

## ARTICLE

## Open Access

# Transcriptome analysis reveals significant differences between primary plasma cell leukemia and multiple myeloma even when sharing a similar genetic background

Elizabeta A. Rojas<sup>1,2</sup>, Luis A. Corchete<sup>1,2</sup>, María Victoria Mateos<sup>1,2,3</sup>, Ramón García-Sanz<sup>1,2,3,4</sup>, Irena Misiewicz-Krzeminska<sup>1,2,5</sup> and Norma C. Gutiérrez<sup>1,2,3,4</sup>

## Abstract

Primary plasma cell leukemia (pPCL) is a highly aggressive plasma cell dyscrasia characterised by short remissions and very poor survival. Although the 17p deletion is associated with poor outcome and extramedullary disease in MM, its presence does not confer the degree of aggressiveness observed in pPCL. The comprehensive exploration of isoform expression and RNA splicing events may provide novel information about biological differences between the two diseases. Transcriptomic studies were carried out in nine newly diagnosed pPCL and ten MM samples, all of which harbored the 17p deletion. Unsupervised cluster analysis clearly distinguished pPCL from MM samples. In total 3584 genes and 20033 isoforms were found to be deregulated between pPCL and MM. There were 2727 significantly deregulated isoforms of non-differentially expressed genes. Strangely enough, significant differences were observed in the expression of spliceosomal machinery components between pPCL and MM, in respect of the gene, isoform and the alternative splicing events expression. In summary, transcriptome analysis revealed significant differences in the relative abundance of isoforms between pPCL and MM, even when they both had the 17p deletion. The mRNA processing pathway including RNA splicing machinery emerged as one of the most remarkable mechanisms underlying the biological differences between the two entities.

## Introduction

Plasma cell leukemia (PCL) is an uncommon and aggressive plasma cell dyscrasia, characterised by the presence of more than 20% of plasma cells (PCs) and an absolute number of  $\geq 2 \times 10^9/L$  of PCs in peripheral blood<sup>1,2</sup>. PCL is classified as primary (pPCL) when detected de novo in patients with no evidence of previous multiple myeloma (MM), or as secondary (sPCL) in patients with relapsed or refractory MM that progresses

to a leukemic phase<sup>1,3,4</sup>. Around 60% and 40% of PCLs are pPCL and sPCL, respectively<sup>5</sup>. pPCL accounts for less than 3% of all malignant plasma cell disorders<sup>6–8</sup>. As a consequence, the current knowledge regarding the molecular basis of pPCL is quite limited by the small number of cases included in most series.

Various studies have reported that pPCL patients show different clinical and biological features from those of MM. pPCL is associated with a dismal prognosis, whereby median survival is about 10 months. Extramedullary plasmacytomas, renal failure and massive bone marrow infiltration are more frequently observed in pPCL than in MM. Conversely, pPCL exhibits a lower prevalence of bone disease<sup>1,2,9–11</sup>.

Cytogenetic studies have shown that pPCL features elevated genomic instability, especially with respect to,

Correspondence: Norma C. Gutiérrez ([normagu@usal.es](mailto:normagu@usal.es))

<sup>1</sup>Cancer Research Center-IBMCC (USAL-CSIC), Salamanca, Spain

<sup>2</sup>Institute of Biomedical Research of Salamanca (IBSAL), Salamanca, Spain  
Full list of author information is available at the end of the article.

These authors contributed equally: Irena Misiewicz-Krzeminska, Norma C. Gutiérrez

© The Author(s) 2019



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

karyotypic complexity, and higher prevalence of 17p13 deletions and 1q gains. Of the translocations involving the immunoglobulin heavy chain locus (*IGH*), t(11;14) and t(14;16) are more frequent in pPCL than in MM<sup>11–14</sup>. In recent years, the development of high-throughput technologies has given rise to a detailed knowledge about molecular characteristics of pPCL. Several studies have investigated the gene mutation patterns<sup>15</sup>, and differential gene and miRNAs expression profiles<sup>16–20</sup>, establishing differences and similarities between pPCL and sPCL, and MM.

Many of the genomic differences detected between pPCL and MM could be attributed to the dissimilar distribution of genetic abnormalities between the two entities<sup>15</sup>. For instance, the overrepresentation of 17p deletions in pPCL could explain some of the differences observed in the molecular and genomic patterns of pPCL and MM. However, a more unfavourable genetic background does not fully explain the ominous prognosis of this plasma cell neoplasm. In fact, although 17p deletion is associated with poor outcome and extramedullary disease in MM, the presence of this abnormality does not confer the degree of aggressiveness observed in pPCL.

We surmised that the comparison of the transcriptome profiles of pPCL and MM cases bearing similar cytogenetic abnormalities could provide valuable insights into new molecular mechanisms responsible for the different clinical outcomes of both diseases, and not reliant on particular chromosomal abnormalities. For this purpose, we analysed the transcriptome of pPCL and MM patients, using samples with 17p deletion and a similar cytogenetic profile, and focusing not only on gene expression profiling previously studied by other authors, but also on isoform expression and alternative splicing patterns.

## Materials and methods

For more specific information, see the *Online Supplementary File*.

### Patients and samples

Nine patients with newly diagnosed pPCL and ten with newly diagnosed MM were selected for the study based on the presence of 17p deletion. The median age of patients was 61 years (range: 54–89 years) for MM patients and 64 years (range: 45–83 years) for pPCL patients. The research ethics committee of the University Hospital of Salamanca approved the study in accordance with the Declaration of Helsinki principles.

Plasma cells were isolated from bone marrow samples using the AutoMACs immunomagnetic system (Miltenyi Biotec), as previously described<sup>21</sup>. Purity was greater than 95% in all MM and pPCL cases. Interphase fluorescence *in situ* hybridisation studies were performed as previously described<sup>21</sup>.

### Nucleic acid extraction, quantitative real-time PCR and TP53 mutation analysis

RNA and DNA were extracted using AllPrep DNA/RNA Mini Kit (Qiagen). RNA integrity was assessed using Agilent 2100 Bioanalyzer. Total RNA (200 ng) was reverse-transcribed to cDNA using the SuperScript First-Strand Synthesis System (Thermo Fisher). Gene expression was evaluated with TaqMan (Applied Biosystems) qRT-PCR assays, and isoform and exon expression were evaluated using SYBR Green (Bio-Rad) qRT-PCR mRNA assays. *PGK1* and *18S* genes were used as endogenous controls for gene and isoform/exon expressions, respectively, and the relative expression was expressed as  $2^{-\Delta Ct}$ .

*TP53* mutation status was determined using 100 ng of genomic DNA by Sanger sequencing, as previously reported<sup>22</sup>.

### Human transcriptome arrays

Total RNA was amplified, labelled and hybridised to GeneChip® Human Transcriptome Array 2.0 from Affymetrix. Complete microarray data are available from the Gene Expression Omnibus under accession number GSE131216. Unsupervised multidimensional scaling (MDS) was performed using the Euclidean distance as the distance measure and the group average as the linkage method. Samples in the MDS from the CoMMpass dataset were classified in subgroups using model-based methods from the mclust R package.

Differential expression was analysed using the samr package (v2.0) in R. Gene and isoform level probesets with expression values less than the mean of the microarray antigenomic control expression values in all samples were excluded from further analysis. Genes and isoforms with a value of  $q < 0.05$  were considered statistically significant and were selected for the KEGG pathway overrepresentation analysis in the Webgestalt suite<sup>23</sup>. Alternative splicing (AS) analysis was performed using the Affymetrix Transcriptome Analysis Console (TAC) version 4.0.1.36.

### Computational RNA-binding site prediction

The SpliceAid database was used to identify potential exonic splicing enhancer (ESE) motifs recognised by human serine/arginine-rich (SR) proteins<sup>24</sup>.

### Statistical analysis

The significance of differences in experiments was assessed by the two-sided Student's *t* test for unpaired samples. Pearson correlation coefficients were calculated to measure associations between the results of HTA 2.0 and qRT-PCR assays. Fisher's exact tests with two-tailed were calculated using the GraphPad QuickCalcs website: <https://www.graphpad.com/quickcalcs/contingency1/>. Values of *p* and *q* less than 0.05 were considered

statistically significant. All statistical analyses were conducted using IBM SPSS Statistics 22.0 and the SIMFIT v7.0.9 package (Bardsley, University of Manchester, UK).

