Transcriptional characterization of a prospective series of primary plasma cell leukemia revealed signatures associated with tumor progression and poorer outcome

Katia Todoerti,1,2,3* Luca Agnelli,1,2,* Sonia Fabris,1,2 Marta Lionetti,1,2 Giacomo Tuana,1,2 Laura Mosca,1,2 Luigia Lombardi,1,2 Vitina Grieco,3 Gabriella Bianchino,3 Fiorella D'Auria,3 Teodora Statuto,3 Carmela Mazzoccoli,4 Luciana De Luca,4 Maria Teresa Petrucci,5 Fortunato Morabito,6 Massimo Offidani,3 Francesco Di Raimondo,8 Antonietta Falcone,9 Paola Omede,10 Pierfrancesco Tassone,11 Mario Boccadoro,10 Antonio Palumbo,10 Antonino Neri,1,2 and Pellegrino Musto3

1Department of Clinical Sciences and Community Health, University of Milan, Milano, Italy; 2Hematology 1 CTMO, Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico, Milano, Italy; 3Department of Onco-Hematology, IRCCS-CROB, Referral Cancer Center of Basilicata, Rionero in Vulture, Italy; 4Laboratory of Pre-Clinical and Translational Research, IRCCS-CROB, Referral Cancer Center of Basilicata, Rionero in Vulture, Italy; 5Hematology, Department of Cellular Biotechnologies and Hematology, “La Sapienza” University, Roma, Italy; 6Hematology Unit, Azienda Ospedaliera di Cosenza, Cosenza, Italy; 7Hematologic Clinic, Azienda Ospedaliero-Universitaria Ospedali Riuniti di Ancona, Ancona, Italy; 8Department of Biomedical Sciences, Division of Hematology, Ospedale “Ferrarotto”, University of Catania, Catania, Italy; 9Hematology Unit, IRCCS “Casa Sollievo della Sofferenza” Hospital, San Giovanni Rotondo, Italy; 10Division of Hematology, University of Torino, A.O.U. S.Giovanni Battista, Torino, Italy; 11Medical Oncology Unit, Department of Experimental and Clinical Medicine, Magna Graecia University and “T. Campanella” Cancer Center, Catanaro, Italy

* These authors contributed equally to this work

Running title: Transcriptional analysis of plasma cell leukemia

Keywords: plasma cell leukemia; myeloma; gene expression

Corresponding author: Antonino Neri, Department of Clinical Sciences and Community Health, University of Milan and Hematology 1, Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico, via F. Sforza, 35 - 20122 Milan Italy; e-mail: antonino.neri@unimi.it

Disclosure of potential conflicts-of-interest: The Authors declare no competing financial interests.

Text word count: 4501
Abstract word count: 244
Number of figures: 4
Number of tables: 1
Number of references: 37
Translational relevance

Primary plasma cell leukemia (pPCL) is a very aggressive and rare hematological malignancy. So far, genomic and clinical differences between pPCL and multiple myeloma (MM) have been demonstrated, mainly based on retrospective studies. Here, we took advantage of a prospective series of pPCLs included in a phase II clinical trial to investigate pPCL global mRNA expression profiles. A 503-gene signature that distinguished pPCL from MM was identified; a fraction of these genes have been suggested whose expression is putatively associated with the aggressiveness of clinical presentation. In addition, a 27-gene model has been identified with potential clinical relevance that distinguished, within pPCL, those with worst outcome. Since pPCL represents a high-risk clinical entity per se, our findings are helpful to provide insights into molecular features that could be able to prognostically stratify patients with this aggressive form of plasma cell dyscrasia.

Abstract

Purpose: Plasma cell leukemia (PCL) is a rare form of plasma cells dyscrasia that presents either as a progression of previously diagnosed multiple myeloma (MM), namely secondary PCL, or as initial manifestation of disease, namely primary PCL (pPCL). Although presenting signs and symptoms include those seen in MM, pPCL is characterized by several aspects that define more aggressive course. Herein, we have investigated the transcriptome of pPCLs and correlated differential expression profiles with outcome, to provide insights into the biology of the disease.

Experimental design: The expression profiles of 21 newly-diagnosed pPCLs included in a multicenter prospective clinical trial were generated using high-density microarray, then evaluated in comparison with a representative series of multiple myeloma (MM) patients and in association with clinical outcome.
Results: All but one of the pPCLs had one of the main IGH translocations, whose associated transcriptional signatures resembled those observed in MM. A 503-gene signature distinguished pPCL from MM, from which emerged 26 genes whose expression trend was associated with progressive stages of plasma cells dyscrasia in a large dataset from multiple institutions, including samples from normal donors throughout PCL. Finally, three genes were identified having expression levels correlated with response to the first-line treatment with lenalidomide/dexamethasone, whereas a 27-gene signature was associated with overall survival independently of molecular alterations, hematological parameters and renal function.

Conclusions: Overall, our data contribute to a fine dissection of pPCL and may provide novel insights into the molecular definition of patients with poorer prognosis.
**Introduction**

Plasma cell leukemia (PCL) is a rare but highly aggressive disease that represents 3–5% of all plasma cell (PC) disorders. Its diagnosis is based on the Kyle’s criteria, which requires circulating PCs to account for at least 20% of peripheral blood leukocytes and/or an absolute circulating PC count of $2 \times 10^9$/l, with evidence of monoclonal gammopathy. Clinically, the PCL can be distinguished into primary PCL (pPCL), originating de novo without any prior history of multiple myeloma (MM), or secondary PCL (sPCL), arising from a pre-existing MM tumor that eventually progressed to the leukemic phase. The prognosis of pPCL is very poor and is even worst for sPCL, with median survival durations of 7 to 11 months and 2 to 7 months, respectively. The pathogenetic mechanisms underlying either the de novo presentation or the aggressive transformation of relapse/refractory MM are largely unknown (1).

