

# Single-cell transcriptional profile of CD34<sup>+</sup> hematopoietic progenitor cells from del(5q) myelodysplastic syndromes and impact of lenalidomide

Received: 1 November 2023

Accepted: 6 June 2024

Published online: 20 June 2024

 Check for updates

A list of authors and their affiliations appears at the end of the paper

While myelodysplastic syndromes with del(5q) (del(5q) MDS) comprises a well-defined hematological subgroup, the molecular basis underlying its origin remains unknown. Using single cell RNA-seq (scRNA-seq) on CD34<sup>+</sup> progenitors from del(5q) MDS patients, we have identified cells harboring the deletion, characterizing the transcriptional impact of this genetic insult on disease pathogenesis and treatment response. Interestingly, both del(5q) and non-del(5q) cells present similar transcriptional lesions, indicating that all cells, and not only those harboring the deletion, may contribute to aberrant hematopoietic differentiation. However, gene regulatory network (GRN) analyses reveal a group of regulons showing aberrant activity that could trigger altered hematopoiesis exclusively in del(5q) cells, pointing to a more prominent role of these cells in disease phenotype. In del(5q) MDS patients achieving hematological response upon lenalidomide treatment, the drug reverts several transcriptional alterations in both del(5q) and non-del(5q) cells, but other lesions remain, which may be responsible for potential future relapses. Moreover, lack of hematological response is associated with the inability of lenalidomide to reverse transcriptional alterations. Collectively, this study reveals transcriptional alterations that could contribute to the pathogenesis and treatment response of del(5q) MDS.

Deletion of the long arm of chromosome 5 (del(5q)) is the most frequently observed cytogenetic alteration in de novo myelodysplastic syndromes (MDS). It affects around 10-15% of MDS patients and represents a distinct hematological and pathological subgroup due to its unique clinical features, such as macrocytosis, anemia, normal or high platelet count, and hypolobulated megakaryocytes<sup>1</sup>. With the advent of high-throughput sequencing technologies, two commonly deleted regions (CDR) were identified and mapped to del(5q): the 1.5 megabase region at 5q32-q33, which contains 41 coding genes<sup>2,3</sup>, and the 5q31 region containing 45 coding genes. Thus, the characterization of the molecular pathogenesis of del(5q) MDS has mainly focused on

the identification of genes within these CDRs that are involved in the pathophysiology of the disease, such as *RPS14*, *miR-145*, *miR-146* and *EGRI*<sup>4-7</sup>. However, the presence of gene regulatory networks (GRNs) altered in these patients, as well as the transcriptional profile of hematopoietic progenitors from patients with del(5q), have not been analyzed in detail, in part due to the difficulty of identifying cells with and without the deletion within the same patient.

Lenalidomide represents the first therapeutic approach for del(5q) MDS, and induces in these patients a prolonged red blood cell transfusion independence and cytogenetic response<sup>8</sup>. Interestingly, even at the time of complete clinical and cytogenetic remission,

✉ e-mail: [fprosper@unav.es](mailto:fprosper@unav.es); [tezponda@unav.es](mailto:tezponda@unav.es); [mhernaes@unav.es](mailto:mhernaes@unav.es)

persistent del(5q) progenitor cells have been identified, providing a reasonable explanation for the loss of responsiveness that patients experience in a 2- to 3-year interval<sup>9</sup>. However, the real impact of the deletion during lenalidomide treatment remains unknown, hinting the need of characterizing the molecular mechanisms driven by the deletion that could be impairing treatment response.

Previous studies have demonstrated that transcriptional alterations play a key role in MDS pathogenesis<sup>10–14</sup>. For example, our group recently identified the transcription factor *DDIT3* as a key erythropoietic regulator that is overexpressed in MDS, and showed its potential as a therapeutic target<sup>15</sup>. Furthermore, gene expression profiles have shown to be affected by different types of alterations, including cytogenetic abnormalities and mutations, among other factors<sup>16,17</sup>. In this sense, several studies have performed gene expression analyses in del(5q) MDS, and have identified several pathways deregulated in the disease, including those related to Wnt/ $\beta$  catenin or integrin signaling, as well as genes that could be potentially contributing to disease pathogenesis, such as *SPARC*<sup>14,18</sup>. Other studies have depicted the effect of lenalidomide on the gene expression profile of del(5q) patients, evidencing its potential to restore erythroid differentiation, modulate the bone marrow microenvironment, and revert the aberrant expression of putative pathogenic microRNAs<sup>19–22</sup>. However, the hematopoietic system of these patients is composed of a mixture of cells with and without the deletion (termed del(5q) and non-del(5q) cells, respectively), which masks the expression profile of the del(5q) cells, and limits the ability to define the transcriptional impact of the deletion.

In this work, we performed single-cell RNA-sequencing (scRNA-seq) in primary CD34<sup>+</sup> hematopoietic progenitor cells from del(5q) MDS patients at diagnosis and patients after lenalidomide treatment, and applied copy number alteration (CNA) analyses to link the del(5q) genotype to the transcriptional profile of each individual cell. This approach yielded a well-characterized del(5q) MDS atlas. Leveraging the generated atlas, we detected transcriptional alterations both in del(5q) cells and also in non-del(5q) cells at diagnosis and after lenalidomide treatment. We demonstrate that non-del(5q) cells present aberrant behavior compared to the healthy hematopoietic system similar to that observed in del(5q) cells. We also show that although lenalidomide restores some of the detected alterations of del(5q) and non-del(5q) cells of responder patients, other lesions identified at diagnosis remain after treatment. Furthermore, our results evidence that lenalidomide is not able to reverse part of the transcriptional lesions carried by del(5q) cells of a non-responder patient, which seems to be associated with the lack of hematological response.

## Results

### Single-cell RNA-sequencing of hematopoietic progenitor cells of del(5q) MDS patients

To identify the transcriptional alterations characterizing hematopoietic progenitors harboring del(5q), we initially performed scRNA-seq of CD34<sup>+</sup> cells of four newly diagnosed patients with del(5q) MDS (Patient\_1-4), and three age-matched healthy donors (Healthy\_1-3) using the 10X Genomics technology (Fig. 1A) (gating strategy can be found in Supplementary Fig. 1). The clinical and genomic characteristics of the MDS patients and healthy donors are shown in Supplementary Table 1. The percentage of cells with del(5q) based on the cytogenetic analysis varied between 35 to 90%. In all cases, the common deleted region encompassed bands 5q(13-33) (genes shown in Supplementary Data 1).

A total of 55,119 and 45,311 cells from patients and healthy donors, respectively, were profiled and integrated. After applying quality filters, 46,772 and 43,442 cells were eventually included in the downstream analysis. Data was integrated, clustered and manually annotated (Fig. 1B, C, Supplementary Fig. 2A, B) based on curated markers (Fig. 1D, Supplementary Fig. 2C), obtaining 14 and 13 clusters (patients and donors, respectively) representing all the expected

hematopoietic progenitor subtypes. Contribution of every MDS patient and donor to the composition of all the clusters was identified (Fig. 1E, Supplementary Fig. 2D), and each individual showed different proportions of hematopoietic progenitors (Fig. 1F, Supplementary Fig. 2E). Although there were some differences in the percentage of hematopoietic progenitors between MDS patients and healthy donors (e.g., HSC), these differences were not statistically significant which might be related to the high variability in cell composition across samples (Fig. 1G).

### Identification of CD34<sup>+</sup> cells harboring del(5q) in MDS patients

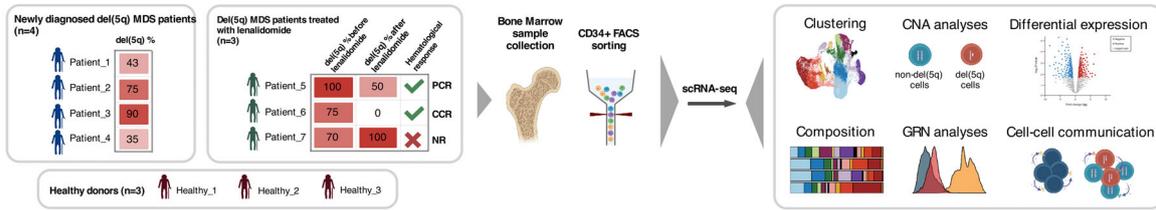
Identifying single-arm copy number variations (CNVs) at the single-cell level presents challenges due to potential compensatory mechanisms of alleles, as well as to the sparse and noisy nature of single-cell data. In this study, we employed two different and complementary approaches: CopyKat<sup>23</sup> (Fig. 2A), which relies on gene expression, and CaSpER<sup>24</sup> (Fig. 2B), which relies on allele frequencies (see Methods). This combined strategy aimed to enhance the sensitivity and accuracy of identifying cells harboring 5q deletion. To avoid false positive detection, we only classified the cells as harboring the del(5q) if the same cell was characterized as such by the two different algorithms (Fig. 2C). To validate this classification, we analyzed the expression pattern of genes encoded in the deleted region in individual cells. Due to the sparsity of scRNA-seq data, we were only able to detect six genes as highly variable, *CD74*, *RPS14*, *BTF3*, *COX7C*, *HINT1* and *RPS23*, whose expression was decreased in del(5q) when compared to non-del(5q) cells at sample level (Fig. 2D), further confirming our del(5q) cell classification. Once the classification was performed, we applied a Wilcoxon signed-rank test between cells classified as del(5q) and non-del(5q), revealing in the underexpressed fraction of the genes an enrichment for the genes located on the deleted locus (Supplementary Fig. 3A, B). To further validate the classification, we randomly shuffled the labels from the classified cells, and repeated the same differential expression analysis, revealing how the genes located on the deleted region started to fade away (Supplementary Fig. 3C). Based on this classification, interestingly, for each individual patient, the proportion of del(5q) in the CD34<sup>+</sup> progenitor cells was consistent with that obtained by karyotype in total bone marrow (Fig. 2E).

We then interrogated the distribution of del(5q) cells across the different hematopoietic progenitors. Cells with the deletion were detected in all the defined hematopoietic progenitor clusters (Fig. 3A), although a high heterogeneity of distribution was observed among patients (Fig. 3B, C). Despite the observed heterogeneity, a statistically significant accumulation of del(5q) cells was detected in early erythroid progenitors across all individuals (hypergeometric test,  $p$ -value < 0.05). Additionally, three out of four patients exhibited statistically significant enrichment of del(5q) in granulocyte-monocyte progenitors (GMP), megakaryocyte and late erythroid progenitors (Fig. 3D). Collectively, these results indicated a bias of del(5q) cells towards specific myeloid compartments, mainly towards erythroid cells, which is consistent with the association between this genetic lesion and the anemia that characterizes patients with del(5q) MDS.

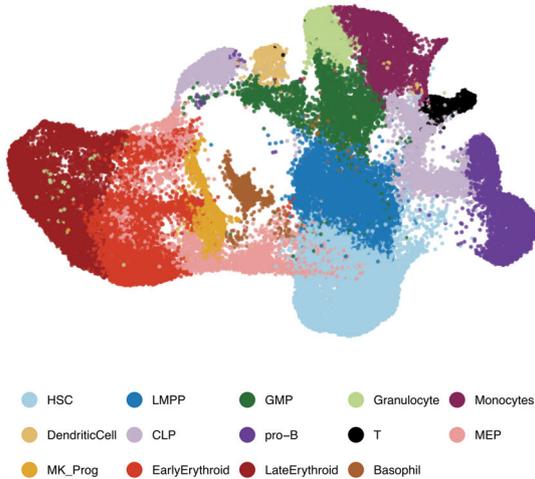
### Transcriptional differences between del(5q) and non-del(5q) cells within MDS patients are driven by few specific transcriptional programs playing key roles in MDS

To delve into the transcriptional program associated with del(5q) cells in patients with MDS, we performed a pseudobulk differential expression (DE) analysis between del(5q) and non-del(5q) cells for each cell population, as traditional Wilcoxon signed-rank test-based DE analysis in single cell data has recently shown to yield high false positive rates<sup>25</sup>. Intriguingly, considering every type of hematopoietic progenitors, only seven genes were differentially expressed (downregulated) in del(5q) in comparison with non-del(5q) cells (Fig. 4A). Some of the downregulated genes played a key role in MDS and other