## Results

### Genetic abnormalities in pPCL and MM patients

All pPCL and MM samples carried the 17p deletion. A balanced distribution of other genetic abnormalities in the two entities was also taken into account when selecting the samples. Genetic abnormalities analysed by FISH are summarised in Table 1. The frequencies of *IGH* translocations according to 14q32 partners, chromosome 1 abnormalities and 13q deletions are shown in Table 1. Fisher's exact test revealed no statistically significant differences between MM and pPCL (Table 1).

To better characterise the *TP53* status, mutations were analysed in all patients with an available genomic DNA sample (18/19 samples). *TP53* mutations were identified in four of eight pPCLs (50%) and three of ten MM (30%) ( $p = 0.63$ , Fisher's exact test). Six of the seven identified mutations were missense substitutions (86%), and were localised in the DNA binding domain, distributed between exons 5 and 8. The mutated amino acid residues were in areas with hotspot properties and direct DNA contact site. According to their functionality as transcription factor the missense mutations detected in the seven samples predicted an inactive p53 (Supplementary Table 1).

### Differentially expressed genes between pPCL and MM

To investigate the transcriptome signature of pPCL and MM patients, Human Transcriptome Array 2.0 analysis was performed. These microarrays allow us to explore gene expression profile, as well as splicing events and isoform expression, using millions of probes that cover exon-exon splice junctions, and coding and non-coding isoforms.

First, we explored the distribution of the samples in clusters based on similarities in gene expression data

using unsupervised multidimensional scaling (MDS). The first component in this analysis clearly separated MM from pPCL samples (Fig. 1a), although two pPCLs were closer to MM samples. We asked whether different clusters could also be identified across a large series of MM cases with del17p. We used RNA-Seq data from 73 MM patients with del17p from the MMRF CoMMpass trial (NCT01454297) to address this question. We found no clusters, but samples were apparently randomly distributed (Supplementary Fig. 1).

Gene expression analysis from HTAs identified a total of 3584 genes differentially expressed between the pPCL and MM groups, of which 367 were overexpressed and 3217 genes were underexpressed in pPCL relative to MM patients (Fig. 1b). The list of the 40 genes with the greatest and least fold change (FC) is shown in Fig. 1c. Functional analysis using the KEGG pathway database revealed statistically significant enrichment for spliceosome, aminoacyl-tRNA biosynthesis, protein processing in endoplasmic reticulum (ER) and RNA transport (Fig. 1d). The heatmap, using only the genes from overrepresented KEGG pathways, distinguished two entities with different gene expression patterns and with a clear prevalence of underexpression in pPCL compared with MM (Supplementary Fig. 2).

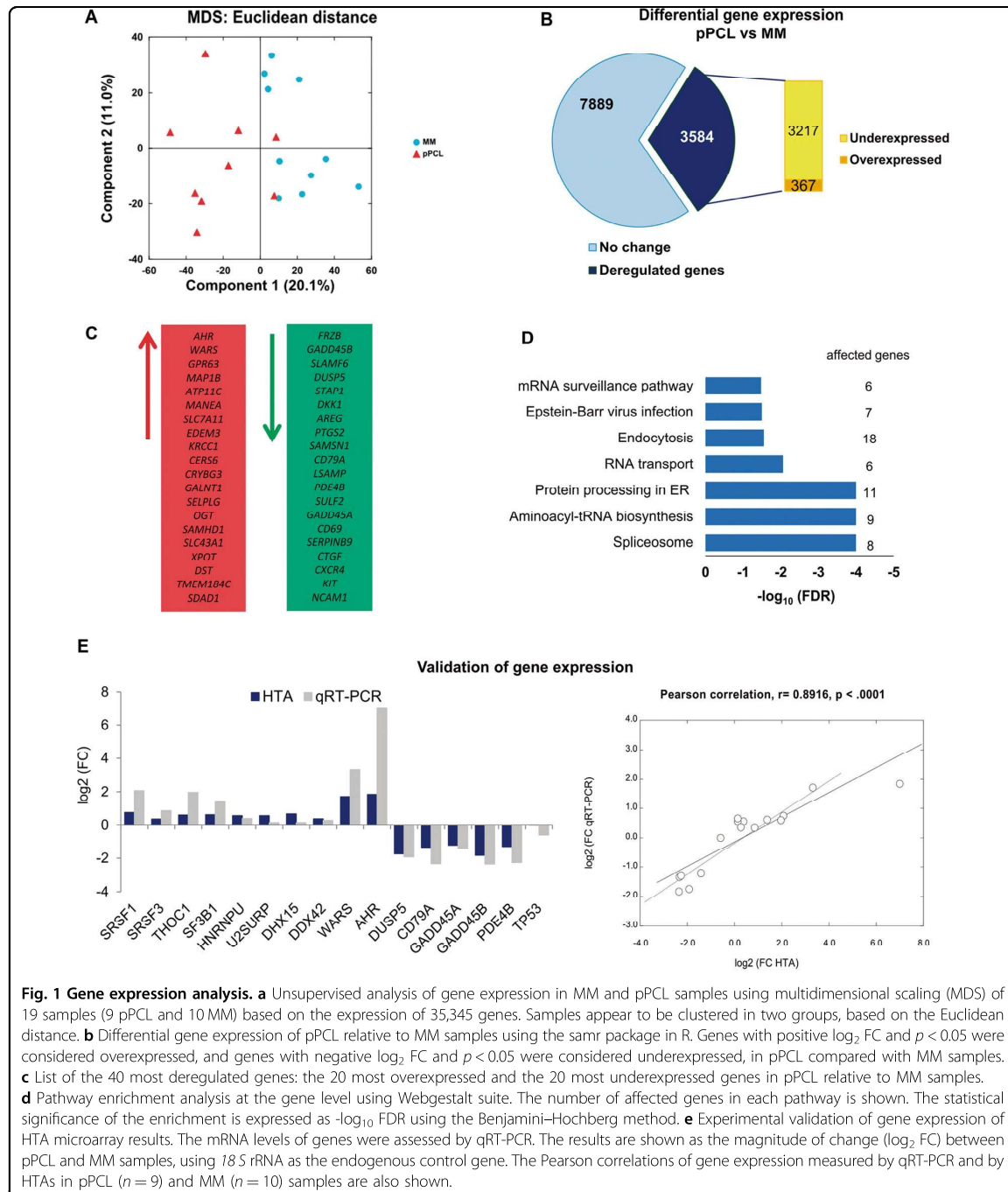
Results of gene expression profiling were validated by qRT-PCR. We chose 15 genes, eight components of the spliceosome selected from the functional analysis using KEGG database (*SRSF1*, *SRSF3*, *SF3B1*, *DDX42*, *DHX15*, *HNRNPU*, *THOC1*, and *U2SURP*), and seven genes from the list of the 40 most deregulated genes, whose role in cancer is less well understood (*WARS*, *AHR*, *DUSP5*, *CD79A*, *GADD45A*, *GADD45B*, and *PDE4B*). *TP53* gene expression was also quantified. The primers used in these assays are shown in Supplementary Table 2. As shown in Fig. 1e, we observed a high concurrence in the FC direction between HTA and qRT-PCR analysis for all the 16 genes considered. These results also revealed a strong positive and statistically significant correlation ( $r = 0.89$ ,  $p < 0.0001$ ) between the HTA microarray and qRT-PCR results (Fig. 1e).

**Table 1** Prevalence of genetic aberrations detected by FISH in pPCL and MM

	pPCL (n = 9)	MM (n = 10)	p-value (Fisher's exact test)
<b>17p del</b>	9/9 (100%)	10/10 (100%)	1.00
<b>13q del</b>	7/9 (78%)	6/10 (60%)	0.63
<b>1q gain</b>	5/9 (56%)	4/10 (40%)	0.65
<b>1p del</b>	2/9 (22%)	1/10 (10%)	0.58
<b>t(11;14)</b>	3/9 (33%)	2/10 (20%)	0.63
<b>t(4;14)</b>	2/9 (33%)	0/10 (0%)	0.21
<b>t(14;16)</b>	1/9 (11%)	0/10 (0%)	0.47

### Differentially alternative splicing events in pPCL and MM

Next, we aimed to identify alternative splicing events (ASEs) differentially expressed between pPCL and MM patients. We detected 2873 genes affected by at least one ASE, and the expression of 3303 splicing events (Fig. 2a). The most abundant ASEs were the cassette exon, with 2225 exons included or excluded from mRNAs, followed by alternative 5' donor site, alternative 3' acceptor site, and intron retention (Fig. 2a). The differential expression of these four types of ASEs in pPCL relative to MM is shown in Fig. 2b. In this regard, among all ASEs we found a large number of alternatively spliced exons differentially