So far, few studies on both pPCL and sPCL have been performed with the attempt of characterizing the molecular alterations underlying PCL biology, mainly based on immunophenotypic, fluorescence in situ hybridization (FISH), and cytogenetic or array-based Comparative Genomic Hybridization (aCGH) analyses (2-6). All of them, albeit suffering restrictions of numerically limited cohorts of cases, agreed in identifying significant molecular differences among the MM and PCL presentation, specifically in regards to the surface immunophenotype (the differential expression of CD20, CD56, CD9, CD117 and HLA-DR antigens) and the increased number of genomic aberrations identified in PCL (7). Indeed PCL shows higher incidence of translocations involving IGH@ at 14q32, in particular t(11;14), whereas very few patients with hyperdiploidy have been reported differently from MM (1). Among PCL, only t(4;14) was found associated with shorter overall survival (3), while controversial data exist on the favorable prognostic significance of t(11;14) (1). MYC rearrangement, a late genetic event correlated with disease progression in MM, has been frequently found in PCL also in association with a worse trend of the disease (5, 6).
To date, high resolution transcriptional profiling studies of PCL are limited to a retrospective small series reported by us (8), and the recently published investigation from Usmani et al (9) who described pPCL cases included in the Total Therapy protocols database. Both studies indicated that transcriptional features may distinguish the two entities, moreover in the latter study the Authors additionally indicated that pPCL is an independent adverse prognostic factor in comparison with MM.

Herein, we investigated a series of 21 previously untreated pPCL patients included in a prospective clinical trial using whole-transcript profiling arrays, and correlated the transcriptional profiles with the primary and secondary outcome endpoints to evaluate whether different molecular or clinical entities may exist within pPCL. In addition, we compared the transcriptional profiles of the 21 pPCL with those of a representative dataset of 55 newly-diagnosed primary MM tumors; the differentially expressed genes were then evaluated in a large dataset, profiled on a different generation array platform, from multiple institutions including normal donors, different forms of plasma cell dyscrasia from asymptomatic to aggressive stages, and a collection of human myeloma cell lines, with the aim to identify transcripts that might contribute to the pathogenesis of pPCL as well as to tumor progression in plasma cell dyscrasia.

Materials and Methods

Patients
Pathological specimens were collected from 23 untreated pPCL patients included in a multicenter Italian clinical trial (RV-PCL-PI-350, EudraCT N° 2008-003246-28), an open label, exploratory, single arm, two-stage study aimed at evaluating safety and anti-tumor activity of combined lenalidomide and dexamethasone (LD) in first-line treatment of previously untreated pPCLs. All patients gave their informed consent for molecular analyses. The primary endpoint of the study was the response rate after 4-cycle therapy over a 4-month
schedule, according to International Uniform Criteria (10, 11); secondary endpoints were (i) time to progression, (ii) progression free and overall survival (OS), (iii) eligibility to undergo autologous or allogeneic stem cells transplantation after LD treatment, and (iv) serious/severe adverse event rate. Until February 2012, the median follow-up was 23 months (range 9-32). To the purpose of the present work, we considered response rate and OS. Highly purified (≥90%) bone marrow PCs were obtained as previously reported (8). Demographic information and details on molecular alterations of the whole cohort were reported in a previous manuscript aimed at the genomic characterization of the samples (12). Fifty-five newly diagnosed MM patients were included in the study; these cases have been previously reported (13) and were selected based on the representativeness of the main molecular characteristics.

**Generation of gene expression data**

Total RNA was available from 21 pPCL and 55 MM patients. MM samples were part of a larger cohorts previously profiled by us on old-generation array (GeneChip® HG-U133A) (13), herein resampled on newer generation arrays and selected based on their representativeness of molecular features, to allow the comparison with pPCL profiles. Preparation of DNA single-stranded sense target, hybridization to GeneChip® Gene 1.0 ST arrays (Affymetrix Inc., Santa Clara, CA) and scanning of the arrays (7G Scanner, Affymetrix Inc.) were performed according to manufacturer’s protocols. Log2-transformed expression values were extracted from CEL files and normalized using NetAffx Transcript Cluster Annotations, Release 32 (June 2011) and Robust Multi-array Average (RMA) procedure in Expression Console software (Affymetrix Inc.). Non-annotated transcript clusters were discarded. The expression values of transcript cluster ID specific for loci representing naturally occurring read-through transcriptions or mapped to more than one chromosomal location were summarized as median value for each sample. The data have been deposited at
NCBI Gene Expression Omnibus repository (http://www.ncbi.nlm.nih.gov/geo) and are accessible through GEO Series accession number GSE39925. The MM samples were stratified into the five TC (Translocation/Cyclin) groups as previously described (14).