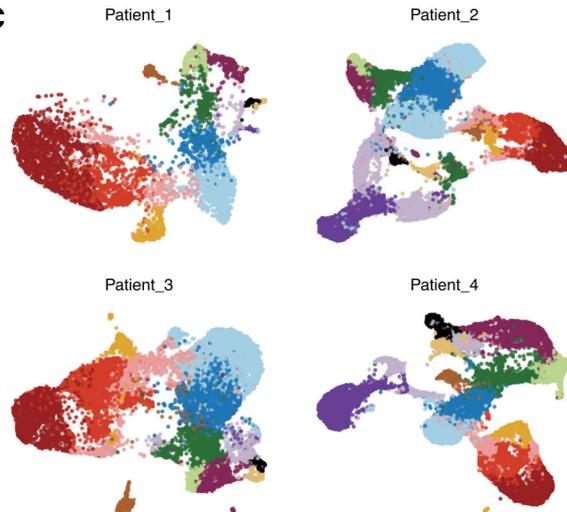
**A**



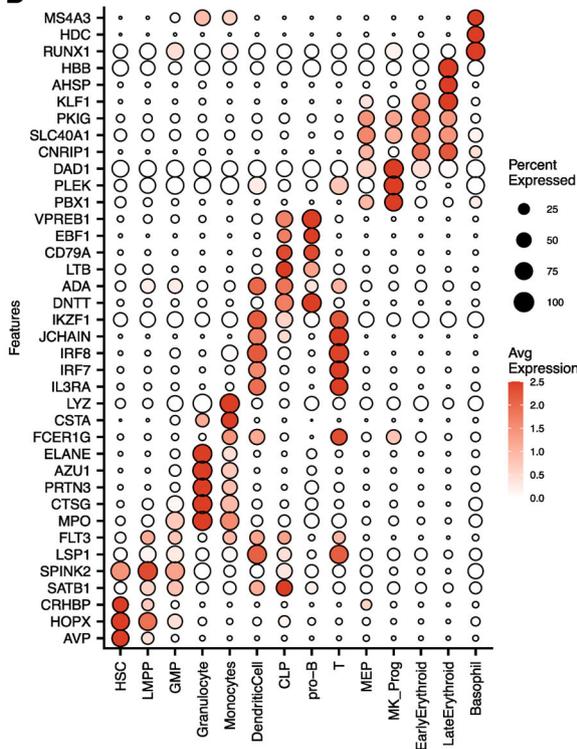
**B**



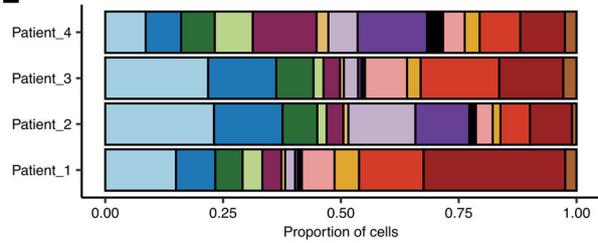
**C**



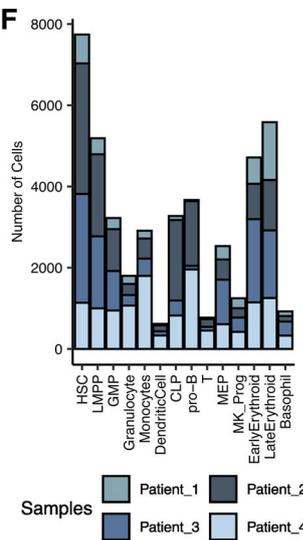
**D**



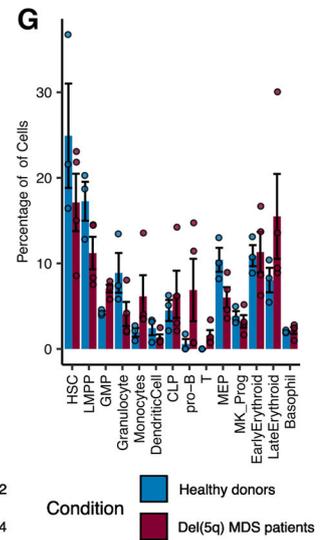
**E**



**F**



**G**



non-hematological tumors, such as *PRSS21*, which encodes for a tumor suppressor frequently hypermethylated in cancer<sup>26</sup>, *MAP3K7CL*, whose downregulation serves as a biomarker in other types of cancer<sup>27</sup>, and *CCL5*, whose downregulation is associated with high-risk MDS<sup>28</sup>.

Due to the unexpected transcriptional similarity between del(5q) and non-del(5q) cells within MDS patients, we next performed a DE

analysis between del(5q) MDS cells and CD34<sup>+</sup> cells from healthy donors. This comparison yielded 20 to 988 differentially expressed genes (FDR < 0.05 and |logFC| > 2), depending on the progenitor cell (Supplementary Fig. 4A, Supplementary Data 2). Although most of these genes were cell-type-exclusive, they were enriched in similar pathways in most of the cell types (Fig. 4B, left panel). Genes

**Fig. 1 | Hematopoietic CD34<sup>+</sup> cells from four independent del(5q) MDS patients were assayed by scRNAseq.** **A** CD34<sup>+</sup> cells were obtained from bone marrow aspirates of newly diagnosed del(5q) MDS patients, healthy donors and patients treated with Lenalidomide, and were subjected to single-cell RNA sequencing and analysis. PCR partial cytogenetic responder, CCR complete cytogenetic responder, NR non-responder. Part of this figure was created with BioRender.com. **B** Uniform Manifold Approximation and Projection (UMAP) of 42,494 cells representing the expected 14 hematopoietic progenitors: HSC hematopoietic stem cells, LMPP lymphoid-primed multipotent progenitors, GMP granulocyte-monocyte progenitors; granulocyte progenitors; monocyte progenitors; dendritic cell progenitors; CLP common lymphoid progenitors; B-cell progenitors; T-cell progenitors; MEP megakaryocyte-erythroid progenitors; MK\_Prog megakaryocyte progenitors; early erythroid progenitors; late erythroid progenitors; basophil progenitors. **C** Per patient UMAP showing the identity of the cells projected from the integrated space.

**D** Dotplot showing the percentage and value of the normalized expression of the canonical marker genes used to assign the cell identity to each cluster. **E** Barplot representing the contribution of cells from each patient to the different cell types. **F** Barplot representing the number of cells assigned to each cell type for the studied patients. **G** Barplot representing the percentage of cells assigned to each cell type for del(5q) MDS patients and healthy samples.  $N = 7$  biologically independent samples were used ( $n = 3$  healthy donors and 4 del(5q) MDS patients). Data are presented as mean values  $\pm$  SD. Two-sided Wilcoxon signed-rank test was used to calculate statistical significance. Exact  $p$ -values for the differential abundance of each hematopoietic progenitor between the del(5q) MDS and the healthy condition were the following: HSC:  $p = 0.63$ ; LMPP:  $p = 0.23$ ; GMP:  $p = 0.06$ ; Granulocyte:  $p = 0.23$ ; Monocytes:  $p = 0.06$ ; DendriticCell:  $p = 0.63$ ; CLP:  $p = 1$ ; pro-B:  $p = 0.23$ ; T:  $p = 0.04$ ; MEP:  $p = 0.11$ ; MK\_Prog:  $p = 0.63$ ; EarlyErythroid:  $p = 1$ ; LateErythroid:  $p = 0.23$ ; Basophil:  $p = 0.4$ .

overexpressed in del(5q) cells were enriched in cell cycle and mitosis-related signatures, such as DNA replication and mitotic nuclear division, and showed increased expression of DNA repair related genes, suggesting that loss of 5q confers increased proliferative potential. Additionally, del(5q) erythroid progenitors, LMPPs, GMPs, DCs, and monocyte progenitors showed significant upregulation of genes involved in the p53 signaling and, genes involved in the apoptosis pathway were significantly upregulated in LMPP, MEP and late erythroid progenitors, but not in early erythroid progenitor cells. Our results are in line with the increased levels of apoptosis described for del(5q) patients<sup>29–31</sup>. Downregulated genes showed enrichment in ribosomes and translation related pathways in all hematopoietic progenitors, in line with previous works that have described del(5q) MDS as a ribosomopathy<sup>2,30,32</sup>. Interestingly, besides the cytoplasmic translation, we also observed altered expression of genes associated with mitochondrial translation altered in HSCs, GMPs and granulocyte progenitors. The comparison of non-del(5q) and healthy cells resulted in 64–736 altered genes per progenitor (Supplementary Fig. 4B, Supplementary Data 2). Enriched processes were also homogeneous among most hematopoietic progenitors and, as expected, were similar to the ones observed in del(5q) vs healthy comparison (Fig. 4B, right panel).

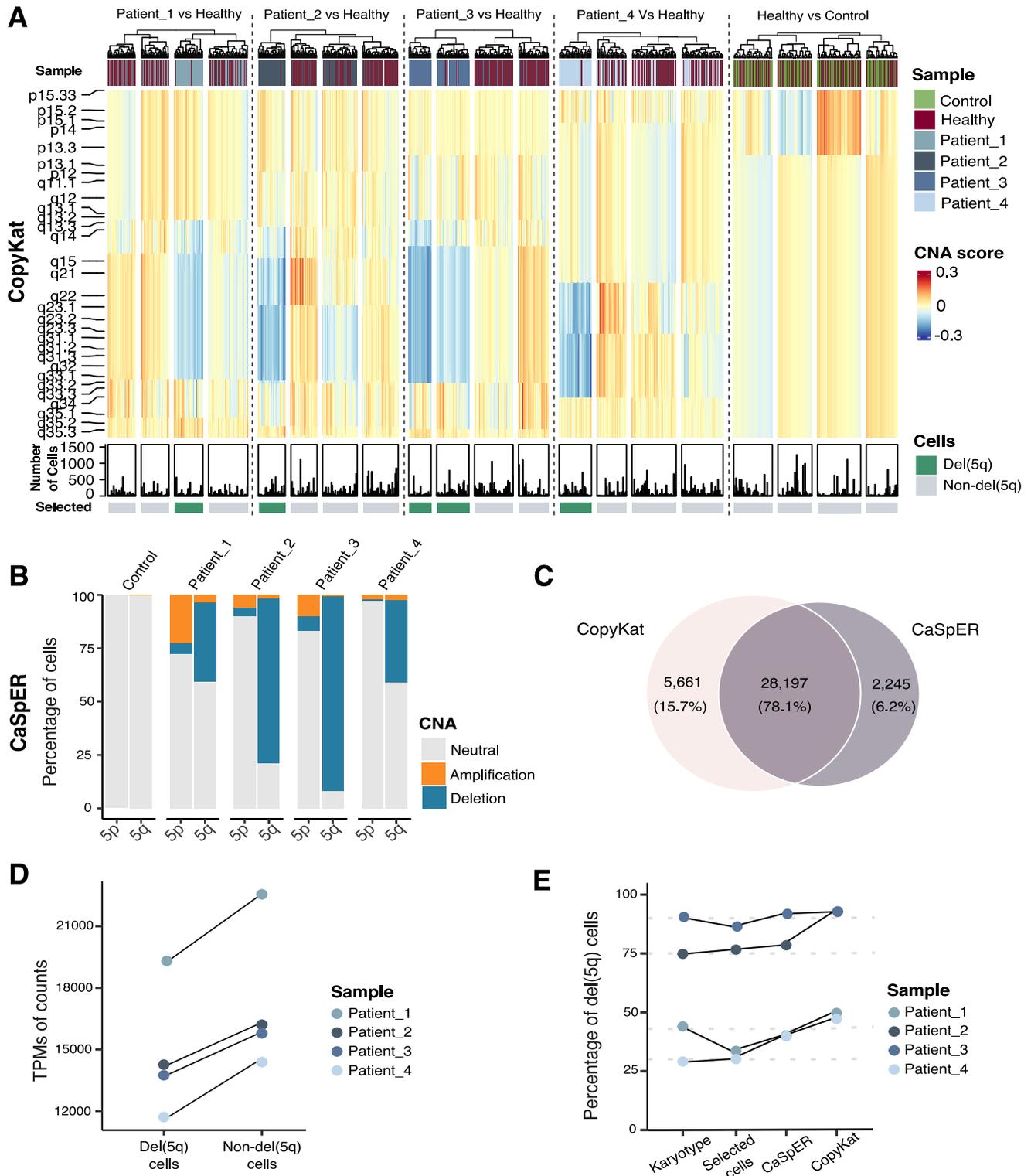
Despite the low number of DE genes between del(5q) and non-del(5q) cells, we were interested in understanding whether differences in GRN might be observed between these two populations. Unlike DE analysis, which is performed in a gene-by-gene manner, GRN studies use data-driven grouping of genes to enable the identification of mechanistic transcriptional differences between conditions. Thus, we applied SimiC<sup>33</sup> to compute the regulatory activity of regulons and observed that although some regulons behaved uniformly (low regulatory dissimilarity score, in Supplementary Fig. 5A black-purple color) between the three conditions (del(5q), non-del(5q) and healthy cells), a group of regulons showed differential activity (high regulatory dissimilarity score, in Supplementary Fig. 5A yellow-orange color) across the conditions. Among them, three different regulon activity patterns arose. Firstly, a group of regulons that showed similar activity between non-del(5q) and del(5q) cells, and different to healthy cells, in line with DE analyses, such as the ones driven by *ZNF451*, *YBX1* and *PSPCI* (Fig. 4C). Secondly, there were regulons with differential activity between the three conditions, such as those driven by *JARID2*, *IRF1* and *KAT6B*, among others (Fig. 4D). The three regulons showed high activity in cells from healthy age-matched controls (61–84 years), whereas they presented a progressively lower activity in non-del(5q) cells, and their lowest activity in del(5q) cells. *JARID2* acts as a tumor suppressor and plays a crucial role in the leukemic transformation of myeloid neoplasms<sup>34</sup>, and its deletion promotes an ineffective hematopoietic differentiation<sup>35</sup>, suggesting that the low activity of this regulon may negatively impact the hematopoietic differentiation of these patients. *IRF1* is located in 5q31.1 and its deletion in one or both alleles has been observed in MDS and AML patients with chromosome 5

abnormalities<sup>36</sup>. *IRF1* has been described as a master HSC regulator, and its loss impairs HSC self-renewal and increases stress-induced cell cycle activation, suggesting that its low activity in patients could confer proliferative advantage<sup>37</sup>. Decreased expression of *KAT6B* in aged hematopoietic stem cells has been associated with impaired myeloid differentiation<sup>38</sup>, suggesting that its almost non-existent activity in del(5q) cells may contribute to aberrant differentiation of these cells. Lastly, we detected regulons exhibiting differential activity between del(5q) and non-del(5q) cells and that showed no activity in healthy cells. In particular, regulons driven by *RERE* and *KDM2A* showed higher activity in del(5q) cells than in non-del(5q) cells (Fig. 4E). *RERE* negatively regulates the expression of target genes, and such genes are enriched in cytoplasmic translation, ribosome biogenesis and ribonucleoprotein complex biogenesis pathways, among others (Supplementary Fig. 5B). The *KDM2A* regulon was enriched in protein stabilization, regulation of cellular protein catabolic process and regulation of protein stability (Supplementary Fig. 5B). The association of *KDM2A* and ribosomal genes has been already described by previous studies, postulating that *KDM2A* overexpression reduces the transcription of rRNA<sup>39,40</sup>.

