medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Restored Hematopoietic Stem Cell Transcriptional Program in Fanconi Anemia Patients Following Gene Therapy

AUTHORS

Miren Lasaga[1]*, Paula Río[2,3,4]*, Amaia Vilas-Zornoza[5,6,]*, Nuria Planell[1] , Susana Navarro [2,3,4], Diego Alignani[7], Josune Zubicaray[8], Jonathan D. Schwartz[9], Julián Sevilla[3,8], Marina Ainciburi[5,6], Asier Ullate-Agote[5,6], Felipe Prosper[5,6,&], David Gomez-Cabrero[1,10,11,&] and Juan A. Bueren[2,3,4,&]

AFFILIATIONS:

[1]Navarrabiomed, Complejo Hospitalario de Navarra (CHN), Universidad Pública de Navarra (UPNA), IdiSNA, Pamplona, Spain.

[2]Hematopoietic Innovative Therapies Division, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT)

[3]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Madrid, Spain.

[4]Instituto de Investigaciones Sanitarias. Fundación Jiménez Díaz, Madrid, Spain.

[5]Area de Hemato-Oncología, Centro de Investigación Médica Aplicada (CIMA), Universidad de Navarra, IDISNA, Pamplona, Spain.

[6]Centro de Investigación Biomédica en Red de Cáncer, CIBERONC.

[7]Flow Cytometry Core, CIMA, Universidad de Navarra, Pamplona, Spain.

[8]Hemoterapia y Hematología Pediátrica, Fundación para la Investigación Biomédica, Hospital Infantil Universitario Niño Jesús, Madrid, Spain.

[9]Rocket Pharmaceuticals Inc, New York, NY, USA

[10]Biological and Environmental Sciences and Engineering Division (BESE), KingAbdullah University of Science and Technology KAUST, Thuwal, 23955, Saudi Arabia.

[11]Mucosal & Salivary Biology Division King's College London Dental Institute, London, United Kingdom.

*Equal first author contribution; [&]Senior authors with equal contribution

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

SUMMARY PARAGRAPH (240/200 words)

Fanconi anemia (FA) is a monogenic inherited disease associated with mutations in genes that encode for proteins participating in the FA/BRCA DNA repair pathway. Mutations in FA genes result in chromosomal instability and cell death, leading to cancer risks and progressive cell mortality, most notably in hematopoietic stem and progenitor cells (HSPC). Recently, we showed the first clinical evidence that gene therapy confers engraftment and proliferative advantage of gene-corrected HSPCs in FA patients[1]. Despite this and many other gene therapy advances, the question of whether the molecular pathways affected in monogenic diseases can be reverted by lentiviral-mediated gene therapy has never been addressed. This is even more challenging in DNA repair syndromes such as FA since in these cases, transcriptional defects in affected cells might not be restored due to DNA damage accumulated prior to gene therapy. Using single-cell RNA sequencing in HSPCs from FA-A patients previously treated by ex vivo gene therapy, we demonstrate that lentiviral-mediated gene therapy prior to severe bone marrow failure not only restores the expression of the defective gene, but also induces a long-term correction of the transcriptional program in FA HSPCs, which then acquire a signature characteristic of healthy HSPCs. Our results reveal new molecular evidence showing the potential of gene therapy to fully rescue phenotypic defects in FA, a devastating HSPC disease characterized by defective DNA repair.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

TEXT

The potential of gene therapy (GT) to correct a variety of human diseases -- in particular human hematopoietic disorders including primary immunodeficiencies and hemoglobinopathies -- has been well demonstrated in previous clinical studies [2]. Additionally, we have recently shown that lentiviral-mediated GT can also enable correction of more complex diseases, such as Fanconi anemia (FA), in which marked phenotypic defects are evident in primitive self-renewing HSCs. In that clinical trial, the infusion of corrected $CD34^+$ cells in non-conditioned FA patients resulted in a marked HSPC proliferative advantage, evidenced by progressive increases in the percentage of gene-corrected cells both in bone marrow (BM) and peripheral blood (PB), with concomitant increases in hematopoietic cell resistance to DNA cross-linking agents[1].