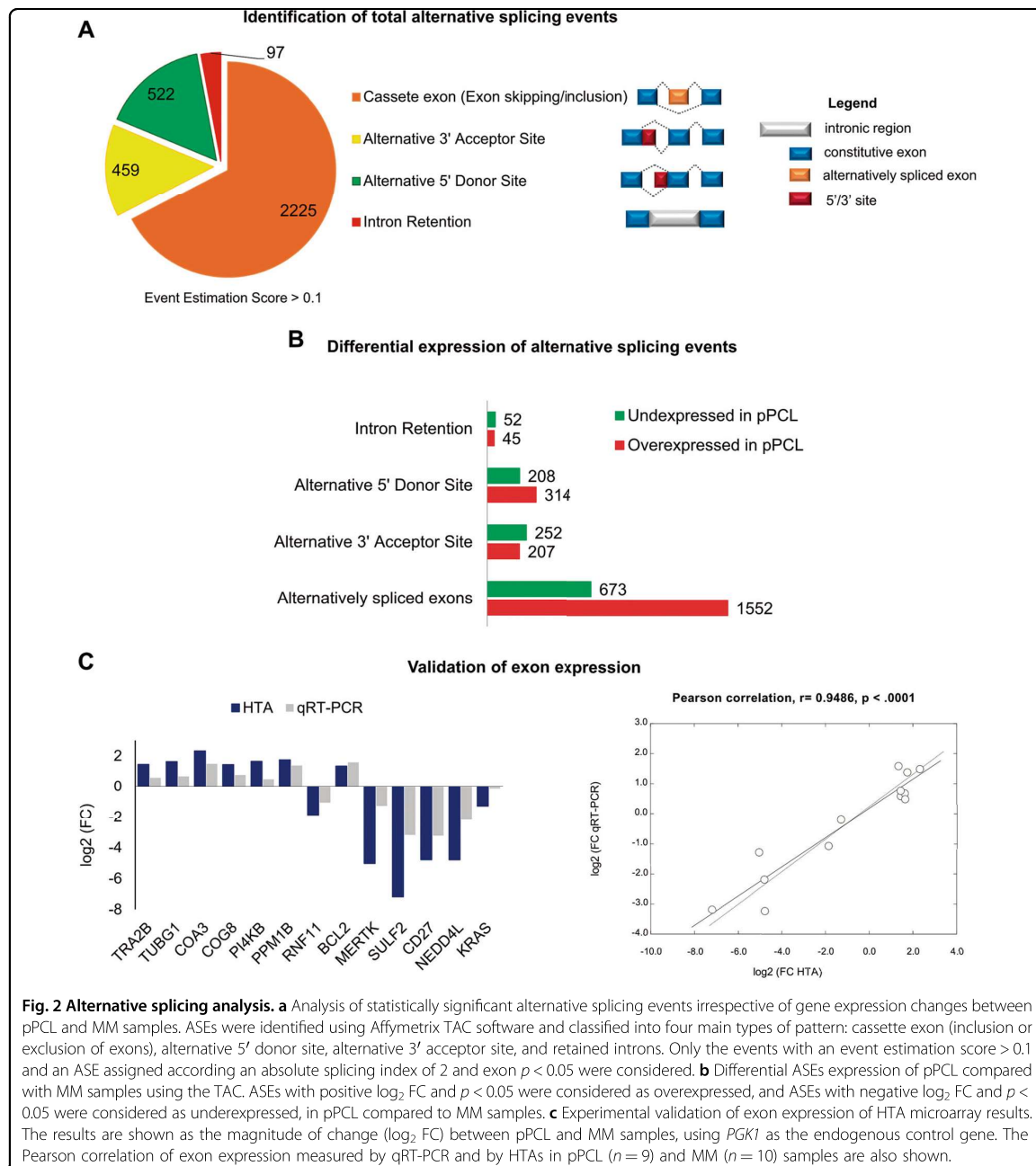


**Fig. 1 Gene expression analysis.** **a** Unsupervised analysis of gene expression in MM and pPCL samples using multidimensional scaling (MDS) of 19 samples (9 pPCL and 10 MM) based on the expression of 35,345 genes. Samples appear to be clustered in two groups, based on the Euclidean distance. **b** Differential gene expression of pPCL relative to MM samples using the samr package in R. Genes with positive  $\log_2$  FC and  $p < 0.05$  were considered overexpressed, and genes with negative  $\log_2$  FC and  $p < 0.05$  were considered underexpressed, in pPCL compared with MM samples. **c** List of the 40 most deregulated genes: the 20 most overexpressed and the 20 most underexpressed genes in pPCL relative to MM samples. **d** Pathway enrichment analysis at the gene level using Webgestalt suite. The number of affected genes in each pathway is shown. The statistical significance of the enrichment is expressed as  $-\log_{10}$  FDR using the Benjamini–Hochberg method. **e** Experimental validation of gene expression of HTA microarray results. The mRNA levels of genes were assessed by qRT-PCR. The results are shown as the magnitude of change ( $\log_2$  FC) between pPCL and MM samples, using 18 S rRNA as the endogenous control gene. The Pearson correlations of gene expression measured by qRT-PCR and by HTAs in pPCL ( $n = 9$ ) and MM ( $n = 10$ ) samples are also shown.

expressed between the two groups (Fig. 2b). KEGG pathway analysis of genes affected by deregulation of alternatively spliced exons, revealed that FoxO signalling, RNA transport, glucagon signalling, phospholipase D signalling and

Epstein–Barr virus (EBV) infection pathways were the most significantly enriched (Supplementary Fig. 3).

We validated the differential expression of the cassette exon event of 13 selected genes (*TRA2B*, *TUBG1*, *BCL2*,



*COA3*, *COG8*, *PI4KB*, *PPM1B*, *RNF11*, *MERTK*, *SULF2*, *CD27*, *KRAS* and *NEDD4L*) by qRT-PCR (Supplementary Table 3). The selection criteria were statistically significant exons ( $p < 0.05$ ) with the highest and lowest FC between pPCL and MM, and longer than 70 nucleotides. The primers used for these assays are shown in

Supplementary Table 4. Thus, the downregulation of exons of *SULF2*, *CD27*, *NEDD4L*, *RNF11*, *MERTK* and *KRAS* genes, and the upregulation of exons of *TRA2B*, *TUBG1*, *BCL2*, *COA3*, *COG8*, *PI4KB*, and *PPM1B* genes observed in the HTA microarrays of pPCL were validated by qRT-PCR (Supplementary Fig. 4). Furthermore, the

Pearson correlation analysis revealed a strong, positive and statistically significant correlation ( $r = 0.9488$ ,  $p < 0.0001$ ) between HTA microarrays and qRT-PCR results of the 13 selected alternatively spliced exons (Fig. 2c).

It is well known that changes in the expression of different spliceosome genes can affect alternative splicing profile of cancer cells<sup>25–27</sup>. In particular, some of the components of the SR protein family have been shown to be upregulated in various cancers, in which regulate the splicing of many genes. These data strongly emphasise the important role of alternative splicing in tumorigenesis<sup>28,29</sup>. Therefore, we investigated whether the deregulated gene expression of spliceosome components found in pPCL as compared to MM, was correlated with the occurrence of deregulated splicing events detected in this study. We focused on *SRSF1* and *SRSF3* genes, whose overexpression in pPCL was validated by qRT-PCR (as shown in Fig. 1e and Supplementary Fig. 5A). A high correlation between *SRSF1* and *SRSF3* was observed (Supplementary Fig. 5A), which is noteworthy because the expression of *SRSF1* and *SRSF3* genes is mutually regulated and coexpressed in normal and tumour cells, and *SRSF3* is able to regulate the alternative splicing of *SRSF1* gene<sup>30</sup>.

We next explored the correlation between *SRSF1* and *SRSF3* genes with the 13 validated exons (Fig. 5b) alternatively spliced and differentially expressed between pPCL and MM (*TRA2B*, *TUBG1*, *BCL2*, *COA3*, *COG8*, *PI4KB*, *PPM1B*, *RNF11*, *MERTK*, *SULF2*, *CD27*, *KRAS*, and *NEDD4L*). We found a negative significant correlation between *SRSF1* and the expression of *CD27* exon 2 ( $r = -0.71$ ,  $p = 0.0006$ ), *NEDD4L* exon 3 ( $r = -0.55$ ,  $p = 0.0130$ ) and *RNF11* exon 3 ( $r = -0.4755$ ,  $p = 0.0396$ ) (Supplementary Fig. 5B, C). In the case of *SRSF3*, a negative significant correlation with *CD27* exon 2 ( $r =$

$-0.66$ ,  $p = 0.0019$ ) and *MERTK* exon 3 ( $r = -0.5587$ ,  $p = 0.0129$ ) was also detected (Supplementary Fig. 5B, D).