Altogether, our results suggest a low transcriptional impact of 5q loss, with del(5q) and non-del(5q) cells presenting very similar gene expression alterations when compared to healthy controls, with such alterations being involved in processes that could contribute to abnormal hematopoietic differentiation. Nevertheless, although limited in number, genes and regulons specifically altered in del(5q) cells, such as those driven by *JARID2*, *KAT6B*, *RERE* or *KDM2A*, seem to be relevant for proliferation and myeloid differentiation, supporting the concept that cells harboring the deletion may have a more prominent role in the promotion of altered hematopoiesis.

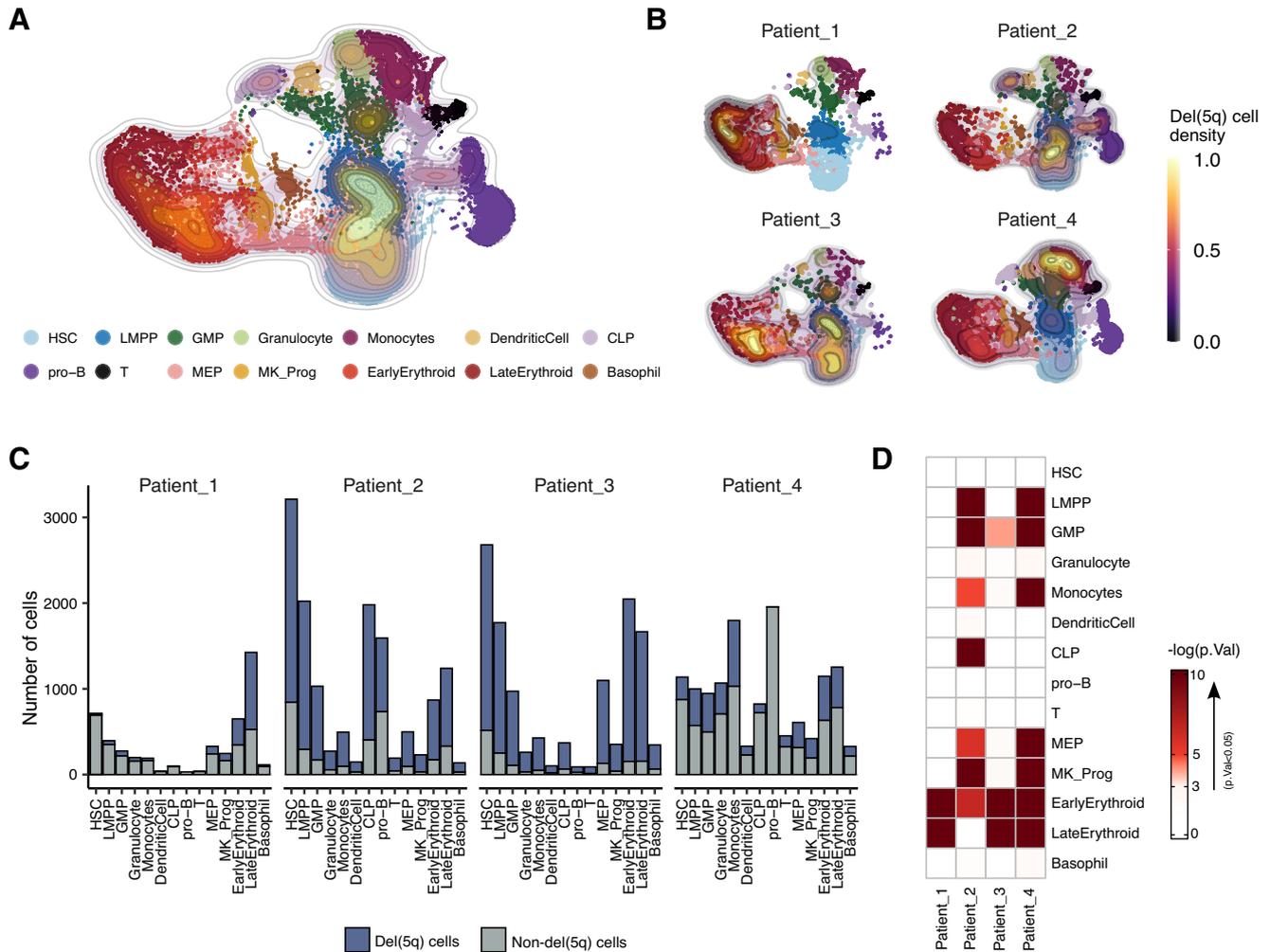
### Abnormal cell-to-cell communication in del(5q) progenitors

To investigate whether the 5q deletion has a detrimental effect on cell-cell interactions between CD34<sup>+</sup> progenitors, thus contributing to disease development, we performed a cell-to-cell communication analysis using Liana<sup>41</sup> in both del(5q) and healthy controls datasets. We identified 4,534 interactions in healthy controls, and 314 interactions that were common to all del(5q) MDS patients, most of them overlapping with those found in healthy cells (Fig. 5A). Despite this strong overlap, several differences between del(5q) MDS and healthy individuals were detected: in patients, monocyte progenitors were the most communicative cells, interacting mainly with early erythroid progenitors (Fig. 5B). However, in healthy donors, HSCs, GMPs, DC, monocyte and granulocyte progenitors were the most interactive compartments, with a notable communicative pattern between granulocyte and GMP/DC progenitors (Fig. 5C). Furthermore, genes involved in these differential interactions were overrepresented in different biological processes in each phenotype. For instance, interactions driven by healthy hematopoietic progenitors were enriched in negative regulation of apoptosis, HSC proliferation, leukocyte/DC differentiation, and



**Fig. 2 | Identification of cells harboring del(5q) deletion in MDS patients.**  
**A** Heatmap of the results of CopyKat showing the copy number alteration score given to each 200 kb bins in chromosome 5. In order to represent cells, a clustering has been performed within each sample (kmeans with  $k = 80$ ), and a posterior clustering has been applied to detect the clusters containing the cells harboring the deletion. The control sample used by the algorithm is an MDS sample with normal karyotype, while the healthy sample with normal karyotype represents an additional negative control for the analysis. **B** Barplot representing the percentage of cells inferred by CaSpER that harbor an amplification, a deletion or a normal number of copy number variation in each branch of chromosome 5 per patient. The

control corresponds to an MDS sample with normal karyotype, which is used as a reference by the algorithm. **C** Venn diagram representing the number and percentage of cells classified as del(5q) by both algorithms. **D** Pseudobulk normalized expression of the 6 CDR-genes with higher expression in our dataset (*CD74*, *RPS14*, *BTF3*, *COX7C*, *HINT1* and *RPS23*) separated by genotype.  $N = 4$  biologically independent samples were used. The number of del(5q) and non-del(5q) cells were used to generate the pseudobulks for each patient can be found in the Source Data. **E** Graph depicting the percentages of del(5q) cells inferred by karyotype, CaSpER and CopyKat for each patient. Selected cells correspond to the cells classified as del(5q) cells by both computational algorithms.



**Fig. 3 | Distribution of del(5q) cells within the CD34<sup>+</sup> compartment of MDS patients.** **A** UMAP representing all the MDS samples integrated and colored by cell type. The density map represents the distribution of the cells classified as del(5q) by the two algorithms. **B** UMAPs with density maps representing the distribution of del(5q) cells per individual patient. **C** Barplots showing the number of del(5q) and non-del(5q) cells composing each cell type for each MDS patient. **D** Heatmap

representing the enrichment of del(5q) cells ( $-\log_{10}(p\text{-value})$ ) in each cell type. Any color different from white represents a statistically significant enrichment of del(5q) cells ( $p\text{-value} < 0.05$ ).  $p$ -values were calculated using the one-sided hypergeometric test. The number of biologically independent replicates (cells) used for the hypergeometric test and the exact enrichment  $p$ -values can be found in the Source Data.

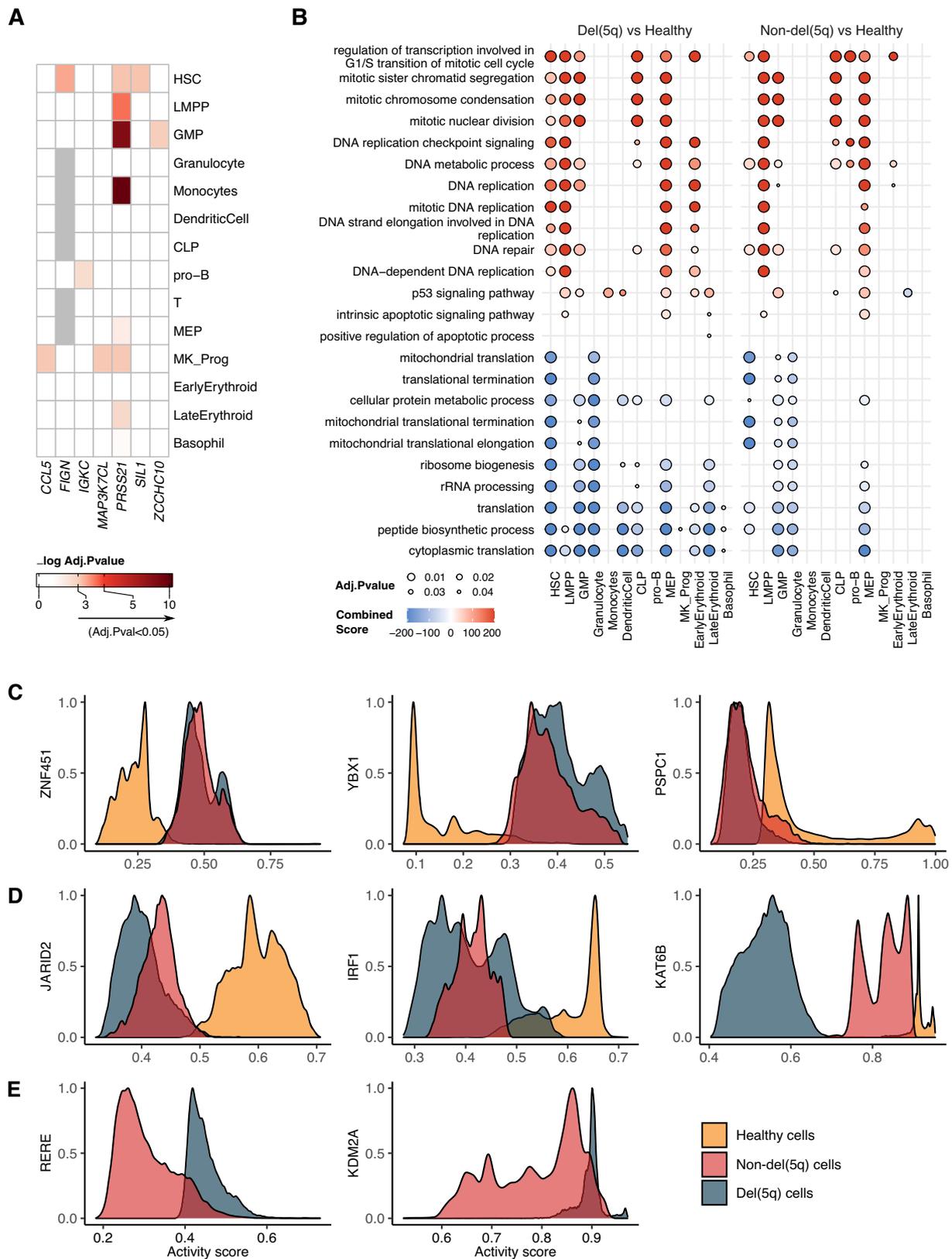
hemopoiesis, whereas those found in MDS progenitors were enriched in negative regulation of translation, oncogenic MAPK signaling and HIF-1 signaling (Fig. 5D). Focusing on interactions driven by del(5q) and non-del(5q) cells within the patients (Fig. 5B), we observed very subtle differences regarding the communicational pattern and the number of interactions observed for each of the compartments, and there were no interactions specifically established between del(5q) cells, corroborating the high similarity already described between del(5q) and non-del(5q) cells. Overall, our results are consistent with the previously described lack of significant differences in gene expression between del(5q) and non-del(5q) cells, suggesting that deregulation of hematoipoiesis in patients with 5q MDS affects all CD34<sup>+</sup> cells.