The dataset of samples profiled on GeneChip® HG-U133A arrays (Affymetrix Inc.) was generated by two proprietary (GSE13591 and GSE6205) and two publicly available datasets (GSE6477 and GSE6691)(15, 16), which globally included 24 healthy donors (N), 33 monoclonal gammopathy of undetermined significance (MGUS), 24 smoldering MM (SMM), 200 newly-diagnosed and 26 relapsed MM, 9 PCL patients (3 of whom primary), together with 23 HMCLs fully characterized in a previous study (17). Expression values were extracted from CEL intensity files using GeneAnnot custom chip definition files version 2.2.0, and normalized by Robust Multi-array Average procedure, as previously described (18). To prevent batch effects, namely the inclusion of low-quality or non-reproducible data, Normalized Unscaled Standard Error (NUSE) and Relative Log-Expression (RLE) distributions were generated in aroma.affymetrix package for all samples; samples were removed if the 25th or the 75th percentile of NUSE and RLE exceeded the value of ± 1.05 or ± 0.5, respectively, which led to a definitive amount of 323 comparable samples.

Microarray data analysis

Hierarchical agglomerative clustering was performed on the list of genes whose average change in expression levels varied at least 2-fold from the mean across the considered data set, as previously reported (8). Euclidean and Ward were respectively used as distance and linkage methods in hclust function in R software.

The differentially expressed genes discriminating the considered classes were identified using the Significant Analysis of Microarrays software version 4.00 as previously described (19). The functional annotation analysis on the selected lists was performed using Gene Set Enrichment Analysis (20) to identify significantly enriched curated gene sets (version 3.0).
among those included in MSigDB database (http://www.broadinstitute.org/gsea/msigdb/index.jsp). False Discovery Rate lower than 25% was used as enrichment statistic and the rank was permuted with 1000 permutations. Only gene sets including 5 through 1000 genes were considered in the analysis.

Jonckheere-Terpstra test was used with clinfun package in R software to investigate the significance of the trend from normal donors through PCL cases in the expression levels of selected gene list generated on HG-U133A arrays. To avoid biases conceivably due to variability in the target detection technology, the genes whose trend in MM and PCL cases on HG-U133A arrays was incoherent with the data generated on HuGene1.0st arrays were not considered. For robustness, to reduce biases that might be due to numerical unbalances within the groups (being MM group largely over-represented) and to gain the most significant trends, an additional criterion was imposed that only steadily ascending or descending median values were allowed; in addition, a differential expression should be observed in at least one condition (Kruskal-Wallis tests performed using the stat package in R software). Benjamini and Hochberg correction was used to adjust significance of multiple tests.

**Survival analysis**

Cox proportional hazards model and 100,000 permutations were used in the globaltest function of R software to search for a correlation between the gene expression levels and overall (OS) (21). Patients were stratified into two groups using the prognostic gene signatures and hierarchical clustering with Euclidean and Ward as distance and linkage metrics, respectively. The groups identified by this approach were tested for association with survival using the Kaplan-Meier estimator and log-rank test, and P-values were calculated according to the standard normal asymptotic distribution. Survival analysis was conducted with the survcomp package in R software. Independence between commonly used
prognostic factors and gene signatures was tested using multivariate Cox regression procedure of survival R package.

**Real-Time Quantitative Polymerase Chain Reaction**

The expression levels of \textit{NFKBIA}, \textit{RELB}, \textit{RHOA}, and \textit{SDC1} were analyzed in purified CD138+ cells by means of Real-Time Quantitative PCR (Q-RT-PCR). Total RNA was converted to cDNA using M-MLV reverse transcriptase (Invitrogen). Inventoried or Made-to-Order TaqMan Gene Expression Assays (Hs00153283\_m1 for \textit{NFKBIA}, Hs00232399\_m1 for \textit{RELB}, Hs01051295\_m1 for \textit{RHOA}, and Hs00896423\_m1 for \textit{SDC1}) and the TaqMan Fast Universal Master Mix were used according to manufacturer’s instructions (Applied Biosystems, Foster City, CA). GAPDH-specific pre-developed assay reagent (PDAR) (Applied Biosystems) was used as internal control. The measurement of gene expression was performed using ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). All the samples were run in duplicate. Data were expressed as $2^{-\Delta\text{Ct}}$ (Applied User Bulletin No. 2).

**Results**

**Association between primary IGH translocation and the transcriptional profiles of pPCL tumors**

Herein, we analyzed the transcriptional profile of 21 of the 23 pPCLs (for whom RNA material was available) included in the prospective Italian multicentric clinical trial RV-PCL-PI-350, all of which characterized by FISH for the main genomic aberrations. Specifically, three of them had t(4;14), nine t(11;14), seven t(14;16) and one t(14;20): of note, only one patient lacked any of these major IGH translocations (12). In addition, we generated the gene expression profiles of 55 primary MM tumors, representative of the type and frequency of the major molecular aberrations of this clinical entity. Unsupervised analysis was performed on the 1145 most variable transcripts in the 21 pPCLs dataset to evaluate whether natural
grouping of the transcriptional profiles might be associated with distinct molecular subgroups. Indeed, as shown in Figure 1A, the hierarchical clustering was mostly driven by the presence of the main recurrent IGH chromosomal translocations. Specifically, all cases with translocated \textit{MAF} genes clustered in one main branch ($P = 4.423 \times 10^{-5}$), whereas the t(11;14) and t(4;14) positive pPCLs were clustered together in the second branch. Within the \textit{MAF}-translocated PCL, a sample was clustered lacking major translocations (PCL-020) with expression profile resembling those of \textit{MAF/MAFB} samples, in all likelihood due to the exclusive and spiked expression of \textit{MAFA}, another member of \textit{MAF} family transcription factor (Supplementary Figure 1). Unfortunately, lack of available material did not allow to investigate the occurrence of a translocative event.