Despite significant advances in the field of GT, the full potential of this therapeutic modality will depend on its capacity to reestablish the molecular circuits altered in diseased cells. This is even more relevant in diseases such as FA, characterized by DNA repair defects, and thus associated with progressive accumulation of DNA damage. In patients treated in the FANCOLEN-I clinical trial, the presence of a chimeric population of corrected and uncorrected HSPCs, none of them exposed to any conditioning agent, allowed us to comparatively investigate the molecular pathways in each of these coexistent populations within the same patient. In the current study we demonstrate that in addition to the reversal of the characteristic hypersensitivity of FA HSPCs to genotoxic agents, lentiviral-mediated GT facilitates transcriptional reprogramming of FA HSPCs during the years following GT, such that these cells resemble the molecular circuits of healthy HSPCs.

Gene therapy modifies the HSPC transcriptional program in FA patients

Bone marrow (BM) $CD34^+$ cells from three FA-A patients (FA-02004, FA-02006, and FA-02008) who had received gene therapy 4, 3, and 2 years previously, respectively, were purified and analyzed by single-cell RNA sequencing (scRNAseq), as described in **Fig 1A**. qPCR analyses from these samples showed the presence of 0.45, 0.26, and 0.29 vector copy numbers per $CD34^+$ purified cell, respectively (see Materials and methods), revealing the chimeric nature of these HSPC populations, with estimated average values of 45%, 26%, and 29%, respectively, based on the insertion of 1 copy per corrected cell[1].

$CD34^+$ cells from FA patients were classified according to previously described transcriptional profiles[3,4], which identified twelve different HSPC clusters that corresponded to primitive HSCs and more committed lympho-hematopoietic progenitor cells **(**See analyses

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

from patient FA-02006 in **Fig. 1B;** and patients FA-02004 and FA-02008 in **Fig. S1A, B)**. To investigate the impact of GT on the transcriptional program of FA HSPCs, CD34$^+$ cell populations from GT-treated FA patients were classified as FANCA$^+$ and FANCA$^-$ cells, based on the expression of FANCA by scRNAseq. As shown in **Fig. 1C** and **1D**, the presence of FANCA$^+$ cells was evident in all HSPC populations represented with at least 30 cells per cluster. No significant differences in the proportion of FANCA$^+$ cells present in these subpopulations were observed among the three patients (**Table S1**). Additionally, a wide range of FANCA expression was observed in the different analyzed samples. Only in specific subpopulations were significant differences in FANCA expression levels observed between individuals **(Fig 1E)**.

To investigate the impact of ectopic FANCA expression on the transcriptional program of HSPCs, we performed a differential expression analysis between *FANCA*$^+$ and *FANCA*$^-$ cells in the different HSPC subpopulations present in each GT-treated patient. As shown in **Fig 1F,** the ectopic expression of FANCA induced the upregulation of a high number of genes compared to the number of genes that were downregulated. As expected, genes with statistically differential expression corresponded to subpopulations with a higher representation (i.e., monocytic CD34$^+$ cells; correlation between number of differentially expressed genes and number of cells: p-value< 0.001; dark red or blue bars in **Fig. 1F**)**.** Remarkably, most of the upregulated and downregulated genes in FANCA$^+$ vs. FANCA$^-$ cells showed the same expression pattern in each of the analyzed patients (see fourth bar at the bottom of each patient's bars in **Fig. 1F**). To select the genes with a robust differential expression between corrected FANCA$^+$ versus FANCA$^-$ HSPCs, we considered those genes that showed a significant differential expression (abs(logFC)>0.25 and adjusted p-value<0.05) in at least one cell type in at least two patients. In addition, changes in gene expression should be in the same direction in all the three patients. Based on these criteria, a total number of 382 differentially expressed genes were identified (See **Table S2**).

Taken together, data obtained in these analyses indicate that the ectopic expression of FANCA induced reproducible changes in the gene expression program of HSPCs from FA-A patients previously infused with genetically-corrected autologous CD34$^+$ cells.

Gene therapy reprograms the transcriptional signature of FA HSPCs towards the transcriptional profile characteristic of healthy HSPCs

Next, we investigated whether changes in the transcriptional program of gene-corrected FA-A HSPCs resembled profiles characteristic of healthy HSPCs. To this end, scRNAseq data from GT-treated FA CD34$^+$ cells was compared to that obtained from a healthy donor (HD)

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

CD34$^+$ cells. As shown in **Fig. 2A** ectopic expression of FANCA (FANCA$^+$ cells) was detected in every CD34$^+$ cell cluster. In **Fig. 2B** we confirmed the heterogeneity of FANCA expression levels in the different HSPC subpopulations, either from GT-treated FA patients or HDs. Despite this heterogeneity, in eight out of the twelve HSPC subpopulations, physiological levels of FANCA mRNA were significantly higher in HD samples compared to the ectopic expression of FANCA in corrected FA HSPCs (**Fig. 2B)**.