To uncover specific interactions that may contribute to the disease, we next focused on those interactions that had been gained or lost in MDS versus controls. There were 17 interactions identified in patients that were totally absent in healthy individuals, suggesting that additional communications arise when developing the disease. For instance, *AGTRAP* expressed in monocyte and late erythroid progenitors interacted with *RACK1* in HSCs, LMPPs, MEPs, pro-B and basophil progenitors (Fig. 5E). *AGTRAP* is known to be implicated in hematopoietic cell proliferation and survival<sup>42</sup>, whereas *RACK1* has been postulated as a potential therapeutic target for promoting proliferation in

other myeloid neoplasms<sup>43,44</sup>. The fact that these molecules are highly expressed in MDS could potentially be contributing to the enhanced proliferation observed in MDS cells. In contrast, there were 37 interactions that appeared in the healthy donors and were absent in the patients, including the one established between *HMGB1* expressed in CLPs, DC, granulocyte, basophil, megakaryocyte, early erythroid and late erythroid progenitors, and *CXCR4* present in HSCs (Fig. 5F). *HMGB1-CXCR4* interaction is known to trigger the recruitment and activation of inflammatory cells in tissue regeneration<sup>45,46</sup>, thus its loss could have a negative impact on the bone marrow niche. In summary, these analyses may allow the identification of potential interactions implicated in the pathogenesis of the disease that could represent new therapeutic targets.

**Effect of Lenalidomide treatment on the transcriptional programs of del(5q) and non-del(5q) cells from MDS patients**

We next aimed to understand the effect of treatment with the standard-of-care, lenalidomide, on the transcriptional alterations observed in del(5q) and non-del(5q) cells. We performed scRNA-seq on CD34<sup>+</sup> cells of two patients (Patient\_5-6), which had achieved hematological response (one with partial cytogenetic response (PCR), and the other one with complete cytogenetic response (CCR), respectively) (clinical

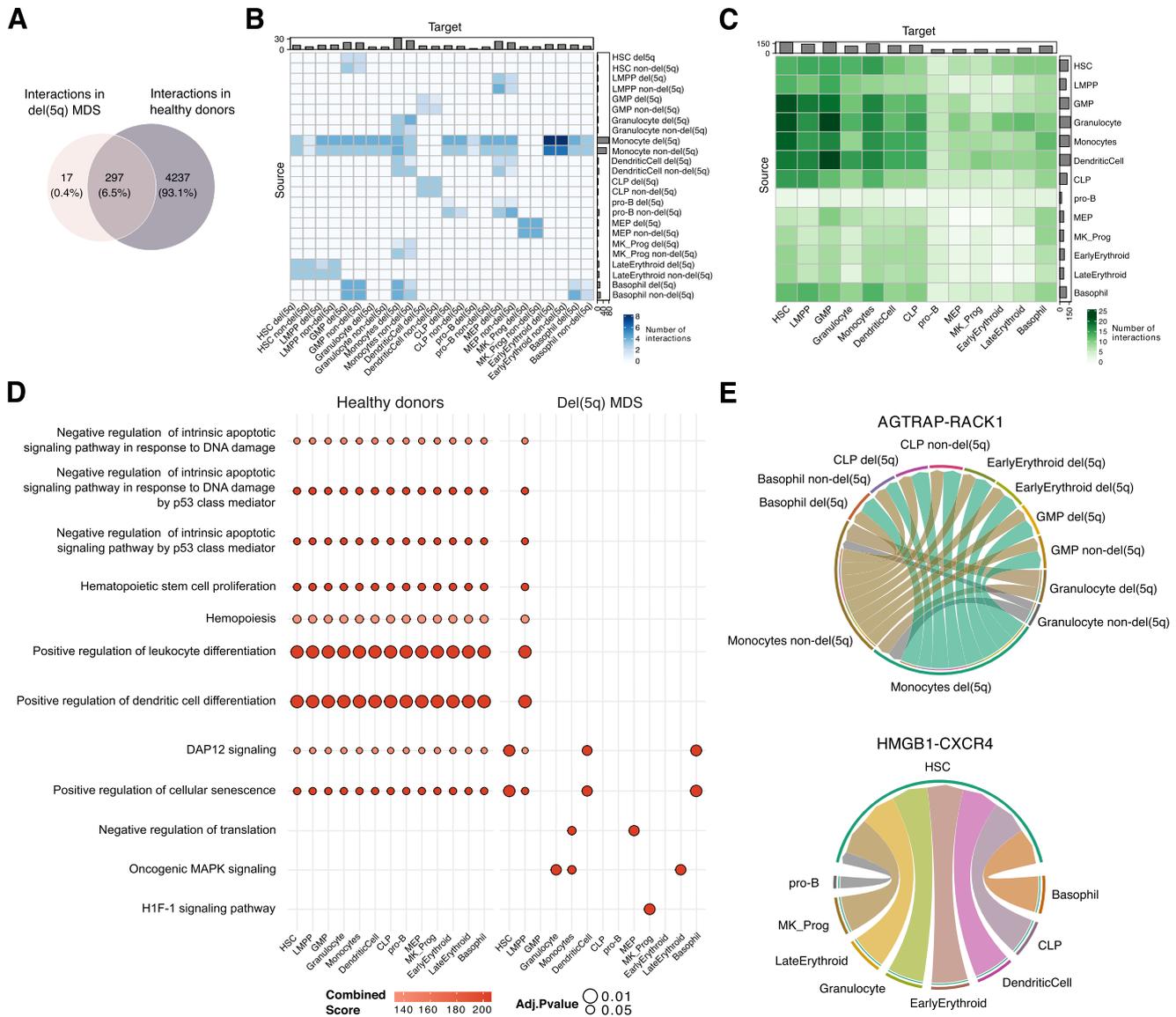


information in Supplementary Table 1). Data were integrated, clustered, manually annotated, and del(5q) cells were identified as described before (Fig. 6A, B). Patients showed different percentages of del(5q) cells which were consistent with karyotype results (Fig. 6C): the patient with PCR showed 1939 cells with del(5q) (37.13%), whereas the patient showing CCR presented only 11 cells with del(5q) after

treatment (0.15%), validating the persistence of del(5q) progenitor cells at the time of complete clinical and cytogenetic remission<sup>9</sup>. Similar to what we observed at diagnosis, the distribution of del(5q) cells was heterogeneous among patients (Fig. 6A, B), and both responders exhibited a statistically significant del(5q) enrichment in GMPs and erythroid progenitors. Interestingly, patient with PCR also

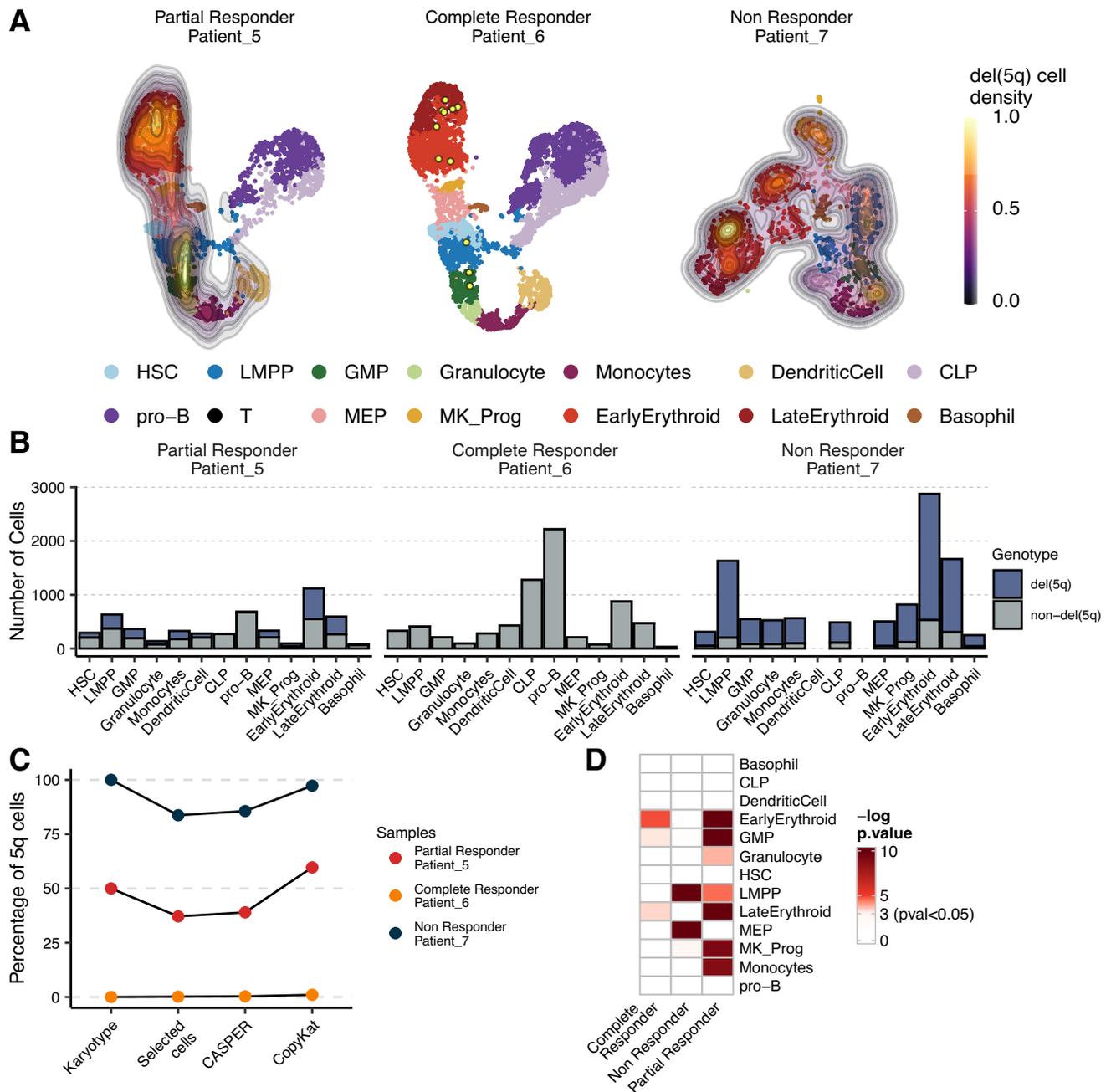
**Fig. 4 | Differential expression analysis between del(5q) and non-del(5q) cells within MDS samples exposes transcriptional similarities.** **A** Heatmap representing the differentially expressed genes (Benjamini–Hochberg-adjusted  $p$ -values  $< 0.05$  and  $|\log_{2}FC| > 2$ ) between del(5q) and non-del(5q) cells within each hematopoietic progenitor. The heatmap was created by combining  $n = 4$  del(5q) MDS patients and generating pseudobulks per cell type. The two-sided edgeR’s Likelihood Ratio Test was used to calculate  $p$ -values. The exact number of biologically independent replicates (del(5q) and non-del(5q) progenitor cells), as well as the specific  $p$ -values for each differentially expressed gene can be found in the Source Data. **B** Dotplot representing statistically significant biological processes and pathways (Benjamini–Hochberg-adjusted  $p$ -values  $< 0.05$ ) for differentially

expressed genes obtained in del(5q) versus Healthy and the non-del(5q) versus Healthy contrasts. The one-sided hypergeometric test was used to calculate  $p$ -values. Del(5q) and non-del(5q) cells were derived from  $n = 4$  del(5q) MDS patients, whereas healthy cells were derived from  $n = 3$  healthy donors. Biologically independent replicates (del(5q), non-del(5q) and healthy progenitor cells) used for each comparison are specified in the Source Data. **C–E** Histograms representing the activity score in all the cells separated by conditions. Some regulons behaved similarly in the MDS samples (non-del(5q) and del(5q) cells) compared to healthy cells (**C**), while other regulons behaved differently in the three different conditions (**D**). Some inferred regulons had an activity score on the MDS samples, while lacking on the healthy samples (**E**).



**Fig. 5 | Cell-to-cell communication analysis reveals shared and unique interactions in del(5q), non-del(5q) and healthy cells.** **A** Venn diagram showing the number of unique interactions in del(5q) MDS and healthy samples. Healthy unique interactions were considered as those present in at least one of the healthy individuals, while MDS unique interactions were those that were present in all the patients. Interactions were inferred from  $n = 4$  del(5q) MDS patients and  $n = 3$  healthy donors. **B** Heatmap depicting the number of interactions triggered by del(5q) and non-del(5q) MDS cells, **C** as well as those established among healthy hematopoietic progenitors. The Source represents the cell types that express the

ligand, whereas the Target represents the cells that express the receptor. **D** Dotplot representing statistically significant biological processes and pathways (Benjamini–Hochberg-adjusted  $p$ -value  $< 0.05$ ) in which are enriched the encoding genes taking part in the healthy and MDS interactions. The one-sided hypergeometric test was used to calculate  $p$ -values, whose exact values can be found in the Source Data. **E** Chord diagram representing the unique MDS interaction AGTRAP-RACK1 among different del(5q) and non-del(5q) progenitors. **F** Chord diagram depicting the unique healthy interaction HMGB1-CXCR4 established by healthy hematopoietic progenitors.