We performed a multi-class supervised analysis to identify genes that might specifically differentiate pPCL cases stratified on the main IGH chromosomal translocations. Forty-seven genes were differentially expressed in the three classes (Figure 1B; Supplementary Table 1). As expected, the direct targets of t(11;14) (\textit{CCND1}), t(4;14) (\textit{WHSC1}), and \textit{MAF} genes were found respectively upregulated in each corresponding pPCL subgroup (to the aim of the analysis, PCL-020 was considered part of the \textit{MAF} group). \textit{FGFR3} gene was not identified in the gene signature of the t(4;14), likely due to missing derivative chromosome 14 in one of the three cases. Of note, the \textit{CCND2} transcript showed the highest score and positively marked both the \textit{MAF} cases and, although at a lesser extent, the t(4;14)-translocated pPCLs. As depicted in Figure 1B, the signatures of the IGH translocation largely resembled those observed in MM and in part previously described.

\textbf{Identification of a 503-gene transcriptional signature distinguishing pPCL from MM}

Based on their clinical presentation, it is conceivable that pPCL is characterized by distinct transcriptional patterns from intramedullary MM. We therefore investigated the relationship between the transcriptional profiles of pPCL and MM tumors. First, we performed
hierarchical clustering of the 1166 most variable genes in the dataset including the 21 pPCL and 55 MM cases (Figure 2). Again, the clustering was mainly driven by the IGH translocations: in fact, either t(14;16)/t(14;20) or t(4;14) or t(11;14) cases clustered together independently of belonging to MM or pPCL groups, suggesting that such translocation events had stronger consequences on the transcriptional fingerprint than those due to leukemic phenotype.

Next, in order to identify the genes that specifically distinguished the two clinical entities, we performed a supervised analysis between the 21 pPCL and the 55 MM cases. We found 503 differentially expressed genes, 366 of which upregulated in pPCLs (Supplementary Table 2). Functional annotation analysis was performed using GSEA (Supplementary Table 3). Among the identified gene sets, it was worth mentioning those associated with NF-kB pathway, structural organization of the cell and migration, together with some gene sets that suggested putative alterations of CD40, TGFB, AKT and FAS signaling pathways. Of note, the SDC1 gene, codifying the CD138 surface antigen, showed significantly lower expression level in pPCLs. Prompted by these findings, we evaluated based on gene expression data whether pPCL cases were likely to show more activated NF-kB pathway or higher proliferation pattern. To this aim, we took advantage of the NF-kB indexes based on expression data as described by Annunziata et al. (22) and Keats et al. (23); and the proliferation indexes as described by Hose et al. (24), Bergsagel et al. (25), and Zhan et al. (26). Interestingly, both the NF-kB indexes calculation procedures highlighted increased NF-kB pathway activity in pPCL (Supplementary Figure 2A and B); conversely, no significant difference raised according to the gene proliferation indexes (Supplementary Figure 2C-E). As a methodological validation of our results, four of the genes identified and included in the mentioned gene sets (NFKBIA, RELB, RHOA and SDC1) were validated by Q-RT-PCR in all samples for which RNA was available (15 pPCL and 40 MM patients) (Supplementary Figure 3). The correlation coefficients of the expression of each gene as determined by microarray or
Q-RT-PCR indicated an optimal concordance for all genes (median Pearson’s correlation coefficient 0.73).

**Modulation of genes included in MM-pPCL signature throughout the different stages of plasma cell dyscrasia**

We hypothesized that the identified 503-gene signature distinguishing pPCL and MM cases could be accountable to the occurrence of overt/aggressive disease. Therefore, we tried to elucidate whether the genes of this list, or at least a part of them, may be modulated through (if not discriminate) the different stages of the disease; namely, whether the expression trend of these transcripts could underlie the transition from indolent forms to aggressive stages. To verify this, we generated a comprehensive dataset including independent samples from three different cohorts, encompassing PC samples from normal subjects, MGUS, smoldering MM, newly-diagnosed MM, and PCL patients, all profiled on the HGU133A microarray chip. Of the public available samples, 274 passed the quality controls applied to prevent biases due to inter-cohort analysis. In this dataset, we investigated the expression profiles of the 360/503 differentially expressed transcripts represented on the HG-U133A array, and tested whether a trend existed in their expression levels through all the progressively malignant stages of plasma cell dyscrasia. We selected the 26 most significant genes (Supplementary Figure 4A) whose trend in expression levels correlated with progressive disease. Larger expression level spreads were observed in MM subgroups than other stages, in line with the well-accepted notion that MM represents a widely heterogeneous entity. However, no significant associations were found between the expression levels of the 26 genes and the molecular stratification of the myeloma samples based on the occurrence of IGH translocations, with few marginally significant exceptions (Supplementary Figure 4B). In addition, we evaluated the expression profiles of those 360 genes in 26 relapsed MM and 23 HMCLs. Of note, although not included in the analysis, the trend of the expression values of all the genes in
HMCLs was concordant with that observed in primary tumors, thus reinforcing the suggestion that the deregulation of these genes (either with increasing or decreasing pattern) is compatible with the “strength” of neoplastic transformation. Among the identified transcripts, is worth mentioning the proteasome-associated gene PSMD6 and the polycomb group protein codifying gene EZH2, whose expression has been associated with tumor burden in MM (27).