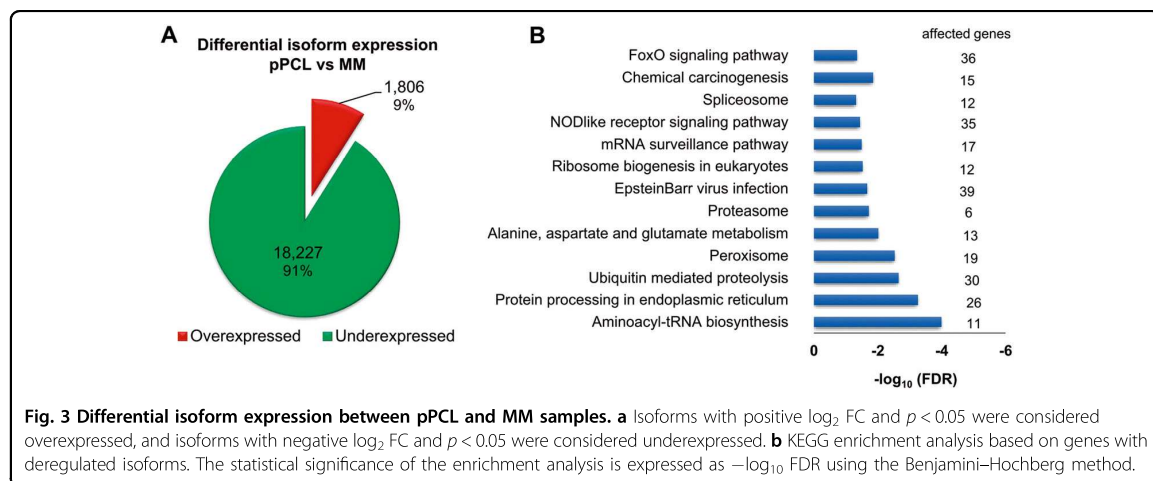
These results indicate that the alternative splicing of *CD27*, *NEDD4L*, and *RNF11* genes could be regulated by the splicing factor *SRSF1*, and the alternative splicing of *MERTK* and *CD27* genes could be controlled by *SRSF3*.

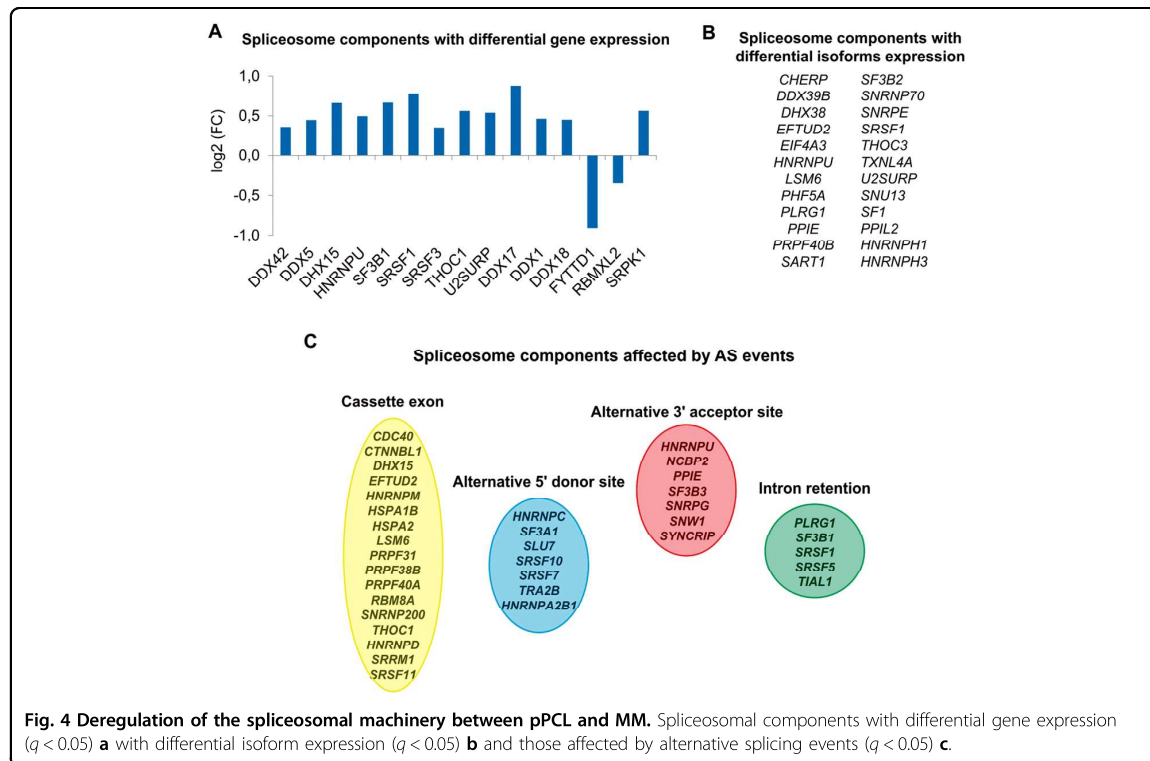
#### Differentially expressed isoforms between pPCL and MM

Alternative splicing, carried out by the spliceosome machinery, leads to mRNA isoforms that encode different proteins. Since we had observed differential expression of ASEs between MM and pPCL, we searched for differences in mRNA isoform expression. We identified 20033 deregulated isoforms associated with 6254 gene IDs curated by the HGNC (Hugo Gene Nomenclature Committee), of which 1806 isoforms were overexpressed and 18227 were underexpressed in pPCL patients relative to MM patients (Fig. 3a). The enrichment analysis identified EBV infection, spliceosome and proteasome pathways, among others, as statistically significantly overrepresented KEGG pathways (Fig. 3b). In the unsupervised analysis based on isoform expression, the samples tended to be separated into two groups, as was the case with gene expression analysis (Supplementary Fig. 6).

It should be pointed out that across the entire transcriptome analysis we found significant differences in the expression of the components of the spliceosomal machinery between pPCL and MM. We found 15 spliceosome genes differentially expressed between the two groups. Moreover, 24 spliceosome components displayed differential expression of some of their isoforms, and 36 were affected by AS events (Fig. 4).

In recent years, some studies have reported the advantage of the analysis of isoform expression pattern compared with standard gene expression profiling in cancer research<sup>31,32</sup>. Therefore, we also focused on those