**Fig. 6 | Distribution of del(5q) cells within CD34<sup>+</sup> progenitors after lenalidomide treatment.** **A** UMAP depicting the del(5q) density across the different hematopoietic progenitors obtained in three patients after lenalidomide treatment. HSC hematopoietic stem cells, LMPP lymphoid-primed multipotent progenitors, GMP granulocyte-monocyte progenitors; granulocyte progenitors; monocyte progenitors; dendritic cell progenitors, CLP common lymphoid progenitors; B-cell progenitors; T-cell progenitors, MEP megakaryocyte-erythroid progenitors, MK\_Prog megakaryocyte progenitors; early erythroid progenitors; late erythroid progenitors; basophil progenitors. **B** Barplots showing the number of

del(5q) and non-del(5q) cells composing each cell type for each MDS patient. **C** Percentage of the cells identified as del(5q) by karyotype, CASPER, CopyKat, and the selection by intersecting the two algorithms. **D** Heatmap representing the enrichment of del(5q) cells ( $\log_{10}(p\text{-value})$ ) in each cell type. Any color different from white represents a statistically significant enrichment of del(5q) cells ( $p\text{-value} < 0.05$ ).  $P$ -values were calculated using the one-sided hypergeometric test. The number of biologically independent replicates (cells) used for the hypergeometric test and the exact enrichment  $p$ -values can be found in the Source Data.

showed an enrichment in LMPPs, megakaryocyte, monocyte, and granulocyte progenitors (Fig. 6D).

We have demonstrated in the previous analyses that at diagnosis both del(5q) and non-del(5q) progenitors displayed transcriptional profiles linked to an aberrant hematopoiesis. Since both PCR and CCR patients were in hematological response, we hypothesized that the remaining CD34<sup>+</sup> cells after lenalidomide treatment, which are mainly composed of non-del(5q) progenitors, must be able to promote

improved hematopoiesis and thus restore the transcriptional profile of normal progenitor cells. To demonstrate that lenalidomide, besides the potential apoptosis of del(5q) cells, could reverse transcriptional alterations harbored by non-del(5q) cells in responder patients, we performed a DE analysis between non-del(5q) cells from the CCR and PCR and the four patients at diagnosis. This comparison revealed significant transcriptional changes after lenalidomide treatment, resulting in 622–3609 genes in the different progenitor populations

for the CCR (Supplementary Fig. 4C, Supplementary Data 2) and between 458–2409 genes for the PCR (FDR < 0.05 and  $|\log_{2}FC| > 2$ ) (Supplementary Fig. 4D, Supplementary Data 2). Note that these transcriptional differences are significantly greater than those related to patient heterogeneity at diagnosis (see previous sections), indicating that most uncovered altered genes after treatment are probably due to treatment effect rather than patient heterogeneity. Genes altered upon treatment were enriched in ubiquitination and proteasome-mediated catabolic processes, and in phosphatidylinositol related pathways, which is in line with the mechanism of action described for lenalidomide in MDS patients<sup>47,48</sup>. Moreover, we detected an enrichment in autophagy-related processes. Overall, our results suggested an increase of the two most important protein degradation pathways in non-del(5q) cells upon lenalidomide treatment (Fig. 7A, first and second panels, Supplementary Table 2). Furthermore, hematopoietic progenitors exhibited an increased expression of genes involved in erythroid differentiation and erythropoietin signaling after treatment (Fig. 7B), validating the enhanced erythropoiesis in response to treatment<sup>19</sup>. Our analyses also detected a positive enrichment of *PD-L1* expression and *PD-1* checkpoint in non-del(5q) cells of the responder patients after treatment, suggesting a potential immunosuppressive mechanism of these cells in response to lenalidomide (Fig. 7A, first and second panels, Supplementary Table 2).

GRN analyses evidenced that some of the alterations described at diagnosis were potentially reverted after treatment in non-del(5q) cells. *IRF1*, the master HSC regulator located in 5q31.1, which showed abnormally low activity at diagnosis in non-del(5q) cells, showed an increased activity after treatment in both patients, with the PCR not reaching the activity level seen for healthy cells, and the CCR showing an augmented activity comparable to the healthy cells (Fig. 7C). *KAT6B*, whose lower expression has been associated with impaired myeloid differentiation, showed an augmented activity in both patients despite not reaching the activity level of healthy cells. Finally, *CUX1*<sup>49</sup>, a TF frequently mutated in myeloid malignancies and whose knockdown leads to an MDS-like phenotype, presented similar activity in non-del(5q) cells, showing higher activity than at diagnosis (Fig. 7C).

Importantly, although some transcriptional lesions were reverted upon lenalidomide treatment, non-del(5q) cells continue exhibiting altered expression of ribosome-related genes, showing a negative enrichment of processes related to ribosomes, translation, and mitochondrial translation when compared to healthy cells (Fig. 7A, fourth and fifth panels, Supplementary Fig. 4E, F, Supplementary Data 2 and Supplementary Table 2). After treatment, early and late erythroid non-del(5q) progenitors from responding patients showed no statistically significant changes in these pathways. Moreover, GRN analyses detected groups of regulons with similar activity for non-del(5q) cells at diagnosis and after treatment response, but with a different activity to the healthy cells, indicating that lenalidomide did not affect their aberrant activity. Some examples included the tumor suppressor *JARID2*<sup>34,35</sup>, *ZNF451*, a TF whose high expression in leukemic cells has been associated with poor outcome<sup>50</sup>, and *NCOR1*, a regulator of erythroid differentiation<sup>51</sup> (Fig. 7C). Moreover, non-del(5q) cells exhibited, both at diagnosis and after treatment, abnormal high activity of two regulons that were not active in healthy cells: *ADNP* and *SMARCE1* (Fig. 7C). Globally, these results indicate that treatment with lenalidomide has the potential to revert some of the transcriptional alterations present at diagnosis in non-del(5q) cells at least in patients that responded to lenalidomide. Nevertheless, some of the transcriptional alterations present at diagnosis were not modified which could be relevant for abnormal hematopoiesis, and potentially, for the future relapse of the patients.

In line with what has been observed in non-del(5q) cells from the PCR, the remaining del(5q) cells generally exhibited significant upregulation of genes involved in ubiquitin and phosphatidylinositol signaling and, autophagy and apoptosis pathways when compared to

del(5q) cells at diagnosis (Supplementary Figs. 4G, 6A, Supplementary Data 2 and Table 3), which is consistent with the mechanism of action of lenalidomide<sup>47,48</sup>. However, these cells showed reduced expression of genes implicated in ribosomal and mitochondrial translation compared to diagnosis, along with diminished expression of DNA repair associated genes. (Supplementary Fig. 6A). This suggests that lenalidomide does not fully reverse key transcriptional alterations that may underlie the ribosomopathy characterizing the disease.

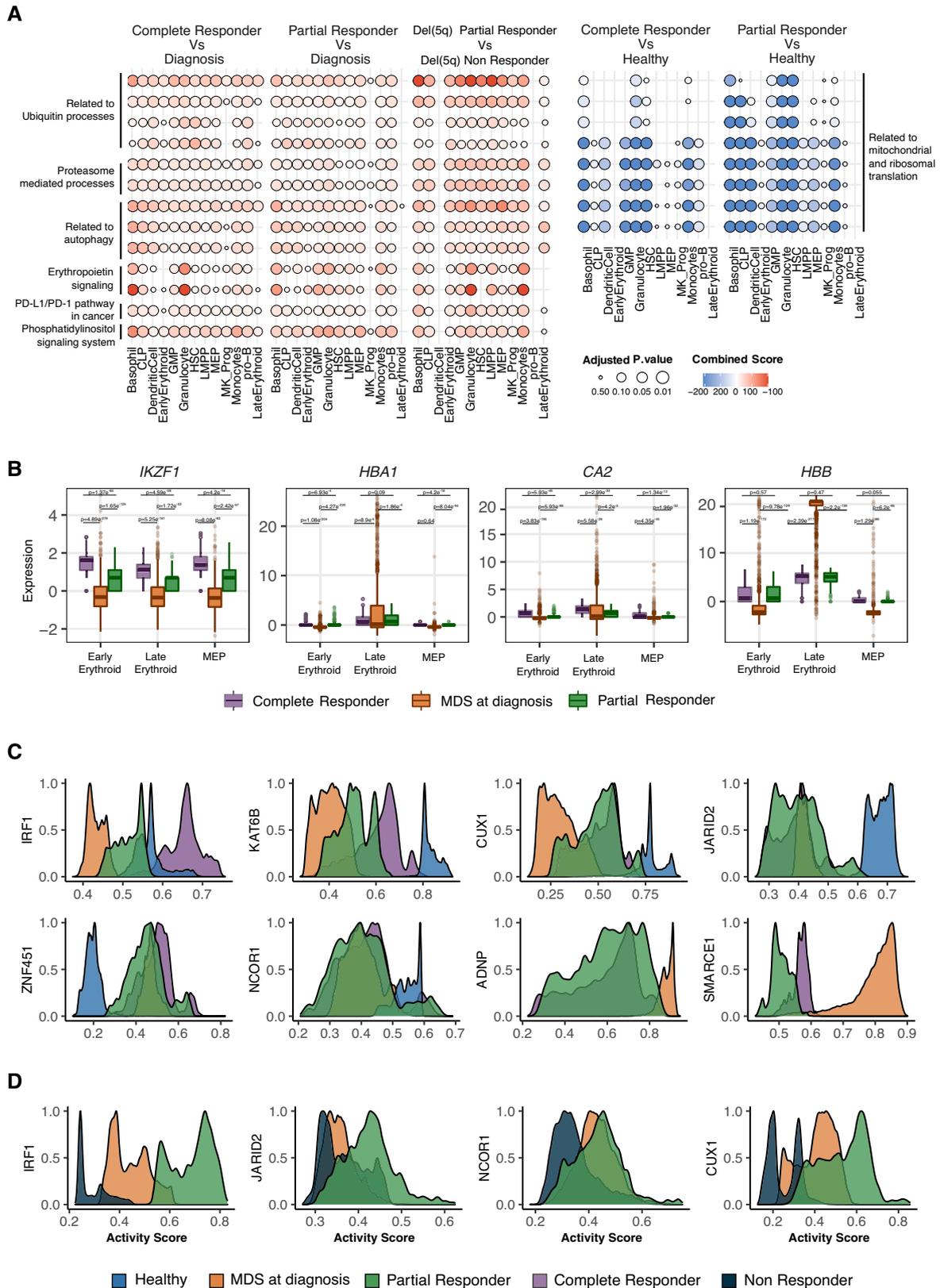
### Lenalidomide does not correct transcriptional alterations of del(5q) cells of a refractory MDS patient

Finally, to understand the transcriptional alterations associated with a lack of hematological response after lenalidomide treatment, we performed scRNAseq on CD34<sup>+</sup> cells of an additional patient (Patient\_7), who was refractory to lenalidomide (non-responder, NR). Data were processed as described previously (clinical information in Supplementary Table 1), showing 83.8% of del(5q) cells (Fig. 6A–C), with a statistically significant increased abundance in LMPPs, MEPs and megakaryocyte progenitors (Fig. 6D). We then analyzed the transcriptional differences between the remaining del(5q) cells of the responder that presented PCR, and those of the NR patient. This analysis identified 116–2244 differentially expressed genes (FDR < 0.05) per progenitor (Supplementary Fig. 4H, Supplementary Data 2). Del(5q) cells from the patient in PCR showed statistically significant enrichment in processes and pathways related to protein ubiquitination, proteasomal protein catabolic process, phosphatidylinositol and autophagosome when compared to the NR. Moreover, these cells also exhibited an increased expression of genes involved in erythropoietin signaling when compared to the cells from the NR (Fig. 7A, third panel). Interestingly, these processes are similar to the ones detected for non-del(5q) cells when comparing these cells to those at diagnosis (see previous section), and have been described as a lenalidomide response in non-del(5q) MDS patients<sup>47,48</sup>. The remaining del(5q) cells from the patient in PCR also exhibited enrichment of *PD-L1* expression and *PD-1* checkpoint pathway when compared to the refractory patient. These analyses suggested low transcriptional alterations promoted by lenalidomide treatment in the NR patient. Accordingly, DE analysis of del(5q) cells at diagnosis and after treatment in the NR patient yielded 20–121 differentially expressed genes per hematopoietic progenitor (Supplementary Fig. 4I, Supplementary Data 2). These few differences resulted in subtle changes in protein ubiquitination and cell cycle-related processes after treatment (Supplementary Fig. 6B, Supplementary Table 4), showcasing that lenalidomide did not have a high transcriptional impact on del(5q) cells of the NR.