Identification of transcriptional profiles with clinical relevance in the definition of pPCL with poorer outcome

Finally, the transcriptional features of pPCL were evaluated in the context of the outcome data of the prospective series of pPCL patients. In particular, we evaluated whether the occurrence of specific gene signature might be associated either with response rate or OS. To this aim, we first investigated the dataset seeking for differentially expressed genes in patients who failed to respond to frontline 4-cycle therapy of LD. The analysis led to the identification of three genes (YIPF6, EDEM3, CYB5D2) able to distinguish non-responder patients from responders [i.e., complete response (CR), very good partial response (VGPR) and partial response (PR)] (Figure 3). When PR, VGPR and CR were considered separately, no specific differentially expressed genes could be evidenced. The identified transcriptional pattern will be helpful to integrate results on efficacy and side effects of the first-line treatment of LD in pPCL.

Next, we assessed the relationship between each of the 1145 most variable genes across the pPCL dataset and OS, using a statistical approach based on linear regression model in which the distribution of the response variable is modelled as a function of the expression levels of each gene. Of the 1145 genes, 27 reached a highly significant correlation ($P<.01$) with OS (Table 1). Based on this 27-gene model, the 18 pPCL cases for whom follow-up information was available could be divided into two groups, of 6 and 12 patients respectively, who showed different outcome (Figure 4A-B). This model retained independency from all the molecular characteristics available (Table 2), as well as from age, sex, LDH levels, renal
function and hematologic parameters (data not shown). However, it is worth mentioning that none of the cytogenetic aberration was associated \textit{per se} with OS (12). It is also worth underlying that the 27-gene model was not independent of patients being subjected to autologous stem cell transplantation (ASCT), indicating that this therapeutic approach points definitively towards a more favorable outcome.

We finally tested the independency of the 27-gene model from other gene-risk models based on gene expression data described in multiple myeloma (28-30). Specifically, to this aim, we first stratified the samples included in our dataset according to the criteria defined by the UAMS 70-gene and 17-gene models from Shaughnessy \textit{et al.} (30), the IFM 15-gene model by Decaux \textit{et al.} (28), and the UK 6-gene model by Dickens \textit{et al.} (29). Of them, only UAMS models effectively stratified pPCL patients into two significantly different risk groups (data not shown). Importantly, the 27-gene model retained significant correlation with outcome against both UAMS and other models (Table 2).

**Discussion**

In the present study, we reported a comprehensive molecular and transcriptional analysis of a series of pPCLs patients included in a prospective clinical trial aimed at exploring efficacy and safety of lenalidomide and dexamethasone combination as first line therapy in previously untreated patients. The analysis was particularly aimed at identifying those transcriptional features that may contribute to explain the aggressive phenotype of PCL, as well as at correlating their gene expression profiles with the clinical outcome.

According to previous reports (2, 4-6), a high incidence of the 14q32 translocations was found in our series of pPCLs. The transcriptional analysis further supported that the main IgH chromosomal translocations drive pPCL patients clustering and are associated with specific signatures, as it occurs in MM. The combined analysis of pPCL and MM patients confirmed that pPCLs and MMs clustered together in specific IgH translocation groups, thus suggesting
that the transcriptional effects related to these translocations have impact that overcomes the signatures related to the specific disease. Moreover, the supervised analysis revealed that IGH translocations have transcriptional effects in pPCL cases that actually resembled in a virtually identical fashion those observed in MM, identifying transcripts that have been demonstrated as direct or indirect targets of the translocations in MM, as previously described (14). Conversely, no association has been found in our dataset between cytogenetic abnormalities and clinical outcome. Therefore, occurrence of IGH translocations in a very large fraction, if not almost the totality, of pPCL cases was not conceivable to justify per se the more aggressive phenotype of leukemic forms than the large majority of MM.

We thus aimed our analysis at investigating the transcriptional pathways differentially expressed in pPCL with regard to MM condition. Of the 503-gene signature identified, several genes have been identified as related to cytoskeleton functions and Rho protein signaling pathway (31-33), involved in cell adhesion/migration processes. Our analysis also highlighted the involvement of NF-kB pathway-associated genes, specifically acting as functional regulators in NF-kB pathway or potentially responsive to NF-kB transcriptional modulation. In particular, RelB subunit involved in NF-kB non-canonical pathway, whose activation was found associated to cell adhesion-mediated drug resistance (CAM-DR) in MM (34), was upregulated in pPCLs as well as the inhibitory subunits NFKBIE (IκBε) and NFKBIA (IκBα). In addition, TRAF1, TRAF4 and TRAF5, several members of the TNF receptor associated factor family with different roles in NF-kB pathway, were expressed at higher level in pPCLs (35, 36). Overall, these considerations, together with the concurrent findings of higher NF-kB indexes in pPCL cases, are suggestive and compatible with more activated NF-kB pathway and aggressive phenotype in pPCL. Conversely, a limited, if not absent, discrepancy between pPCL and MM has been identified, according to published expression-based indexes (24-26), as regards proliferation. This might be explained in that such indexes were mainly derived by the comparison between primary tumors and cell lines, that might therefore account for more
enhanced proliferation activity in relapsed/refractory myeloma evolving to a more aggressive and uncontrolled leukemic progression. The transcriptional pattern of pPCL was recently investigated by Usmani et al., who characterized the clinic-biological features of a series of pPCL patients, partly included in Total Therapy protocols trials (9), suggesting that pPCL represents a highly significant adverse feature of myeloma tumors in relationship with both overall and progression-free survival and complete remission duration. This was further supported by the occurrence of high-risk variables (advanced age; abnormal albumin, LDH and β2-microglobulin levels; chromosome abnormalities) that undoubtedly confer to pPCL an highly aggressive phenotype with very poor outcome. They described a 203-gene signature, obtained from 20 pPCL cases, that only in a minimal part (~15%) overlapped with our 503-gene list. As possible explanations to this finding, it could be considered the new generation arrays used in our analysis (leading to only 125 transcript commonly represented on the two platforms) and the different balance between pPCL and non-pPCL samples in the two datasets. Moreover, the Authors did not provide specific information on the molecular alterations (translocations, deletions, amplifications) of their cases; therefore, we could not exclude that different stratification within pPCL cases might affect the differential expression analysis. Furthermore, their data were not available in public repositories, which prevented any meta-analyses, validation or verification of similarities and differences between the two datasets.