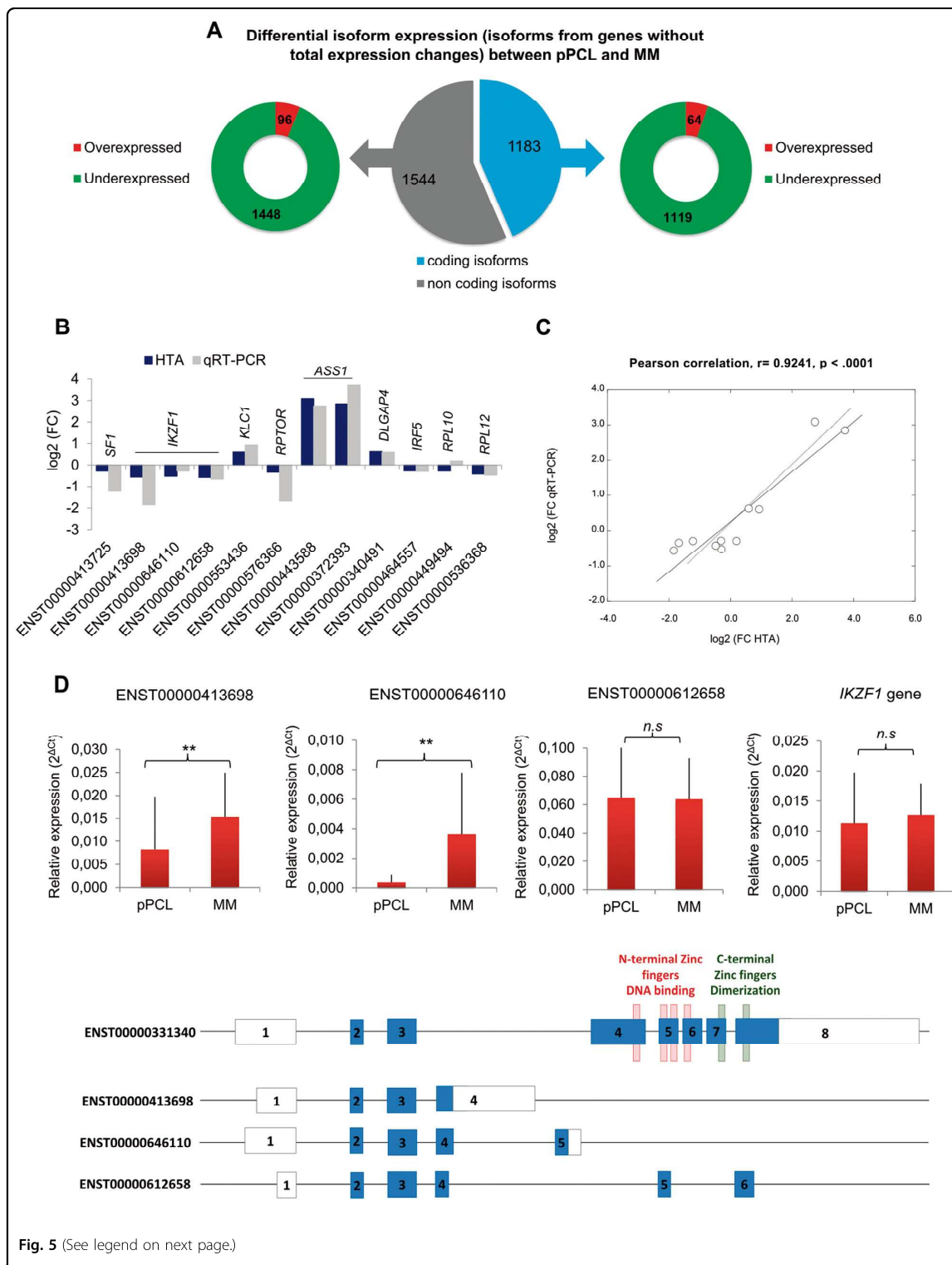


expressed genes (selection criteria for “expressed genes” in Materials and Methods section) that were not differentially expressed, but whose isoforms were differentially expressed. We found 2727 deregulated isoforms belonging to 1249 expressed genes, 1183 of which were coding isoforms (Fig. 5a). A prevalence of underexpressed isoforms in pPCL patients relative to MM was also found.

To validate the differential isoform expression detected by HTA, we focused on the coding isoforms, of which 64 were overexpressed and 1119 underexpressed in pPCL compared to MM patients (Supplementary Fig. 7A). We selected 12 coding isoforms differentially expressed ( $q < 0.05$ ) and distributed according to high, medium and low FC, to be validated by RT-PCR (Supplementary Fig. 7A). In the case of genes with two or three deregulated isoforms a specific sequence was needed in each isoform for primer design. The selected isoforms and the primers used are shown in Supplementary Tables 5 and 6, respectively.

The differential expression of ten of the twelve isoforms tested by qRT-PCR was consistent with the results obtained by HTA (Fig. 5b). Subtle discrepancies were observed for isoforms ENST00000449494 (*RPL10* isoform) and ENST00000612658 (*IKZF1* isoform). A strong, positive and statistically significant correlation ( $r = 0.92$ ,  $p < 0.0001$ ) between the two methods was observed (Fig. 5c).

The overexpressed isoforms ENST00000553436 (*KLC1* gene), ENST00000443588 and ENST00000372393 (*ASS1* gene), and ENST00000340491 (*DLGAP4* gene) are isoforms coding for protein domains, which should be biologically effective and able to exert their function. On the other hand, the isoforms ENST00000413698 and ENST00000646110 (*IKZF1* gene), ENST00000576366 (*RPTOR* gene) and ENST00000413725 (*SF1* gene) are among the underexpressed isoforms, which include regions coding for protein domains whose lack probably lead to loss of some of the protein functions. We were struck in particular by the *IKZF1* gene (Ikaros protein), which was among the genes containing various differentially expressed isoforms. The *IKZF1* gene is composed of eight exons and produces 23 splice variants by alternative splicing, which are differentiated by alternative use of exons and subsequently in their functional domain composition, with distinct biological functions (Supplementary Fig. 6B). Six *IKZF1* isoforms were underexpressed ( $q < 0.05$ ) in the HTA analysis comparing pPCL and MM (Supplementary Fig. 7C). Among the four protein coding isoforms, three could be quantified by qRT-PCR, considering the specific sequences in each isoform. The coding isoforms ENST00000413698 and ENST00000646110 were significantly underexpressed in pPCL ( $p < 0.01$ ) as revealed by qRT-PCR, while the *IKZF1*





(see figure on previous page)

**Fig. 5 Identification of differentially expressed isoforms of genes without total expression changes.** **a** Classification of deregulated isoforms according to coding (in blue) and non-coding (in grey) isoforms among the non-differentially expressed genes. The right side shows the coding isoforms that were overexpressed and underexpressed in pPCL relative to MM. The left side shows the non-coding isoforms that were overexpressed and underexpressed in pPCL relative to MM. For both, coding and non-coding isoforms with positive  $\log_2$  FC and  $p < 0.05$  were considered overexpressed, and isoforms with negative  $\log_2$  FC and  $p < 0.05$  were considered underexpressed. Genes with a value of  $p > 0.05$  were considered not to be differentially expressed. **b** Experimental validation of differential isoform expression from genes without total expression changes in the HTA microarrays. The mRNA levels of isoforms (Ensembl ID) were assessed by qRT-PCR. The results are shown as the  $\log_2$  FC between pPCL and MM samples, using the *PGK1* gene as the endogenous control. The expression difference of ten isoforms between pPCL and MM was statistically significant, and the expression difference of ENST00000449494 (*RPL10* isoform) and ENST00000612658 (*IKZF1* isoform) between pPCL and MM were not statistically significant. **c** Pearson correlation of isoform expression measured by qRT-PCR and by HTAs in pPCL ( $n = 9$ ) and MM ( $n = 10$ ) samples. **d** *IKZF1* isoform analysis. The mRNA levels of ENST00000413698, ENST00000646110, ENST00000612658 and *IKZF1* gene were assessed by qRT-PCR. The results are shown as the relative expression (calculated as  $2^{-\Delta\Delta C_t}$ )  $\pm$  s.d between pPCL and MM samples, using *PGK1* as endogenous control gene. Statistically significant differences between pPCL and MM samples are represented as \*\* $p < 0.01$ , \* $p < 0.05$  and *n.s.* non-significant (unpaired two-sided *Student's t* test). Schematic representation of the three downregulated *IKZF1* coding isoforms analysed by qRT-PCR in pPCL compared with MM samples is provided below. The longest isoform (ENST00000331340, NM\_006060) contains four N-terminal zinc finger motifs (red), and two C-terminal zinc fingers (green). Exons are numbered from 1 to 8. Coding exons are represented by blue boxes and UTR regions by black boxes.

gene expression was equally expressed in both plasma cell dyscrasias, corroborating the HTA results (Fig. 5d). However, we found similar expression of the isoform ENST00000612658 in pPCL and MM (Fig. 5d).

#### Identification of potential exonic splicing enhancers

Alternative splicing produces different isoforms from a single gene, contributing significantly to proteomic diversity. This process is mediated by the splicing machinery, the spliceosome, which comprises more than 300 splicing factors that bind to specific sequence elements in the pre-mRNA sequence. Among the splice site signals in this sequence, exonic splicing enhancers (ESEs) are short and degenerative sequences that enhance splice-site recognition in constitutively and alternatively spliced exons<sup>33</sup>. ESEs act as binding sites for SR RNA-binding proteins, a family of conserved splicing factors that participate in multiple steps of the spliceosome assembly, during RNA splicing<sup>34</sup>. A broad range of sequences can function as ESEs in a particular context, and distinct SR proteins have different ESE specificities. The development of several resources for identifying ESEs binding sites and the global deregulation of the spliceosomal machinery prompted us to use one of these databases<sup>24</sup> to look for potential ESE motifs recognised by human SR proteins in our transcriptome dataset.