GRN analysis demonstrated a large number of regulons that showed changes in activity after treatment in del(5q) cells from the patient in the PCR but not in the refractory patient. For example, regulons driven by *IRF1*, *JARID2*, *NCOR1*, and *CUX1*, which showed aberrant low activity at diagnosis that was partially recovered upon treatment, presented very reduced activity in the NR patient, which was lower than that observed in the PCR, and at diagnosis (Fig. 7D). Collectively, these results suggest that in NR patients, lenalidomide treatment is not able to reverse part of the transcriptional lesions carried by (5q) cells, which seems to be associated with the lack of hematological response.

## Discussion

Establishing the relationship between genomic and transcriptional abnormalities in hematological malignancies has allowed researchers to characterize the molecular pathogenesis of diseases such as MDS and to identify new potential targets<sup>10–14,18</sup>. However, for the most part, these studies have been performed using bulk sequencing data, which precludes a direct association on a per cell basis between genomic and transcriptomic alterations. Using scRNAseq data from CD34<sup>+</sup> cells from patients with del(5q) MDS, we have been able to identify cells



with del(5q) and non-del(5q), which enabled us to compare the transcriptional profile and GRNs of both populations within the same patient, thereby mitigating potential confounding factors associated with interindividual variability and microenvironmental influences, and thus made strides in the understanding of the molecular pathogenesis of del(5q) MDS.

Single cell RNA-seq coupled with CNA detection algorithms offers the opportunity to link the genetic information to the transcriptional profile of each individual cell. Nonetheless, this type of analyses remain challenging due to the sparse and noisy nature of single-cell data<sup>52</sup>. To ensure robust findings we have employed two CNA detection methods, one based on transcript abundance<sup>23</sup> and another utilizing allelic

**Fig. 7 | Differential expression between treated and untreated patients unravels persistent transcriptional alterations after lenalidomide treatment.**

**A** Dotplot representing statistically significant biological processes and pathways (Benjamini–Hochberg-adjusted  $p$ -value  $< 0.05$  and  $|\log_{2}FC| > 2$ ) for differentially expressed genes obtained in different comparisons: non-del(5q) cells of the complete responder vs at diagnosis (1st panel); non-del(5q) cells of the partial responder vs at diagnosis (2nd panel); del(5q) cells of the partial responder vs the non-responder (3rd panel); non-del(5q) cells of the complete responder vs healthy cells (4th panel); non-del(5q) cells of the partial responder vs healthy cells (5th panel). For  $p$ -value calculation, one-sided hypergeometric test was used. Specific  $p$ -values for statistically significant biological processes can be found in the Source Data. The detailed breakdown of the grouped processes shown can be found in Supplementary Table 2. **B** Boxplot showing the normalized expression of erythroid differentiation-related genes for non-del(5q) cells in MDS at diagnosis ( $n = 4$ ) or after

treatment with lenalidomide (partial responder,  $n = 1$ ; complete responder,  $n = 1$ ). Biologically independent replicates (cells) for hematopoietic progenitors were: Early Erythroid:  $n = 5196$  (MDS at diagnosis);  $n = 2200$  (Partial Responder);  $n = 3496$  (Complete Responder); Late Erythroid:  $n = 7168$  (MDS at diagnosis);  $n = 1056$  (Partial Responder);  $n = 1880$  (Complete Responder); MEP:  $n = 3108$  (MDS at diagnosis);  $n = 824$  (Partial Responder);  $n = 844$  (Complete Responder). Two-sided Wilcoxon signed-rank test was used to calculate  $p$ -values, that were then Benjamini–Hochberg-adjusted. Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers) as well as outliers (single points). Exact  $p$ -values are shown within the figure. **C** Graphs representing the activity scores of proliferation and differentiation-associated transcription factors in healthy cells and non-del(5q) cells of MDS patients at diagnosis and after lenalidomide treatment, as well as **D** in del(5q) cells of MDS patients at diagnosis, with a partial response and with no response to lenalidomide. Specific activity scores can be found in the Source Data.

imbalance in heterozygous SNPs<sup>24</sup>. While expression-based approaches operate under the assumption that amplifications or deletions generally result in up- or down-regulation of genes within the affected region of the genome, allele-based methodologies focus on analyzing deviations in heterozygous allele frequency. We acknowledge the potential for minor misclassification due to the inherent complexity of applying CNA detection algorithms to scRNA-seq data, however, we believe that this would be minimal, as clearly demonstrated in the results section (Supplementary Fig. 2). Our confidence in employing two complementary methods stems from the belief that this approach enhances classification accuracy and mitigates false-positive results arising from local expression variations unrelated to genomic copy numbers, as well as from data sparsity and allele-specific transcriptional stochasticity<sup>23,24</sup>.

The stem-cell origin of del(5q) MDS has been previously demonstrated<sup>9,53</sup> as well as the presence of del(5q) cells even at the stage of cytogenetic response<sup>9</sup>. By being able to dissect the presence of del(5q) in individual cells and identifying the different types of progenitors, our study suggests uneven distribution of del(5q) in different progenitor cell populations, and also an heterogeneous distribution according to the patients. Although we cannot establish a correlation between the phenotype of the disease, the clinical symptoms, and the specific distribution of del(5q), we were able to identify an enrichment of del(5q) cells in GMP, megakaryocyte and erythroid progenitors in patients, not described to date. Future studies with larger cohorts of cases may uncover the nature of the detected enrichment and provide possible associations between the distribution of del(5q) cells and clinical characteristics of the patients.

The comparison between the transcriptional profiles of del(5q) cells and non-del(5q) within the same patient provides insights regarding the impact of 5q deletions on pathogenesis of del(5q) MDS. The presence of similar transcriptional alterations in both del(5q) and non-del(5q) suggests that in del(5q) MDS patients, both types of cells are implicated in the disease and contribute to the promotion of aberrant hematopoietic differentiation. Nevertheless, the boosted aberrant activity of some regulons in del(5q) cells, may exert a more prominent role of these cells in the aberrant hematopoiesis of these patients. Furthermore, this fact might indicate the presence of other additional factors, such as a disrupted microenvironment, as evidenced by recent studies elucidating its impact on the initiation and progression of MDS<sup>54,55</sup>. Moreover, the presence of additional shared genetic lesions between del(5q) and non-del(5q) cells may play a key role in the observed transcriptional similarities. Future studies with larger cohorts of cases may elucidate the underlying nature of these observed transcriptional similarities.

Specifically, del(5q) and non-del(5q) cells presented alterations in biological processes described in del(5q) MDS, including cell proliferation<sup>56,57</sup>, p53 signaling pathway<sup>14</sup> and apoptosis. In this regard, previous murine models generated by the inactivation of *RPS14* described a p53 dependent apoptosis in the erythroid lineage between

basophilic/early chromophilic erythroblasts to poly/orthochromatophilic erythroblasts<sup>30</sup>, but our data suggests that it could already occur at the CD34<sup>+</sup> progenitor level. Furthermore, in addition to the negative enrichment in ribosome and cytoplasmic translation processes identified in our data, which are consistent with previous studies<sup>2,30,32</sup>, we observed an altered function of mitochondrial ribosomes in HSCs and granulocyte-lineage progenitors. Mitochondrial translation and mitoribosomes fulfill a pivotal function in cell cycle regulation and apoptosis signaling<sup>58</sup>, and several works have found gene expression alterations in mitoribosome or associated proteins in relation to different cancers<sup>59</sup>. Hence, the altered mitochondrial translation in del(5q) cells could be contributing to the molecular pathogenesis of del(5q) MDS. Despite these findings, a deeper GRN analysis led to the characterization of a number of transcriptional differences between del(5q) and non-del(5q) cells, identifying a group of regulons with different activity between both genotypes. Some of the GRNs specifically altered in del(5q) cells, such as those driven by *JARID2*, *KAT6B*, *RERE* or *KDM2A*, seem to be relevant for proliferation and myeloid differentiation, supporting the concept that cells harboring the deletion may have a more prominent role in the promotion of altered hematopoiesis. Altogether, our data manifests a general transcriptional dysregulation in all the progenitor cells of del(5q) MDS patients, and shows that progenitors harboring the del(5q) deletion show additional alterations.

Lenalidomide represents the standard-of-care for transfusion-dependent del(5q) MDS patients. Our analyses suggest a close relationship between lenalidomide-mediated reversal of the transcriptional alterations harbored by hematopoietic progenitors and hematological response. In this sense, we observed that lenalidomide treatment alters the transcriptional profile of both del(5q) and non-del(5q) cells of responder patients, reverting part of the alterations observed in these cells. For example, the remaining non-del(5q) cells of responder patients show increased levels of proteasome-mediated catabolic and autophagy related processes after treatment, suggesting a compensatory increase of the two main mechanisms of intracellular protein degradation as a result of the treatment. Moreover, we detected a direct effect of lenalidomide in erythropoietin signaling, promoting the upregulation of genes involved in erythropoiesis, a mechanism that so far has been described for non-del(5q) MDS<sup>60</sup>. Nevertheless, certain alterations that were initially present at the time of diagnosis in del(5q) and non-del(5q) cells remain detectable at this stage of clinical evaluation, which could explain the relapses seen after the treatment. This association between reversal of transcriptional alterations and clinical response is further supported by the fact that both del(5q) and non-del(5q) cells of a non-responder patient presented very few transcriptional alterations upon lenalidomide treatment. Altogether, these results may be at odds with previous data suggesting that lenalidomide induces synthetic lethality to suppress the malignant clone without significant effect on the growth of cytogenetically normal CD34<sup>+</sup> cells<sup>61</sup>. Our data demonstrate that

lenalidomide affects both del(5q) and non-del(5q) cells, and suggest that the hematological response of del(5q) MDS patients may be due to the correction of transcriptional alterations in both types of cells.

Finally, the characterization of the transcriptional profile of remaining cells after lenalidomide treatment could help to unveil novel therapeutic strategies that could be effective in the eradication of malignant cells. In this sense, our results showed an enrichment in *PD-1/PDL-1* pathway after the treatment in non-del(5q) cells which was not detected in the NR patient, suggesting a putative immune evasion mechanism of these cells. Interestingly, lenalidomide has shown increased cytotoxic activity in combination with immune-checkpoint blockade in multiple myeloma<sup>62</sup>. Thus, future studies will confirm if, as our results suggest, this combination could serve as a promising therapeutic strategy for these patients.

## Methods

The research performed in this work complies with all relevant ethical regulations: the study was approved by the research ethics committee of University of Navarra, and informed consent was obtained from all patients and healthy donors.

### Sample collection

Bone marrow aspirates were obtained from healthy age-matched controls [ $n = 3$ , median age, 72 years, range, 61–84 years] and from patients with MDS [ $n = 7$ , median age, 84 years, range, 80–91 years] from the Clinica Universidad de Navarra and collaborating hospitals after the study was approved by the research ethics committee of University of Navarra, and informed consent was obtained. MDS patients and healthy donors were not economically compensated for the samples donated. Patient's data were fully anonymized, and all patients provided informed written consent for the use and publication of data from their medical records such as age, sex, and diagnosis for research purpose. Healthy controls were patients undergoing orthopedic surgery. As del(5q) MDS is a subtype of MDS which is predominantly found in females, and these samples are not very abundant, in the present study we included all the patients for which we were able to obtain samples. Thus, we were not able to select patients by sex or to carry out a sex or gender analysis. The patients' clinical characteristics are shown in Supplementary Table 1.

### Fluorescence-activated cell sorting

For purification of CD34<sup>+</sup> cells, BM samples were lysed in 1X of BULK lysis buffer (150 mM NH<sub>4</sub>Cl, 10 mM CH<sub>3</sub>CO<sub>3</sub> and 0.1 mM EDTA in deionized water) for 15 min at a sample to bulk ratio of 1:10, and centrifuged for 5 min at 500 × *g* to eliminate red blood cells. Next, cells were stained using CD34-APC (clone 581; Beckman Coulter, #1M2472, lot number 200504, dilution 5:100) and CD45-PerCPCy5.5 (clone HI30; Biolegend, #304028, lot number 200504, dilution 1:100) for 15 min at RT. CD34<sup>+</sup> CD45<sup>+</sup> cells were then sorted in a BD FACSAria II (BD Biosciences) and directly used for scRNA seq analysis. BD FACSDIVA v8 software was used for flow cytometry data analysis.

### scRNA-seq library preparation

The transcriptome of the bone marrow CD34<sup>+</sup> cells were examined using NEXTGEM Single Cell 3' Reagent Kits v3 and v3.1 (10X Genomics) according to the manufacturer's instructions. Between 5000 and 17,000 cells, depending on the donor, were loaded at a concentration of 700–1200 cells/μL onto a Chromium Controller instrument (10× Genomics) to generate single-cell gel bead-in-emulsions (GEMs). In this step, each cell was encapsulated with primers containing a fixed Illumina Read 1 sequence, a cell-identifying 16-bp 10× barcode, a unique molecular identifier (UMI), and a poly-dT sequence. Upon cell lysis, reverse transcription yielded full-length, barcoded cDNA, which was then released from the GEMs, amplified using polymerase chain reaction, and purified using magnetic beads (SPRIselect, Beckman

Coulter). Enzymatic fragmentation and size selection were used to optimize the cDNA size prior to library construction. Fragmented cDNA was then end-repaired, A-tailed, and ligated to Illumina adapters. A final polymerase chain reaction amplification using barcoded primers was performed for sample indexing. Library quality control and quantification were performed using a Qubit 3.0 Fluorometer (Life Technologies) and an Agilent 4200 TapeStation System (Agilent), respectively. Sequencing was performed on NextSeq500 and NextSeq2000 instruments (Illumina) at an average depth of 30,000 reads/cell.