Based on the list of differentially expressed genes between pPCL and MM, we took advantage of a composite dataset from multiple Institutions to investigate whether the modulation of expression levels of some transcripts included in the 503-gene signature might be correlated with progression in plasma cell dyscrasias, provided that the increased or decreased expression could be due to or associated with features characterizing highly injured plasma cells and/or aggressive clinical course. Among the 26 genes with a significant trend (8 decreasing and 18 increasing) from normal PC condition to PCL through the different forms
of PC dyscrasia, is worth mentioning the methyl-transferase EZH2, which expression is induced and correlates with tumor burden during MM disease progression and which is constitutive in IL-6-independent cell lines (27), and that was among the most overexpressed genes associated to proliferation (PR) MM subgroup in UAMS classification (26). Overall, the 26 genes identified here, showed significant modulation in the expression levels through the different progressive presentation of plasma cell dyscrasia, and exhibited further enhanced expression levels in HMCLs in line with the trend in primary tumors, thus strengthening the putative involvement of these genes in sustaining the intensity/aggressiveness of the neoplastic phenotype.

A further finding of the present study was the identification of distinct transcriptional signatures in groups of pPCL that clearly showed different outcome. The 27 genes carrying a highly significant correlation with OS were prevalently upregulated (17/27, 63%) in the pPCL patients with the poorest outcome. The occurrence of this peculiar pattern is predictive of poorer OS independently from both the major cytogenetic alterations, hematological parameters and known MM gene-risk models. Conversely, the low-risk signature is strongly associated with the ASCT procedure, which points to an overall more favorable prognosis. Further investigations on larger cohorts and validation sets are warranted to better elucidate whether a specific transcription pattern at diagnosis may be suggestive of the eligibility of the patients to ASCT with consequently associated better outcome. Interestingly, none of the 27 genes were recognized in known MM high-risk signatures (24, 28, 30), whereas only a minor fraction were dispersedly included in the gene lists characterizing the 7 MM molecular subgroups in UAMS database (26). Interestingly, the platelet/endothelial cell adhesion molecule 1 (PECAM1), whose expression resulted most significantly associated with OS, encoded a member of the immunoglobulin superfamily involved in angiogenesis and activation of integrins. Notably, a recent work described its involvement in the lenalidomide effect in angiogenesis inhibition and the interaction with cadherin 5 and β-catenin, which is
critical for endothelial cell cord formation (37). It could be speculated that the lower levels of 
\textit{PECAM1} associated with more aggressive disease might lead to less dependency of malignant PCs from microenvironment and reduced substrate for drugs, such as lenalidomide, that target microvessel formation.

The primary endpoint of the study was set to response after 4 cycle of therapy; six patients failed to reach response; the gene expression analysis highlighted three genes that clearly distinguished non-responders from responders (albeit at different treatment response, from partial to complete) to the treatment with LD. Little is known about the biologic role and the function of \textit{CYB5D2}, \textit{EDEM3} and \textit{YIPF6}, which leaves open the investigation of these genes as putatively involved in drug response or associated with PCs resistant phenotype. Additional studies on larger cohorts are warranted on this aspect.

Overall, our data suggest that pPCL might not represent an unique transcriptional and clinical entity, albeit clearly distinguishable from MM based on the expression of a large signature, and provide insights for further investigations of mechanisms underlying the biology of aggressive forms of plasma cell dyscrasia.

**Grant support**

This work was supported by grants from: Associazione Italiana Ricerca sul Cancro (AIRC) grant 5xmille #9980 2010-15 (to P.T., A.N., F.M.); AIRC grant IG10136 and Ministero Italiano dell’Istruzione, Università e Ricerca (MIUR) grant 2009PKMYA2 (to A.N.); Fondazione Matarelli, Milan, Italy. M.L. was supported by a fellowship from the Fondazione Italiana Ricerca sul Cancro (FIRC). P.M. is supported by a Research fund by Celgene.
References

27. Croonquist PA, Van Ness B. The polycomb group protein enhancer of zeste homolog 2 (EZH 2) is an oncogene that influences myeloma cell growth and the mutant ras phenotype. Oncogene. 2005;24:6269-80.
Legends to Figures

Figure 1. (A) Hierarchical clustering analysis of gene expression profiles of 21 pPCL cases. Samples are grouped according to the expression levels of the 1145 most variable genes. Main molecular alterations are shown; black indicates occurrence. (B) Heatmap of the 47 differentially expressed genes identified by multi-class analysis of 21 pPCL patients stratified into the three main IgH translocation groups. Visualization of the expression in MM samples is included. In all the heatmaps, the color scale bar represents the relative gene expression changes normalized by the standard deviation, and the color changes in each row represent gene expression relative to the mean across the samples.