First, we examined exons differentially expressed between pPCL and MM that were validated by qRT-PCR assays. We only considered ESEs with a score of  $\geq 5$  in the SpliceAid database. The prediction algorithm used identified a very large number of ESEs. The analysis of the 13 exon sequences (*TRA2B*, *TUBG1*, *BCL2*, *COA3*, *COG8*, *PIAKB*, *PPM1B*, *RNF11*, *MERTK*, *SULF2*, *CD27*, *KRAS*, and *NEDD4L* genes) based on the SpliceAid database, predicted that all of them had at least one binding site for SR proteins (Fig. 6a). SRp40, SRp30a and SC35 proteins had the highest number of binding sites on the 13 exons,

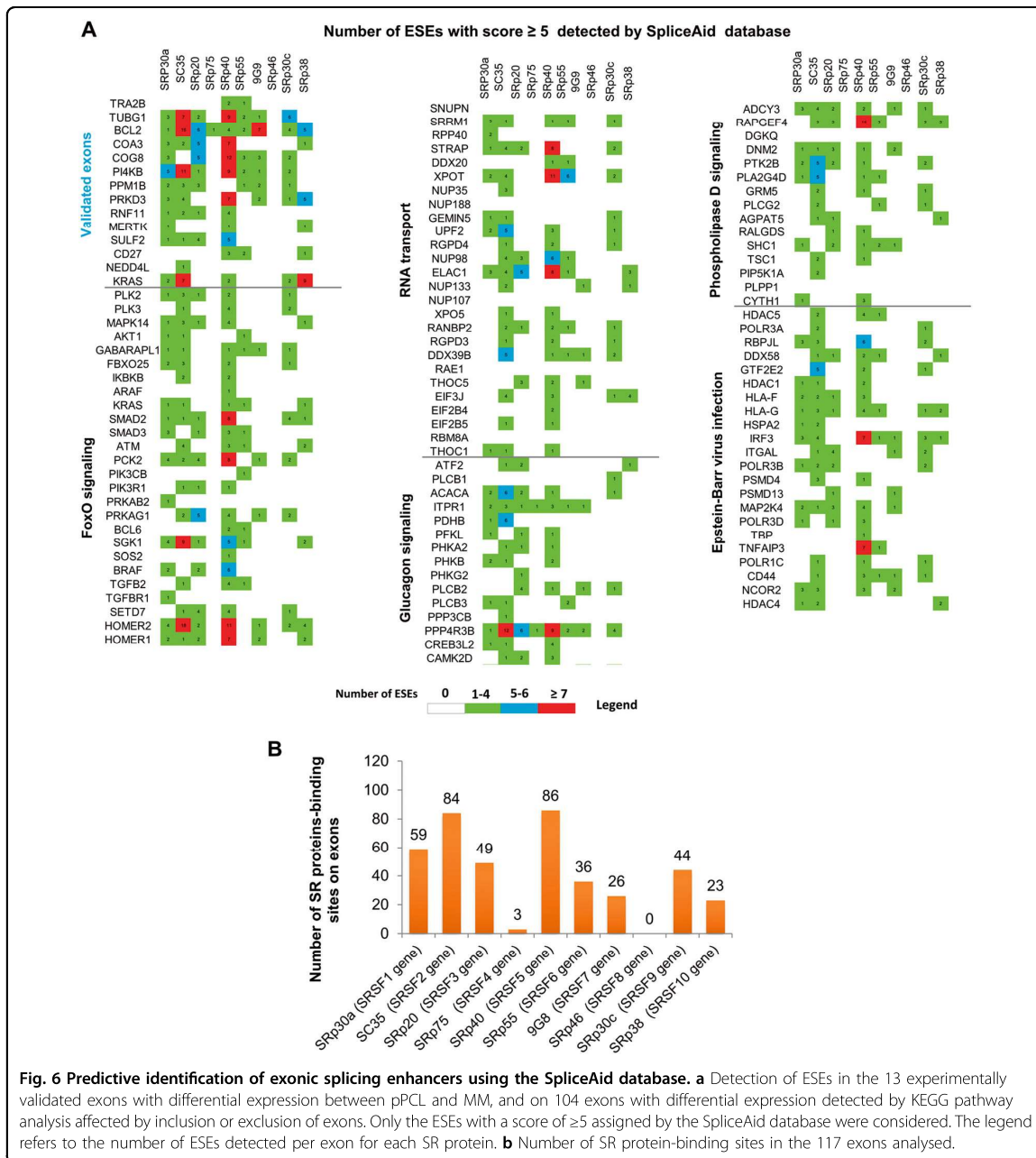
while SRp75 and SRp46 SR proteins had one and zero, respectively (Fig. 6a). These results indicate a possible strong effect of SRp40, SRp30a and SC35 proteins of all the SR proteins involved in the splicing of these pre-mRNAs (with a greater number of binding sites). In fact, SR proteins were overexpressed at the gene level in pPCLs compared with MM patients, although the differences were only statistically significant for *SRSF1* and *SFSF3* (Supplementary Fig. 8).

Secondly, we analysed 104 exons differentially expressed between pPCL and MM that were detected in KEGG pathway analysis of genes affected by inclusion or exclusion of exons (Supplementary Table 6). More than 70% of exons had binding sites for SRp40 and SC35 proteins, followed by SRp30a and SRp20, with 50% and 42%, respectively (Fig. 6b).

Since ESEs can be highly heterogeneous, their composition probably influences the occurrence of pre-mRNA splicing and thereby affects the transcriptome of cells. The higher number of ESEs found using the database indicates that prediction analysis needs to be validated experimentally to identify those that genuinely act by binding to SR proteins in the specific context of pPCL and MM.

#### Discussion

In this study, we compared the transcriptome profiles of pPCL and MM using samples that shared 17p deletion and a similar pattern of other cytogenetic alterations. These selection criteria allowed us to investigate whether molecular mechanisms other than those associated with the simple predominance of a particular chromosomal abnormality could contribute to determine the dramatic aggressive outcome of pPCL. Using HTA microarrays we were able not only to study the gene expression pattern, as other authors have done, but also to investigate the isoform expression and the impact of alternative splicing on the pathology of these diseases.



In spite of the similar distribution of cytogenetic alterations in the two groups, the unsupervised analysis based on gene expression data distinguished two differentiated clusters. In contrast, the same kind of analysis using RNA-Seq data from the CoMMpass study did not identify any cluster in the set of MM patients with the 17p deletion. These results suggest that tumour PCs of pPCL

have a significantly different transcriptome profile from that observed in the tumour PCs of MM, even though both diseases share the same tumour cells and have similar cytogenetic abnormalities. Indeed, we found a large number of genes that are differentially expressed between pPCL and MM, most of which are under-expressed in pPCL patients. Our results are not

comparable with those of previous studies of gene expression profiling in pPCL and MM patients, since we selected patients with a similar genetic background across the two plasma cell dyscrasias<sup>17</sup>. This probably explains why the gene expression-based pathway enrichment analysis showed seven affected pathways, including the spliceosome, aminoacyl-tRNA biosynthesis and EBV infection, which have not been described before. Nevertheless, *SLAMF6*, *SULF2*, *STAP1*, and *AHR* genes, four of the 40 most significantly deregulated genes in pPCL were among the list of 503 genes reported by Todoerti et al.<sup>17</sup> as being differentially expressed between the two groups.

Interestingly, six of the 20 most underexpressed genes founded in pPCL (*FRZB*, *DKK1*, *KIT*, *NCAM1*, *CTGF*, and *CXCR4*) are commonly associated with the bone marrow microenvironment and bone disease in MM. For example, the expression of the WNT inhibitors, *FRZB* and *DKK1* genes, is associated with osteolytic bone lesions in myeloma patients<sup>35</sup>. In fact, ~80% of MM patients develop osteolytic bone lesions, while in pPCL bone lesions are less frequent<sup>36</sup>.