### Single-cell RNA-seq analysis

The scRNA-seq data was demultiplexed and aligned to the human reference genome (GRCh38). The feature-barcode matrix was quantified using Cell Ranger (v6.0.1) from 10X Genomics. Computational analysis was carried out using Seurat<sup>63</sup> (v4.2.0). In order to remove the possible heterotypic and homotypic doublets/multiplets, cells were analyzed with scrublet<sup>64</sup> (v0.2.3) and removed. Cells underwent quality control filters based on the number of detected genes, number of unique molecular identifiers (UMIs), the cell complexity by calculating the ratio between the logarithm of the number of genes and the logarithm of number of UMIs, and the proportion of UMIs mapped to mitochondrial and ribosomal genes per cell. Each sample was analyzed searching for effects of cell cycle heterogeneity by calculating cell cycle phase scores based on canonical markers, without needing to apply any correction to the expression of the cells. Thus, each dataset was normalized, 2000 highly variable genes were identified, and unwanted sources of variation were removed. Integration of all datasets was performed using Seurat's canonical correlation analysis. Samples from del(5q) MDS patients and healthy donors were integrated using the Seurat pipeline selecting 3000 integration features. Counts were log normalized and scaled, and integration was performed on the 2000 genes with the highest variability across samples, using 50 dimensions. Nonlinear dimensionality reduction was conducted using UMAP, after selecting the appropriate number of principal components by determining the number of principal components that exhibits cumulative percent greater than 90% and less than 5% variation associated with. To identify cell populations, we performed iterative Louvain modularity optimization clustering from the Seurat package. We evaluated cluster quality using multiple metrics, including Gini coefficient and Silhouette score. To characterize the cell types and states defined by each cluster, we manually reviewed the differentially expressed genes, uncovered by Seurat's FindMarkers function.

### 5q deletion analysis

To differentiate del(5q) cells from non-del(5q) cells, CopyKat<sup>23</sup> (v1.0.8) and CaSpER<sup>24</sup> (v0.2.0) were applied. CopyKat excels at uncovering large-scale aberrations by effectively identifying groups of significantly deviating copy number segments across the genome. On the other hand, CaSpER utilizes BAF information to offer precise delineation of deletion boundaries by modeling heterogeneity, incorporating prior knowledge, and accounting for dependencies. This synergy helps mitigate the risk of underestimating the deletion extent or missing it altogether, as can occur with single-method approaches. Moreover, the combined analysis could potentially account for biases associated with either method individually. CaSpER was used with a non-del(5q) MDS sample that had normal karyotype<sup>65</sup> as reference, whereas for CopyKat, a reference composed of a combination of three healthy samples was employed. To validate the generated results, we also analyzed a non-del(5q) MDS sample using the same reference as a negative control. The raw results from CaSpER underwent filtering by extracting large-scale events using a threshold of 0.75; by raising this threshold, we increase the needed number of amplifying or deleting events in order to support the detection of such aberration, increasing the robustness of the identified CNA events. Subsequently, results

were binarized as described in the tool's methods, classifying each arm of each chromosome's arm as amplified, neutral, or deleted. For the CopyKat results, a cell clustering was performed on the results from CopyKat based on the values of copy number alterations in 220 Kbp bins of the targeted region, obtaining a cluster composed of cells with negative values. Cells exhibiting negative copy number alterations in the chr5 q15-31 region, as determined by both methods, were classified as del(5q) cells. Conversely, cells for which no alterations were identified by either method were classified as non-del(5q) cells.

### Gene Regulatory Network analysis

For each comparison, 100 transcription factors and 1000 target genes were selected based on their variability, determined by calculating the maximum absolute deviation. To determine the optimal parameters, each analysis involved a cross-validation run of SimiC<sup>33</sup> (v1.0.0). The resulting Gene Regulatory Networks (GRNs) were visualized using the GRN incidence matrices provided by SimiC. Histograms for different regulons were computed using the "regulon activity score" provided by SimiC, allowing to see the regulatory activity in all the cells or by cell type. Additionally, this score was utilized to calculate the regulatory dissimilarity score for the selected cell clusters.

### Differential expression analysis

Different methodologies were employed depending on the specific contrasts being examined. When the contrast allowed for the comparison of cells from different samples within each phenotype, we utilized the Libra<sup>25</sup> (v1.0.0) framework in combination with edgeR-LRT<sup>66</sup> (v4.0.6). This approach involved generating pseudobulks for each cell type, per sample, effectively aggregating the expression profiles of cells from the same phenotype. By utilizing edgeR-LRT, statistical testing was performed to identify genes that exhibited significant differential expression while accounting for the batch effect. In cases where the contrast involved a limited number of samples in any of the phenotypes, we used Libra (v1.0.0) framework in combination with MAST<sup>67</sup> (v1.22.0) methodology. All the results for the different contrasts were filtered by an adjusted *p*-value lower than 0.05. Gene ontology enrichments were determined using enrichR (v3.2) based on biological process and molecular function gene sets from 2023.

### Cell to cell communication analysis

The analysis of cell-to-cell communication using Liana<sup>41</sup>(v0.1.7) was carried out individually for each sample, comparing the healthy samples with del(5q) MDS samples. Interactions were filtered based on their statistical significance (*p*-value < 0.05), using the *p*-value from CellChat and CellPhoneDB, and on the magnitude of the interaction (log<sub>10</sub> of the magnitude rank > 5). Gene ontology enrichments were determined using enrichR (v3.2) based on biological process and molecular function gene sets from 2023.

### Statistics and reproducibility

Statistical analyses were performed in R (v4.2.2) and SimiC<sup>33</sup> was run in python (v3.6.9). Analysis methods for single-cell RNA sequencing data are described in the corresponding method sections and each statistical method used is specified in the corresponding figure legend. The number of human samples or cells analyzed in each subpanel is also indicated in the figure legends. No statistical method was used to predetermine sample size, and no data were excluded from the analyses performed. When possible, investigators were blinded during the data analysis (i.e., the detection of del(5q) cells was performed without prior knowledge of the karyotype result).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All data needed to evaluate the conclusions in the paper are present in the manuscript and/or the Supplementary Materials. The scRNA-seq data generated in this study have been deposited in the Gene Expression Omnibus database (GEO) under the accession code GSE245452. The previous publicly available data used in this study, corresponding to scRNA-seq data of CD34<sup>+</sup> cells from healthy age-matched individuals, are available in GEO under the accession code GSE183328. The GRCh38 assembly of the human genome used is available at NCBI, under the accession code NCBI: GCA\_000001405.27. CellPhoneDB database is stored in <https://github.com/ventolab/CellphoneDB-data>. The biological process and molecular function gene set libraries used for gene ontology analyses are available in <https://maayanlab.cloud/Enrichr/enrichr>. Source data are provided as a Source Data file. Source data are provided with this paper.

### Code availability

The code and scripts for data reproducibility are available at [github.com/ML4BM-Lab/MDS\\_5q\\_2023](https://github.com/ML4BM-Lab/MDS_5q_2023); <https://doi.org/10.5281/ZENODO.10983466><sup>68</sup>.

### References

1. Acha, P. & Solé, F. Myelodysplastic Syndromes with Isolated del(5q): Value of Molecular Alterations for Diagnostic and Prognostic Assessment. *Cancers* **14**, 5531 (2022).
2. Ebert, B. L. Molecular dissection of the 5q deletion in myelodysplastic syndrome. *Semin. Oncol.* **38**, 621–626 (2011).
3. Komrokji, R. S., Padron, E., Ebert, B. L. & List, A. F. Deletion 5q MDS: Molecular and therapeutic implications. *Best. Pr. Res. Clin. Haematol.* **26**, 365–375 (2013).
4. Ebert, B. L. et al. Identification of RPS14 as a 5q2 syndrome gene by RNA interference screen. *Nature* **451**, 335–339 (2008).
5. Starczynowski, D. T. et al. Identification of miR-145 and miR-146a as mediators of the 5q- syndrome phenotype. *Nat. Med.* **16**, 49–58 (2010).
6. Youn, M. et al. MMP9 inhibition increases erythropoiesis in RPS14-deficient del(5q) MDS models through suppression of TGF- $\beta$  pathways. *Blood Adv.* **3**, 2751–2763 (2019).
7. Joslin, J. M. et al. Haploinsufficiency of EGR1, a candidate gene in the del(5q), leads to the development of myeloid disorders. *Blood* **110**, 719–726 (2007).
8. List, A. et al. Lenalidomide in the myelodysplastic syndrome with chromosome 5q deletion. *N. Engl. J. Med.* **355**, 1456–1465 (2006).
9. Tehrani, R. et al. Persistent Malignant Stem Cells in del(5q) Myelodysplasia in Remission. *N. Engl. J. Med.* **363**, 1025–1037 (2010).
10. Hofmann, W. K. et al. Characterization of gene expression of CD34+ cells from normal and myelodysplastic bone marrow. *Blood* **100**, 3553–3560 (2002).
11. Pellagatti, A. et al. Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. *Blood* **108**, 337–345 (2006).
12. Pellagatti, A. et al. Identification of gene expression-based prognostic markers in the hematopoietic stem cells of patients with myelodysplastic syndromes. *J. Clin. Oncol.* **31**, 3557–3564 (2013).
13. Sridhar, K., Ross, D. T., Tibshirani, R., Butte, A. J. & Greenberg, P. L. Relationship of differential gene expression profiles in CD34+ myelodysplastic syndrome marrow cells to disease subtype and progression. *Blood* **114**, 4847–4858 (2009).
14. Pellagatti, A. et al. Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia* **24**, 756–764 (2010).
15. Berastegui, N. et al. The transcription factor DDIT3 is a potential driver of dyserythropoiesis in myelodysplastic syndromes. *Nat. Commun.* **13**, 7619 (2022).