Figure 2. Hierarchical clustering analysis of gene expression profiles of 21 pPCL and 55 MM cases. Samples are grouped according to the expression levels of the 1166 most variable genes. Main molecular alterations are indicated as black boxes, as well as the pPCL type in the last lane.

Figure 3. Boxplot distribution of the expression levels of the 3 differentially expressed genes identified by supervised analysis of pPCL patients stratified according to response rate. NR: non responder; PR: partial remission; VGPR: very good partial remission; CR: complete remission.

Figure 4. (A) Heatmap of the pPCL samples clustered on the 27 genes identified by globaltest as significantly associated with OS. Molecular features are indicated above the heatmap (black indicates occurrence). (B) Kaplan–Meier curves of the 2 groups defined by the 27-gene model. The colors of the groups correspond to those shown in the dendrogram of panel A.
Table 1. List of the 27 genes with p<.01 in globaltest analysis of expression data and OS

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Title</th>
<th>Cytoband</th>
<th>p-value</th>
<th>corr with survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>PECAM1</td>
<td>platelet/endothelial cell adhesion molecule</td>
<td>17q23.3</td>
<td>0.0006</td>
<td>pos</td>
</tr>
<tr>
<td>MKX</td>
<td>mohawk homeobox</td>
<td>10p12.1</td>
<td>0.0020</td>
<td>pos</td>
</tr>
<tr>
<td>FAM111B</td>
<td>family with sequence similarity 111, member B</td>
<td>11q12.1</td>
<td>0.0021</td>
<td>neg</td>
</tr>
<tr>
<td>MCTP1</td>
<td>multiple C2 domains, transmembrane 1</td>
<td>5q15</td>
<td>0.0022</td>
<td>neg</td>
</tr>
<tr>
<td>CALCRL</td>
<td>calcitonin receptor-like</td>
<td>2q32.1</td>
<td>0.0023</td>
<td>pos</td>
</tr>
<tr>
<td>C10orf10</td>
<td>chromosome 10 open reading frame 10</td>
<td>10q11.21</td>
<td>0.0028</td>
<td>neg</td>
</tr>
<tr>
<td>MKX</td>
<td>mohawk homeobox</td>
<td>10p12.1</td>
<td>0.0020</td>
<td>pos</td>
</tr>
<tr>
<td>EFEMP1</td>
<td>EGF-containing fibulin-like extracellular matrix protein 1</td>
<td>2p16.1</td>
<td>0.0030</td>
<td>neg</td>
</tr>
<tr>
<td>C3orf14</td>
<td>chromosome 3 open reading frame 14</td>
<td>3p14.2</td>
<td>0.0031</td>
<td>pos</td>
</tr>
<tr>
<td>ALDH1L2</td>
<td>aldehyde dehydrogenase 1 family, member L2</td>
<td>12q23.3</td>
<td>0.0032</td>
<td>pos</td>
</tr>
<tr>
<td>WARS</td>
<td>tryptophanyl-tRNA synthetase</td>
<td>14q32.2</td>
<td>0.0034</td>
<td>pos</td>
</tr>
<tr>
<td>SLC15A2</td>
<td>solute carrier family 15 (H+/peptide transporter), member 2</td>
<td>3q13.33</td>
<td>0.0043</td>
<td>pos</td>
</tr>
<tr>
<td>FAIM3</td>
<td>Fas apoptotic inhibitory molecule 3</td>
<td>1q32.1</td>
<td>0.0043</td>
<td>neg</td>
</tr>
<tr>
<td>CPEB4</td>
<td>cytoplasmic polyadenylation element binding protein 4</td>
<td>5q35.2</td>
<td>0.0048</td>
<td>neg</td>
</tr>
<tr>
<td>EDN1</td>
<td>endothelin 1</td>
<td>6p24.1</td>
<td>0.0055</td>
<td>neg</td>
</tr>
<tr>
<td>PVALB</td>
<td>parvalbumin</td>
<td>22q12.3</td>
<td>0.0056</td>
<td>neg</td>
</tr>
<tr>
<td>LY86</td>
<td>lymphocyte antigen 86</td>
<td>6p25.1</td>
<td>0.0058</td>
<td>neg</td>
</tr>
<tr>
<td>LAPTM5</td>
<td>lysosomal protein transmembrane 5</td>
<td>1q35.2</td>
<td>0.0065</td>
<td>neg</td>
</tr>
<tr>
<td>RNU5D</td>
<td>RNA, USD small nuclear</td>
<td>1p34.1</td>
<td>0.0077</td>
<td>pos</td>
</tr>
<tr>
<td>PARP15</td>
<td>poly (ADP-ribose) polymerase family, member 15</td>
<td>3q21.1</td>
<td>0.0079</td>
<td>neg</td>
</tr>
<tr>
<td>PLEKHF2</td>
<td>pleckstrin homology domain containing, family F (with FYVE domain) member 2</td>
<td>8q22.1</td>
<td>0.0085</td>
<td>neg</td>
</tr>
<tr>
<td>PDK4</td>
<td>pyruvate dehydrogenase kinase, isozyme 4</td>
<td>7q21.3</td>
<td>0.0089</td>
