible via the Protein Information Resource (PIR) (http://pir.georgetown.edu/pirwww/search/idmapping.shtml). Then, AutoCompare can be used with gene names, as described in § 4, taking advantage of the much larger collection of gene pathways provided with AutoCompare. The disadvantage of these conversion strategies however is that the original amount of data generally increases because of redundancy in databases and gene synonyms. Moreover, since most conversion tools do not filter results by taxonomy, this increase of non relevant data also augments the P values.

Example 2: Comparative analysis of experimental proteomes. The lymphoma cell line Karpas 299 was cultured in vitro for 48 hours in complete medium with and without the bisphosphonate drug zoledronate, the cells were isolated, their protein extract were prepared and the two resulting proteomes were analysed by mass spectrometry: briefly, the proteins were digested by trypsin, the peptides were analysed by nano-electrospray mass spectrometry and identified in SwissProt database using MASCOT (http://www.matrixscience.com/) software (unpublished results). AutoCompare allowed us to compare them to each other and to the proteomes listed the PANTHER pathways. This approach identified 52 matches between lymphoma proteins and one of the reference pathway proteomes. In control lymphoma cells for instance, Autocompare identified among others, 10 proteins of "cytoskeletal_regulation_by_Rho_GTPase" (O15144, O15145, O15511, P23528, P62736, P63261, P63267, P68032, P68133, Q5NBV3), 10 proteins involved in "inflammation mediated by chemokines and cytokines" and 6 proteins from the "Integrin_signalling_pathway". Of note, this approach also indicated that the 5 proteins P62736, P68032, P68133, Q13363 and Q969G3 expressed by the lymphoma cells in control conditions are involved in the Wnt pathway. By contrast, the proteome from cells treated with zoledronate only comprised the P68133 and Q13363 proteins from this pathway, suggesting the treatment had inhibited expression of the 3 others. Hence this example shows

how Autocompare can be used to pinpoint targeting of the morphogen Wnt cascade by the bisphosphonate drug.

Protein PANTHER Pathway	Counts/Total (%)
Integrin signalling pathway	132/1175 (11%)
Inflammation mediated by chemokine and cytokine	135/1417 (10%)
Angiogenesis	118/1231 (10%)
Wnt signaling pathway	142/2085 (7%)
Interleukin signaling pathway	74/596 (12%)
Blood coagulation	51/244 (21%)
Cadherin signaling pathway	78/885 (9%)
Apoptosis signaling pathway	74/839 (9%)
PDGF signaling pathway	65/938 (7%)
p53 pathway	49/548 (9%)
Toll receptor signaling pathway	37/281 (13%)
Oxidative stress response	39/349 (11%)
B-cell activation	39/361 (11%)
TGF-beta signaling pathway	56/810 (7%)
Heterotrimeric G-protein signaling pathway	59/978 (6%)
VEGF signaling pathway	39/416 (9%)
EGF receptor signaling pathway	61/1071 (6%)
Heterotrimeric G-protein signaling pathway	48/670 (7%)
Insulin IGF pathway-mitogen activated protein kinase	32/276 (12%)
Interferon-gamma signaling pathway	29/256 (11%)
Metabotropic glutamate receptor group II pathway	36/423 (9%)
GABA-B receptor II signaling	22/126 (17%)
T-cell activation	43/631 (7%)
Cytoskeletal regulation by Rho GTPase	44/674 (7%)
FGF signaling pathway	53/978 (5%)
Enkephalin release	29/280 (10%)
Endogenous cannabinoid signaling	18/83 (22%)
Endothelin signaling pathway	45/748 (6%)
Nicotinic acetylcholine receptor signaling pathway	51/1011 (5%)
p38 MAPK pathway	17/96 (18%)

Table 1. Top rated PANTHER pathways identified by AutoCompare after conversion of follicular lymphoma genes into protein accession numbers. Counts indicate the number of genes found in the corresponding reference gene sets.

6. Conclusion

In conclusion, this example study shows how the use of Autocompare and nwCompare enables users to get fast access to multidimensional comparisons and to the corresponding analysis of large datasets such as proteomes and transcriptomes. We illustrated here this use by the determination of oncogenes and functions involved in the biology of aggressive human B cell lymphomas. Proteomics data sets (protein names, protein accession numbers) can be compared directly in nwCompare since this software performs strings comparisons.

AutoCompare is provided with a starting collection of PANTHER protein pathways for a direct analysis of proteomic datasets. Proteomics users can additionally take advantage of AutoCompare large gene starting database of about 5500 pathways by converting protein names into gene names.

7. Acknowledgements

Work in JJF's lab is supported by institutional grants from INSERM, Université de Toulouse 3 and CNRS, as well as by grants from Institut National du Cancer (contracts RITUXOP and V9V2TER). We thank L. Pasqualucci (Columbia University, NY) for kindly providing us with clinical classifications of the lymphoma samples from GSE12195 dataset.

8. References

Benjamini, Y., & Yekutieli, D. (2005). Quantitative trait loci analysis using the false discovery rate. *Genetics,* 171, pp 783-790, Print ISSN: 0016-6731; Online ISSN: 1943-2631

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.,* 29, 4, pp 1165-1188, ISSN 0090-5364

Caron, G., Le Gallou, S., Lamy, T., Tarte, K. & Fest, T. (2009). CXCR4 expression functionally discriminates centroblasts versus centrocytes within human germinal center B cells. *J Immunol.* 182, pp 7595-7602.

Compagno, M., Lim, W. K., Grunn, A., Nandula, S. V., Brahmachary, M., Shen, Q., Bertoni, F., Ponzoni, M., Scandurra, M., Califano, A., Bhagat, G., Chadburn, A., Dalla-Favera, R. & Pasqualucci, L. (2009). Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature*, 459, pp 717-721.

Côté, R.G., Jones, P., Martens, L., Kerrien, S., Reisinger, F., Lin, Q., Leinonen, R., Apweiler, R. & Hermjakob, H. (2007). The Protein Identifier Cross-Referencing (PICR) service: reconciling protein identifiers across multiple source databases. *BMC Bioinformatics,* 8, pp 401-414.

Lenz, G. and Staudt, L. (2010). Aggressive lymphomas. *The New England Journal of Medicine,*. 362, pp 1419-1429.

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M. J., Kitano, H. & Thomas* P. D. (2005). The PANTHER database of protein families, subfamilies, functions and pathways. *Nucl. Acids Res*, 33, suppl 1, D284-D288.

Pont, F. & Fournié, JJ. (2010). Sorting protein lists with nwCompare: a simple and fast algorithm for n-way comparison of proteomic data files. *Proteomics,* 10, 5, March 2010, pp 1091-1094. ISSN: 1615-9861.

Pottiez, G., Deracinois, B., Duban-Deweer, S., Cecchelli, R., Fenart, L., Karamanos, Y. & Flahaut, C. (2010). A large-scale electrophoresis- and chromatography-based determination of gene expression profiles in bovine brain capillary endothelial cells after the re-induction of blood-brain barrier properties. *Proteome Sci*., 15, 8, November 2010, pp 57. ISSN: 1477-5956.

IntechOpen