16. Harewood, L. & Fraser, P. The impact of chromosomal rearrangements on regulation of gene expression. *Hum. Mol. Genet.* **23**, R76–R82 (2014).
17. Jia, P. & Zhao, Z. Impacts of somatic mutations on gene expression: an association perspective. *Brief. Bioinform* **18**, 413–425 (2017).
18. Boulton, J. et al. Gene expression profiling of CD34+ cells in patients with the 5q- syndrome. *Br. J. Haematol.* **139**, 578–589 (2007).
19. Ebert, B. L. et al. An erythroid differentiation signature predicts response to lenalidomide in myelodysplastic syndrome. *PLoS Med.* **5**, 0312–0322 (2008).
20. Narla, A. et al. Dexamethasone and lenalidomide have distinct functional effects on erythropoiesis. *Blood* **118**, 2296 (2011).
21. Belickova, M. et al. Changes associated with lenalidomide treatment in the gene expression profiles of patients with del(5q). *Clin. Lymphoma Myeloma Leuk.* **12**, 375–383 (2012).
22. Krejčík, Z. et al. Aberrant expression of the microRNA cluster in 14q32 is associated with del(5q) myelodysplastic syndrome and lenalidomide treatment. *Cancer Genet.* **208**, 156–161 (2015).
23. Gao, R. et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* **39**, 599–608 (2021).
24. Serin Harmanci, A., Harmanci, A. O. & Zhou, X. CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. *Nat. Commun.* **11**, 1–16 (2020).
25. Squair, J. W. et al. Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 5692 (2021).
26. Manton, K. J. et al. Hypermethylation of the 5' CpG island of the gene encoding the serine protease Testisin promotes its loss in testicular tumorigenesis. *Br. J. Cancer* **92**, 760–769 (2005).
27. Niu, L., Guo, W., Song, X., Song, X. & Xie, L. Tumor-educated leukocytes mRNA as a diagnostic biomarker for non-small cell lung cancer. *Thorac. Cancer* **12**, 737 (2021).
28. Kittang, A. O., Sand, K., Brenner, A. K., Rye, K. P. & Bruserud, Ø. The Systemic Profile of Soluble Immune Mediators in Patients with Myelodysplastic Syndromes. *Int. J. Mol. Sci.* **17**, 1080 (2016).
29. Barlow, J. L. et al. A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. *Nat. Med.* **16**, 59–66 (2010).
30. Schneider, R. K. et al. Rps14 haploinsufficiency causes a block in erythroid differentiation mediated by S100A8/S100A9. *Nat. Med.* **22**, 288–297 (2016).
31. Zhang, L., McGraw, K. L., Sallman, D. A. & List, A. F. The role of p53 in myelodysplastic syndromes and acute myeloid leukemia: molecular aspects and clinical implications. *Leuk. Lymphoma* **58**, 1777–1790 (2017).
32. Ruggero, D. & Shimamura, A. Marrow failure: A window into ribosome biology. *Blood* **124**, 2784–2792 (2014).
33. Peng, J. et al. SimiC enables the inference of complex gene regulatory dynamics across cell phenotypes. *Commun. Biol.* **5**, 351 (2022).
34. Celik, H. et al. JARID2 Functions as a Tumor Suppressor in Myeloid Neoplasms by Repressing Self-Renewal in Hematopoietic Progenitor Cells. *Cancer Cell* **34**, 741–756.e8 (2018).
35. Puda, A. et al. Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. *Am. J. Hematol.* **87**, 245–250 (2012).
36. Willman, C. L. et al. Deletion of IRF-1, Mapping to Chromosome 5q31.1, in Human Leukemia and Preleukemic Myelodysplasia. *Science* **259**, 968–971 (1993).
37. Nilsson, A. R., Xian, H., Shalpour, S., Cammenga, J. & Karin, M. IRF1 regulates self-renewal and stress-responsiveness to support hematopoietic stem cell maintenance. *Sci. Adv.* **9**, eadg5391 (2023).
38. Khokhar, E. S. et al. Aging-associated decrease in the histone acetyltransferase KAT6B is linked to altered hematopoietic stem cell differentiation. *Exp. Hematol.* **82**, 43–52.e4 (2020).
39. Okamoto, K. et al. KDM2A-dependent reduction of rRNA transcription on glucose starvation requires HP1 in cells, including triple-negative breast cancer cells. *Oncotarget* **10**, 4743 (2019).
40. Tanaka, Y. et al. JmjC enzyme KDM2A is a regulator of rRNA transcription in response to starvation. *EMBO J.* **29**, 1510–1522 (2010).
41. Dimitrov, D. et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat. Commun.* **2022** **13**, 1–13 (2022).
42. Kwiatkowski, B. A. & Richard, R. E. Angiotensin II Receptor-Associated Protein (AGTRAP) Synergizes with Mpl Signaling to Promote Survival and to Increase Proliferation Rate of Hematopoietic Cells. *Blood* **114**, 3606 (2009).
43. Wu, H. et al. RACK1 deficiency synergizes with all-trans retinoic acid to induce apoptosis in human acute promyelocytic leukemia cells. *Mol. Cell Biochem.* **451**, 155–163 (2019).
44. Zhang, D. et al. RACK1 promotes the proliferation of THP1 acute myeloid leukemia cells. *Mol. Cell Biochem.* **384**, 197–202 (2013).
45. Lee, G. et al. Fully reduced HMGB1 accelerates the regeneration of multiple tissues by transitioning stem cells to Galert. *Proc. Natl Acad. Sci. USA* **115**, E4463–E4472 (2018).
46. Schiraldi, M. et al. HMGB1 promotes recruitment of inflammatory cells to damaged tissues by forming a complex with CXCL12 and signaling via CXCR4. *J. Exp. Med.* **209**, 551–563 (2012).
47. Jones, R. J. et al. Lenalidomide, Thalidomide, and Pomalidomide Reactivate the Epstein-Barr Virus Lytic Cycle Through Phosphoinositide 3-kinase Signaling and Ikaros Expression. *Clin. Cancer Res.* **22**, 4901 (2016).
48. Martinez-Høyer, S. & Karsan, A. Mechanisms of lenalidomide sensitivity and resistance. *Exp. Hematol.* **91**, 22–31 (2020).
49. Supper, E. et al. Cut-like homeobox 1 (CUX1) tumor suppressor gene haploinsufficiency induces apoptosis evasion to sustain myeloid leukemia. *Nat. Commun.* **12**, 2482 (2021).
50. Nguyen, N. H. et al. Genome-wide CRISPR/Cas9 screen identifies etoposide response modulators associated with clinical outcomes in pediatric AML. *Blood Adv.* **7**, 1769 (2023).
51. Han, C. R., Park, S. & Cheng, S. Y. NCOR1 modulates erythroid disorders caused by mutations of thyroid hormone receptor  $\alpha 1$ . *Sci. Rep.* **7**, 18080 (2017).
52. Ziegenhain, C. et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **65**, 631–643.e4 (2017).
53. Nilsson, L. et al. The molecular signature of MDS stem cells supports a stem-cell origin of 5q-myelodysplastic syndromes. *Blood* **110**, 3005–3014 (2007).
54. Calvi, L. M., Li, A. J. & Becker, M. W. What is the role of the microenvironment in MDS? *Best. Pr. Res. Clin. Haematol.* **32**, 101113 (2019).
55. Li, A. J. & Calvi, L. M. The microenvironment in myelodysplastic syndromes: Niche-mediated disease initiation and progression. *Exp. Hematol.* **55**, 3–18 (2017).
56. Fang, J. et al. Myeloid malignancies with chromosome 5q deletions acquire a dependency on an intrachromosomal NF- $\kappa$ B gene network. *Cell Rep.* **8**, 1328 (2014).
57. Liu, T. X. et al. Chromosome 5q deletion and epigenetic suppression of the gene encoding alpha-catenin (CTNNA1) in myeloid cell transformation. *Nat. Med.* **13**, 78–83 (2007).
58. Koc, E. C. et al. A new face on apoptosis: Death-associated protein 3 and PDCD9 are mitochondrial ribosomal proteins. *FEBS Lett.* **492**, 166–170 (2001).
59. Kim, H. J., Maiti, P. & Barrientos, A. Mitochondrial Ribosomes in Cancer. *Semin. Cancer Biol.* **47**, 67 (2017).
60. Talati, C., Sallman, D. & List, A. Lenalidomide: Myelodysplastic syndromes with del(5q) and beyond. *Semin. Hematol.* **54**, 159–166 (2017).
61. Pellagatti, A. et al. Lenalidomide inhibits the malignant clone and up-regulates the SPARC gene mapping to the commonly deleted region in 5q- syndrome patients. *Proc. Natl Acad. Sci. USA* **104**, 11406–11411 (2007).

62. Görgün, G. et al. Lenalidomide Enhances Immune Checkpoint Blockade Induced Immune Response in Multiple Myeloma. *Clin. Cancer Res.* **21**, 4607 (2015).
63. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
64. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281–291.e9 (2019).
65. Ainciburu, M. et al. Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single-cell resolution. *Elife* **12**, 79363 (2023).
66. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
67. Finak, G. et al. MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 1–13 (2015).
68. Serrano, G. et al. Single-cell transcriptional profile of CD34+ hematopoietic progenitor cells from del(5q) Myelodysplastic Syndromes and impact of lenalidomide. *GitHub* <https://doi.org/10.5281/ZENODO.10983466> (2024).

## Acknowledgements

This work was supported by the Instituto de Salud Carlos III and co-financed by ERDF A way of making Europe (PI20/01308, PI23/00516, PI19/00726, PI22/01044, PI 20/00531), CIBERONC (CB16/12/00489), RICORS TERA V (RD21/0017/0009), ERA-PerMed JTC 2019 (MEET-AML), MICCIN/AEI/10.13039/501100011033 and [RTI2018-101708-A-I00], Gobierno de Navarra (AGATA 0011-1411-2020-000010/0011-1411-2020-000011), Fundación La Caixa (GR-NET NORMAL-HIT HR20-00871), Cancer Research UK [C355/A26819], FC AECC and AIRC under the Accelerator Award Program, TRANSCAN (AECC AC 18/000002 and ISCI11), 2021 SGR 00560 (GRC) Generalitat de Catalunya; economical support from CERCA Programme/Generalitat de Catalunya, Fundació Internacional Josep Carreras and Stiftung Carreras Foundation. AECC award from the Fundación AECC (INVEST19059EZPO) (T.E). H2020 Marie S. Curie IF Action, European Commission, Grant Agreement No. 898356, a Ramon y Cajal contract RYC2021-033127-I MCIN/AEI/10.13039/501100011033 (M.H). Partially funded by a Ramon y Cajal contract RYC2019-028578-I, a Gipuzkoa Fellows grant 2022-FELL-000003-01, and grant PID2021-126718OA-I00 funded by MCIN/AEI/10.13039/501100011033 (I.O.) Supported by PhD fellowships from Gobierno de Navarra (0011-0537-2019-000001) (N.B.), and (0011-0537-2020-000022) (A.D.-M.); a PhD fellowship from Ministerio de Ciencia, Innovación y Universidades (FPU18/05488) (M.A.). Funded by Sara Borrell grant (CD22/00027) from the Instituto Carlos III and NextGenerationEU (A.U.-A) and personal técnico de Apoyo fellowship from the Ministry of Science and Innovation (PTA2021-020262) (P.A.-R). We particularly acknowledge the patients and healthy donors for their participation in this study, and the Biobank of the University of Navarra for its collaboration.

## Author contributions

G.S., N.B., A.D.-M., F.P., T.E., and M.H. conceived and designed the research studies; G.S., A.D.-M. and C.R.-R. performed in silico analysis of transcriptomic data; G.S., N.B., and A.D.-M. performed statistical analysis; G.S., N.B., A.D.-M., P.G.O. S.H.-D., A.A.-P., A.U.-A., B.A., C.R.-R., and M.A. analyzed and interpreted data; A.V.-Z., P.S.-M., P.A.-R., J.L.-E., A.A.-P., P.A., O.C., T.J., A.M., J.M., M.D.-C., D.V., and F.S. provided clinical samples and data; A.V.-Z., P.S.-M., P.A.-R., A.U.-A., B.A., M.A. provided technical assistance; S.H.-D., and A.A.-P. provided clinical advice; T.E., I.O., M.H., and F.P. were responsible for research supervision, coordination, and strategy. N.B., A.D.-M., G.S., P.G.-O., F.P., T.E., and M.H. wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-49529-x>.

**Correspondence** and requests for materials should be addressed to Felipe Prósper, Teresa Ezponda or Mikel Hernaez.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Guillermo Serrano <sup>1,2,12</sup>, Nerea Berastegui <sup>3,4,12</sup>, Aintzane Díaz-Mazkarian<sup>1,3,4,12</sup>, Paula García-Olloqui <sup>3,4</sup>, Carmen Rodríguez-Res<sup>1</sup>, Sofia Huerga-Dominguez<sup>5</sup>, Marina Ainciburu <sup>3,4</sup>, Amaia Vilas-Zornoza<sup>3,4</sup>, Patxi San Martín-Uriz <sup>3</sup>, Paula Aguirre-Ruiz<sup>3</sup>, Asier Ullate-Agote<sup>3</sup>, Beñat Ariceta <sup>3,4</sup>, Jose-Maria Lamo-Espinosa<sup>6</sup>, Pamela Acha<sup>7,8</sup>, Oriol Calvete<sup>7</sup>, Tamara Jimenez <sup>4,9</sup>, Antonieta Molero <sup>8</sup>, Maria Julia Montoro<sup>8</sup>, Maria Díez-Campelo <sup>4,9</sup>, David Valcarcel<sup>8</sup>, Francisco Solé <sup>7</sup>, Ana Alfonso-Pierola <sup>4,5</sup>, Idoia Ochoa <sup>10,11</sup>, Felipe Prósper <sup>3,4,5,13</sup> ✉, Teresa Ezponda <sup>3,4,13</sup> ✉ & Mikel Hernaez <sup>1,4,10,13</sup> ✉

<sup>1</sup>Computational Biology Program CIMA-Universidad de Navarra, Cancer Center Clínica Universidad de Navarra (CCUN), IdISNA, Pamplona, Spain. <sup>2</sup>Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. <sup>3</sup>Hematology-

Oncology Program, CIMA, Cancer Center Clínica Universidad de Navarra (CCUN), IdiSNA, Pamplona, Spain. <sup>4</sup>Centro de Investigación Biomédica en Red de Cáncer, CIBERONC, Madrid, Spain. <sup>5</sup>Hematology and Cell Therapy Service, Cancer Center Clínica Universidad de Navarra (CCUN), IdiSNA, Pamplona, Spain. <sup>6</sup>Department of Orthopedics, Clínica Universidad de Navarra, Pamplona, Spain. <sup>7</sup>MDS Research Group, Josep Carreras Leukaemia Research Institut, Universitat Autònoma de Barcelona, Barcelona, Spain. <sup>8</sup>Service of Hematology, Hospital Universitari Vall d'Hebron, Barcelona; Vall d'Hebron Instituto de Oncología (VHIO), Barcelona, Spain. <sup>9</sup>Department of Hematology, Hospital Universitario de Salamanca-IBSAL, Salamanca, Spain. <sup>10</sup>Instituto de Ciencia de los Datos e Inteligencia Artificial (DATAI), University of Navarra, Pamplona, Spain. <sup>11</sup>Department of Electrical and Electronics engineering, School of Engineering (Tecnun), University of Navarra, Donostia, Spain. <sup>12</sup>These authors contributed equally: Guillermo Serrano, Nerea Berastegui, Aintzane Díaz-Mazkaran. <sup>13</sup>These authors jointly supervised this work: Felipe Prósper, Teresa Ezponda, Mikel Hernaez. ✉ e-mail: [fprosper@unav.es](mailto:fprosper@unav.es); [tezponda@unav.es](mailto:tezponda@unav.es); [mhernaez@unav.es](mailto:mhernaez@unav.es)