<td>neg</td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>tumor necrosis factor, alpha-induced protein 3</td>
<td>6q23.3</td>
<td>0.0089</td>
<td>neg</td>
</tr>
<tr>
<td>FAM105A</td>
<td>family with sequence similarity 105, member A</td>
<td>5p15.2</td>
<td>0.0091</td>
<td>neg</td>
</tr>
<tr>
<td>CTH</td>
<td>cystathionase (cystathionine gamma-lyase)</td>
<td>1p31.1</td>
<td>0.0091</td>
<td>pos</td>
</tr>
<tr>
<td>HOOK1</td>
<td>hook homolog 1 (Drosophila)</td>
<td>1p32.1</td>
<td>0.0092</td>
<td>pos</td>
</tr>
<tr>
<td>TCN2</td>
<td>transcobalamin II</td>
<td>22q12.2</td>
<td>0.0094</td>
<td>neg</td>
</tr>
</tbody>
</table>
Table 2. Multivariate analysis comparing the 27-gene model with molecular variables and UAMS, IFM and UK gene-risk models in pPCL series. The model was available for all covariates with the exception of t(4;14), due to the low number of occurrence; and of ASCT, whose occurrence was inversely correlated with being part of the high-risk group, that prevented the calculation of hazard ratio model in relationship with the configuration of survival data versus covariates value.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>HR^a</th>
<th>lo95%CI^b</th>
<th>up95%CI^b</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>27-gene model</td>
<td>33.28</td>
<td>2.77</td>
<td>400.09</td>
<td>5.73E-03</td>
</tr>
<tr>
<td>del(13q)</td>
<td>4.89</td>
<td>0.37</td>
<td>64.92</td>
<td>2.29E-01</td>
</tr>
<tr>
<td>27-gene model</td>
<td>20.54</td>
<td>2.42</td>
<td>174.38</td>
<td>5.61E-03</td>
</tr>
<tr>
<td>del(17p)</td>
<td>0.93</td>
<td>0.19</td>
<td>4.41</td>
<td>9.24E-01</td>
</tr>
<tr>
<td>27-gene model</td>
<td>36.37</td>
<td>2.76</td>
<td>478.67</td>
<td>6.27E-03</td>
</tr>
<tr>
<td>gain(1q)</td>
<td>0.37</td>
<td>0.05</td>
<td>2.84</td>
<td>3.39E-01</td>
</tr>
<tr>
<td>27-gene model</td>
<td>19.07</td>
<td>2.23</td>
<td>163.11</td>
<td>7.10E-03</td>
</tr>
<tr>
<td>del(1p)</td>
<td>0.97</td>
<td>0.2</td>
<td>4.66</td>
<td>9.72E-01</td>
</tr>
<tr>
<td>27-gene model</td>
<td>35</td>
<td>3.19</td>
<td>383.61</td>
<td>3.61E-03</td>
</tr>
<tr>
<td>t(11;14)</td>
<td>3.2</td>
<td>0.47</td>
<td>22</td>
<td>2.36E-01</td>
</tr>
<tr>
<td>27-gene model</td>
<td>20.26</td>
<td>2.35</td>
<td>174.38</td>
<td>6.15E-03</td>
</tr>
<tr>
<td>MAF-translocations</td>
<td>1.28</td>
<td>0.26</td>
<td>6.41</td>
<td>7.65E-01</td>
</tr>
<tr>
<td>27-gene model</td>
<td>16.28</td>
<td>1.79</td>
<td>148.15</td>
<td>1.33E-02</td>
</tr>
<tr>
<td>del(8p)</td>
<td>0.83</td>
<td>0.14</td>
<td>4.81</td>
<td>8.33E-01</td>
</tr>
<tr>
<td>27-gene model</td>
<td>16.30</td>
<td>1.68</td>
<td>158.46</td>
<td>1.61E-02</td>
</tr>
<tr>
<td>UAMS 70-gene model</td>
<td>1.87</td>
<td>0.34</td>
<td>10.21</td>
<td>4.70E-03</td>
</tr>
<tr>
<td>27-gene model</td>
<td>19.92</td>
<td>2.30</td>
<td>172.94</td>
<td>6.66E-03</td>
</tr>
<tr>
<td>UAMS 17-gene model</td>
<td>1.24</td>
<td>0.22</td>
<td>6.92</td>
<td>8.08E-01</td>
</tr>
<tr>
<td>27-gene model</td>
<td>20.54</td>
<td>2.43</td>
<td>173.92</td>
<td>5.56E-03</td>
</tr>
<tr>
<td>IFM 15-gene model</td>
<td>0.79</td>
<td>0.15</td>
<td>4.20</td>
<td>7.79E-01</td>
</tr>
<tr>
<td>27-gene model</td>
<td>20.63</td>
<td>2.40</td>
<td>177.53</td>
<td>5.84E-03</td>
</tr>
<tr>
<td>UK 6-gene model</td>
<td>1.38</td>
<td>0.29</td>
<td>6.51</td>
<td>6.83E-01</td>
</tr>
</tbody>
</table>

^a HR = Hazard Ratio; ^b lo95%CI, up95%CI = lower and upper bound, respectively, of the 95% confidence interval of the hazard ratio
Transcriptional characterization of a prospective series of primary plasma cell leukemia revealed signatures associated with tumor progression and poorer outcome

Katia Todoerti, Luca Agnelli, Sonia Fabris, et al.

Clin Cancer Res  Published OnlineFirst April 18, 2013.