*WARS* (tryptophanyl-tRNA synthetase) and *AHR* (aryl hydrocarbon receptor) genes were the two most overexpressed genes in pPCL. *WARS* is a housekeeping enzyme that participates in the protein synthesis, catalysing the specific ligation of tRNA with tryptophan<sup>37</sup>. The upregulation of *WARS* has been reported to be included in a high-risk transcriptional signature for pPCL correlated with shorter overall survival<sup>17</sup>. Remarkably, a recent study has demonstrated a novel non-canonical function of *WARS* in antiviral defence, so it is rapidly secreted in response to viral infection, leading to the *in vitro* and *in vivo* inhibition of virus replication<sup>38</sup>. Previous studies have reported high levels of expression of *WARS* in cells infected with human cytomegalovirus, hepatitis B virus, and even in mouse intestines infected with *Cholera vibrio*<sup>39–41</sup>. In our study, an unexpected finding was the deregulation of the EBV infection pathway as revealed by the KEGG analysis for gene, isoform and alternatively spliced exon expressions. In addition, a few studies have reported the use of cellular spliceosome machinery by EBV for the transcription of viral proteins in EBV-infected cells with splicing components<sup>42,43</sup>. We can speculate that the infection with EBV might lead to the overexpression of the *WARS* gene as an antiviral defence of the EBV-infected cells. However, the potential role of this response in the pathogenesis of pPCL is unknown. The *AHR* gene is a cytosolic ligand-activated transcription factor considered as a chemical sensor of xenobiotic-induced carcinogenesis<sup>44</sup>. Several studies have demonstrated that *AHR* is overexpressed and constitutively active, even in the absence of environmental ligands in a range of human tumours, such as cutaneous squamous cell carcinomas<sup>45</sup>, breast cancer<sup>46</sup>, adult T-cell

leukemia<sup>47</sup>, pancreatic cancer<sup>48</sup>, and human glioma<sup>49</sup>. Recently, *AHR* levels have been inversely associated with survival of MM patients, indicating that *AHR* inhibition may be an attractive therapeutic option for the treatment of this disease<sup>50</sup>.

We also investigated alternative splicing events and isoforms expression in pPCL and MM patients. More than 2000 events of alternative splicing, predominantly inclusion or exclusion of exons, were detected. The use of a transcriptome array allowed us to identify more than 1000 deregulated coding isoforms, mostly underexpressed in pPCL, belonging to 658 genes whose expression did not significantly differ between pPCL and MM. Of these, the *SF1* (Splicing Factor 1) gene is a RNA-binding component of the spliceosomal machinery required for the ATP-dependent first step of spliceosome assembly. The deregulated *SF1* isoform is shorter than the canonical one and lacks the essential protein domain expression involved in their RNA binding functions. In this way, the *SF1* isoform is probably unable of efficiently exert its canonical function in spliceosome assembly, irrespective of *SF1* gene expression. We also found two deregulated *ASS1* (Argininosuccinate Synthase 1) isoforms that differs in the number of protein domains and features. The *Ass1* protein is one of the enzymes of the urea cycle and catalyses the penultimate step of the arginine biosynthesis pathway. Thereby, *ASS1* gene is critical for the growth of human cancers, due to the role of L-arginine in different aspects of tumour metabolism and the immune system<sup>51,52</sup>. We can reasonably speculate that the overexpression of these two *ASS1* isoforms in pPCL patients could be associated with an increased proliferation of the PCs.

On the other hand, we detected six underexpressed *IKZF1* isoforms in pPCL patients in the HTA data, two of which were validated by qRT-PCR. Ikaros, encoded by the *IKZF1* gene, is a zinc-finger protein essential for lymphocyte development and which is important in MM pathology<sup>53</sup>. Their proteasomal degradation is a critical event in the mechanism of action of immunomodulatory drugs (IMiDs) against MM<sup>54</sup>. The 23 *IKZF1* isoforms produced by alternative splicing are differentiated by the alternative use of exons and subsequently in their functional domain composition, with distinct biological functions. However, only the isoforms that contain at least three of four N-terminal zinc fingers are capable of binding to DNA efficiently, and exert their function as a transcription factor. Nevertheless, the isoforms that contain the C-terminal zinc fingers that are responsible for dimerisation with other Ikaros isoforms and other family members have dominant negative functions by which they interfere with the DNA-binding functions of those proteins, including the inhibition of the activity of the full-length Ikaros protein<sup>55–57</sup>. Interestingly, in this study the

two validated *IKZF1* isoforms are shorter than the canonical isoform and are non-DNA-binding isoforms with a distinct C-terminal compared with the canonical isoform. Likewise, we hypothesised that these two *IKZF1* non-DNA-binding isoforms could also bind to the other members of the zinc-finger protein family through the C-terminal domain and exert a dominant-negative effect by interfering with the activity of those DNA-binding isoforms and, consequently, transcription would not be activated<sup>56,58</sup>.

Overall, our results indicate that pPCL is a unique biological entity, and therefore different from MM. Even comparing MM and pPCL cases that share similar cytogenetic alterations, the transcriptome of pPCL is substantially and significantly different from that of MM. This work is the first of its kind to explore the expression of the spliceosomal components and the occurrence of alternative splicing events in pPCL compared with MM samples. Our findings also suggest the existence of a significant connection between alternative splicing deregulation and the molecular features of pPCL, and provide evidence of candidate genes and as yet unidentified pathways that could be pertinent to the biology of pPCL and that could contribute to the more aggressive clinical course of this neoplasm compared with that of MM. A limitation of this study is the small sample size related to the low incidence of pPCL. As such, further research is needed to validate the most relevant findings in other series of pPCL patients with del17p.

#### Author details

<sup>1</sup>Cancer Research Center-IBMCC (USAL-CSIC), Salamanca, Spain. <sup>2</sup>Institute of Biomedical Research of Salamanca (IBSAL), Salamanca, Spain. <sup>3</sup>Hematology Department, University Hospital of Salamanca, Salamanca, Spain. <sup>4</sup>Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), CB16/12/00233, Salamanca, Spain. <sup>5</sup>National Medicines Institute, Warsaw, Poland

#### Competing interests

Dr. Mateos reports personal fees from Janssen, Celgene, Amgen, Takeda, Pharmamar, GSK, AbbVie and Adaptive, outside the submitted work. Dr. García-Sanz reports personal fees from Amgen, Janssen, Takeda and Pfizer, outside the submitted work. Dr. Gutiérrez reports honoraria from Janssen outside the submitted work. The other authors declare no conflict of interest.

#### Funding

This study was partially supported by the Instituto de Salud Carlos III cofinanced by the European Union FEDER funds (PI13/00111 and PI16/01074), Gerencia Regional de Salud, Junta de Castilla y León (BIO/SA57/13) and the Asociación Española Contra el Cáncer (AECC, GCB120981SAN). E.A.R.R. was supported by the Consejería de Educación de Castilla y León and FEDER funds. I.M.K. and L.A.C. were supported by the Sociedad Española de Hematología y Hemoterapia.

#### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Supplementary Information** accompanies this paper at (<https://doi.org/10.1038/s41408-019-0253-1>).

Received: 4 June 2019 Revised: 17 October 2019 Accepted: 29 October 2019

Published online: 20 November 2019

#### References

- Albarracín, F., Fonseca, R. Plasma cell leukemia. *Blood Rev.* **25**, 107–112 (2011).
- Fernández de Larrea, C. et al. Plasma cell leukemia: consensus statement on diagnostic requirements, response criteria and treatment recommendations by the International Myeloma Working Group. *Leukemia* **27**, 780–791 (2013).
- Kyle, R. A., Maldonado, J. E., Bayrd, E. D. Plasma cell leukemia. Report on 17 cases. *Arch. Intern. Med.* **133**, 813–818 (1974).
- Noel, P., Kyle, R. A. Plasma cell leukemia: an evaluation of response to therapy. *Am. J. Med.* **83**, 1062–1068 (1987).
- International Myeloma Working Group. Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. *Br. J. Haematol.* **121**, 749–757 (2003).
- Gonsalves, W. I. et al. Trends in survival of patients with primary plasma cell leukemia: a population-based analysis. *Blood* **124**, 907–912 (2014).
- Nerj, A. et al. Primary plasma cell leukemia 2.0: advances in biology and clinical management. *Expert Rev. Hematol.* **9**, 1063–1073 (2016).
- Mina, R., D'Agostino, M., Cerrato, C., Gay, F., Palumbo, A. Plasma cell leukemia: update on biology and therapy. *Leuk. Lymphoma* **58**, 1538–1547 (2017).
- Ravi, P. et al. Revised diagnostic criteria for plasma cell leukemia: results of a Mayo Clinic study with comparison of outcomes to multiple myeloma. *Blood Cancer J.* **8**, 116 (2018).
- Simeon, V. et al. Molecular classification and pharmacogenetics of primary plasma cell leukemia: an initial approach toward precision medicine. *Int. J. Mol. Sci.* **16**, 17514–17534 (2015).
- Tiedemann, R. E. et al. Genetic aberrations and survival in plasma cell leukemia. *Leukemia* **22**, 1044–1052 (2008).
- Gutiérrez, N. C. et al. Differences in genetic changes between multiple myeloma and plasma cell leukemia demonstrated by comparative genomic hybridization. *Leukemia* **15**, 840–845 (2001).
- Avet-Loiseau, H. et al. Cytogenetic, interphase, and multicolor fluorescence in situ hybridization analyses in primary plasma cell leukemia: a study of 40 patients at diagnosis, on behalf of the Intergroupe Francophone du Myélome and the Groupe Français de Cytogénétique Hématologique. *Blood* **97**, 822–825 (2001).
- Mosca, L. et al. Genome-wide analysis of primary plasma cell leukemia identifies recurrent imbalances associated with changes in transcriptional profiles. *Am. J. Hematol.* **88**, 16–23 (2013).
- Cifola, I. et al. Whole-exome sequencing of primary plasma cell leukemia discloses heterogeneous mutational patterns. *Oncotarget* **6**, 17543–17558 (2015).
- Usmani, S. Z. et al. Primary plasma cell leukemia: clinical and laboratory presentation, gene-expression profiling and clinical outcome with Total Therapy protocols. *Leukemia* **26**, 2398–2405 (2012).
- Todoerti, K. et al. Transcriptional characterization of a prospective series of primary plasma cell leukemia revealed signatures associated with tumor progression and poorer outcome. *Clin. Cancer Res.* **19**, 3247–3258 (2013).
- Lionetti, M. et al. Biological and clinical relevance of miRNA expression signatures in primary plasma cell leukemia. *Clin. Cancer Res.* **19**, 3130–3142 (2013).
- Todoerti, K. et al. Global methylation patterns in primary plasma cell leukemia. *Leuk. Res.* **73**, 95–102 (2018).
- Mattioli, M. et al. Gene expression profiling of plasma cell dyscrasias reveals molecular patterns associated with distinct IGH translocations in multiple myeloma. *Oncogene* **24**, 2461–2473 (2005).
- López-Corral, L. et al. The progression from MGUS to smoldering myeloma and eventually to multiple myeloma involves a clonal expansion of genetically abnormal plasma cells. *Clin. Cancer Res.* **17**, 1692–1700 (2011).
- Gonzalez, D. et al. Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. *J. Clin. Oncol.* **29**, 2223–2229 (2011).
- Wang, J., Vasaiak, S., Shi, Z., Greer, M., Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).
- Piva, F., Giulietti, M., Nocchi, L., Principato, G. SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. *Bioinf. Oxf. Engl.* **25**, 1211–1213 (2009).

25. Anczuków, O. et al. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol. Biol.* **19**, 220–228 (2012).
26. Anczuków, O. et al. SRSF1-regulated alternative splicing in breast cancer. *Mol. Cell* **60**, 105–117 (2015).
27. DeLigio, J. T. et al. Serine/Arginine-Rich Splicing Factor 3 modulates the alternative splicing of cytoplasmic polyadenylation element binding protein 2. *Mol. Cancer Res.* **17**, 1920–1930 (2019).
28. Das, S., Krainer, A. R. Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Mol. Cancer Res.* **12**, 1195–1204 (2014).
29. Karni, R. et al. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* **14**, 185–193 (2007).
30. Ajiro, M., Jia, R., Yang, Y., Zhu, J., Zheng, Z.-M. A genome landscape of SRSF3-regulated splicing events and gene expression in human osteosarcoma U2OS cells. *Nucleic Acids Res.* **44**, 1854–1870 (2016).
31. Zhang, Z., Pal, S., Bi, Y., Tchou, J., Davuluri, R. V. Isoform level expression profiles provide better cancer signatures than gene level expression profiles. *Genome Med.* **5**, 33 (2013).
32. Pal, S. et al. Isoform-level gene signature improves prognostic stratification and accurately classifies glioblastoma subtypes. *Nucleic Acids Res.* **42**, e64 (2014).
33. Blencowe, B. J. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* **25**, 106–110 (2000).
34. Wang, J., Smith, P. J., Krainer, A. R., Zhang, M. Q. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic Acids Res.* **33**, 5053–5062 (2005).
35. van Andel H., Kocemba K. A., Spaargaren M., Pals S. T. Aberrant Wnt signaling in multiple myeloma: molecular mechanisms and targeting options. *Leukemia* (2019). <https://doi.org/10.1038/s41375-019-0404-1>.
36. van de Donk, N. W. C. J., Lokhorst, H. M., Anderson, K. C. & Richardson, P. G. How I treat plasma cell leukemia. *Blood* **120**, 2376–2389 (2012).
37. Jin, M. Unique roles of tryptophanyl-tRNA synthetase in immune control and its therapeutic implications. *Exp. Mol. Med.* **51**, 1 (2019).
38. Lee H.-C. et al. Released tryptophanyl-tRNA synthetase stimulates innate immune responses against viral infection. *J. Virol.* **93**, (2019). <https://doi.org/10.1128/JVI.01291-18>.
39. Ellis, C. N. et al. Comparative proteomic analysis reveals activation of mucosal innate immune signaling pathways during cholera. *Infect. Immun.* **83**, 1089–1103 (2015).
40. Wieland, S., Thimme, R., Purcell, R. H., Chisari, F. V. Genomic analysis of the host response to hepatitis B virus infection. *Proc. Natl Acad. Sci. USA* **101**, 6669–6674 (2004).
41. Zhu, H., Cong, J. P., Mamtora, G., Gingeras, T., Shenk, T. Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* **95**, 14470–14475 (1998).
42. Lee, N., Yario, T. A., Gao, J. S., Steitz, J. A. EBV noncoding RNA EBER2 interacts with host RNA-binding proteins to regulate viral gene expression. *Proc. Natl Acad. Sci. USA* **113**, 3221–3226 (2016).
43. Nanni, A. V., Lee, N. Identification of host RNAs that interact with EBV non-coding RNA EBER2. *RNA Biol.* **15**, 1181–1191 (2018).
44. Xue, P., Fu, J., Zhou, Y. The aryl hydrocarbon receptor and tumor immunity. *Front. Immunol.* **9**, 286 (2018).
45. Pan, Z.-Y. et al. Activation and overexpression of the aryl hydrocarbon receptor contribute to cutaneous squamous cell carcinomas: an immunohistochemical study. *Diagn. Pathol.* **13**, 59 (2018).
46. D'Amato, N. C. et al. A TDO2-AhR signaling axis facilitates anoikis resistance and metastasis in triple-negative breast cancer. *Cancer Res.* **75**, 4651–4664 (2015).
47. Hayashibara, T. et al. Possible involvement of aryl hydrocarbon receptor (AhR) in adult T-cell leukemia (ATL) leukemogenesis: constitutive activation of AhR in ATL. *Biochem. Biophys. Res. Commun.* **300**, 128–134 (2003).
48. Koliopoulos, A. et al. Increased arylhydrocarbon receptor expression offers a potential therapeutic target for pancreatic cancer. *Oncogene* **21**, 6059–6070 (2002).
49. Gramatzki, D. et al. Aryl hydrocarbon receptor inhibition downregulates the TGF-beta/Smad pathway in human glioblastoma cells. *Oncogene* **28**, 2593–2605 (2009).
50. Bianchi-Smiraglia, A. et al. Inhibition of the aryl hydrocarbon receptor/polyamine biosynthesis axis suppresses multiple myeloma. *J. Clin. Invest.* **128**, 4682–4696 (2018).
51. Szeffel, J., Danielak, A., Kruszewski, W. J. Metabolic pathways of L-arginine and therapeutic consequences in tumors. *Adv. Med. Sci.* **64**, 104–110 (2018).
52. Rodriguez, P. C., Quiceno, D. G., Ochoa, A. C. L-arginine availability regulates T-lymphocyte cell-cycle progression. *Blood* **109**, 1568–1573 (2007).
53. Krönke, J. et al. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305 (2014).
54. Zhu, Y. X. et al. Identification of cereblon-binding proteins and relationship with response and survival after IMiDs in multiple myeloma. *Blood* **124**, 536–545 (2014).
55. Hahm, K. et al. The lymphoid transcription factor LyF-1 is encoded by specific, alternatively spliced mRNAs derived from the Ikaros gene. *Mol. Cell Biol.* **14**, 7111–7123 (1994).
56. Sun, L., Liu, A., Georgopoulos, K. Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. *EMBO J.* **15**, 5358–5369 (1996).
57. Fedele, P. L. et al. IMiDs prime myeloma cells for daratumumab-mediated cytotoxicity through loss of Ikaros and Aiolos. *Blood* **132**, 2166–2178 (2018).
58. Sun, L. et al. Expression of dominant-negative and mutant isoforms of the antileukemic transcription factor Ikaros in infant acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **96**, 680–685 (1999).