To investigate if GT reprogrammed the transcriptional signature of FA HSPCs towards that corresponding to HD HSPCs, we conducted additional studies analyzing the 382 genes that showed robust changes due to the ectopic expression of FANCA in FA HSPCs (See **Table S2**). In these new studies, three different comparisons were performed in each of the twelve HSPC subpopulations: FANCA$^+$ *vs.* FANCA$^-$ HSPCs from GT-treated FA patients (GT-FA HSPCs); HD HSPCs *vs.* FANCA$^-$ GT-FA HSPCs; and finally, HD HSPCs *vs.* FANCA$^+$ GT-FA HSPCs **(Fig. 2C)**. As observed in analyses of **Fig 1**, significant differences (dark red or blue spots in Fig. 2C) were more evident in HSPC subpopulations with a higher representation (p-value<0.001).

In the comparison of FANCA$^+$ *vs.* FANCA$^-$ HSPCs (**Fig. 2C** first column), several differentially expressed genes relevant to FA are shown in the right side of the figure. Interestingly, genes involved in functions such as DNA repair or cell cycle were upregulated in corrected FA-HSPCs (FANCA$^+$ vs FANCA$^-$ cells). On the contrary, genes encoding for three HLA molecules were down-regulated in corrected FA-HSPCs. In the second column of **Fig 2C**, the gene expression pattern of HSPCs from HDs was compared with that corresponding to uncorrected HSPCs from GT-treated patients (HD vs GT-FANCA$^-$). Strikingly, most of the transcriptional changes observed between these populations resembled changes that were seen between corrected and uncorrected FA HSPCs (first column in **Fig 2C**) (see **Table S3** for detailed analyses per cell type using two complementary statistical analyses). Finally, in contrast to the above-mentioned observations, the comparison of the transcriptional program of HD HSPCs vs FANCA$^+$ HSPCs from GT-treated FA patients (third column in **Fig 2C**) showed limited changes, most of which were not significant, and not related to differences noted when HD or corrected FA HSPCs were compared with uncorrected FA HSPCs. Overall, these studies demonstrate that lentiviral-mediated GT reverts the gene expression program of FA HSPCs, which then acquire an expression pattern that resembles the signature characteristic of healthy HSPCs.

Lentiviral-mediated gene therapy reverts molecular pathways characteristic of FA HSPCs

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Having observed that the ectopic expression of FANCA in FA HSPCs reverts the transcriptional signature of these cells to resemble healthy HSPCs, we next performed a gene-set enrichment analysis to determine changes in relevant pathways associated with FA, including those related to DNA damage response and repair, cell cycle checkpoint, cell aging, and telomerase maintenance (see Methods). As shown in **Fig. 3A** and **Table S4**, an upregulated expression of several of these pathways was observed when FANCA$^+$ and FANCA$^-$ HSPCs present in each of the GT-treated patients were compared (See first paired columns in **Fig. 3A**; GT-FANCA$^+$ vs GT-FANCA$^-$). Moreover, an almost identical upregulation of these pathways was observed when HSPCs from HDs were compared with FANCA$^-$ HSPCs from GT-treated patients (HD vs GT-FANCA$^-$; second paired columns in **Fig 3A**).

A deeper comparative expression analysis of genes involved in the control of the cell cycle was then performed between FANCA$^+$ and FANCA$^-$ HSPCs from GT-treated patients (GT-FANCA$^+$ vs GT-FANCA$^-$; see external crowns in **Fig. 3B**). Similar comparisons were also made between HD HSPCs and FANCA$^-$ HSPCs from GT-treated patients (internal crowns in **Fig. 3B**). As shown, the ectopic expression of FANCA was associated with the downregulation of *TGF-β* in every patient, and of *p21* in two of the three GT-treated FA patients. On the other hand, a number of cyclins and minichromosome maintenance (MCM) genes were upregulated in corrected vs uncorrected FA HSPCs (FANCA$^+$ vs FANCA$^-$ HSPCs). Notably, when similar comparisons were performed between HD HSPCs and FANCA$^-$ HSPCs from GT-treated FA patients, almost identical changes in gene expression were observed (**Fig 3B)**, indicating that the behavior of cell cycle and DNA checkpoint pathways in corrected FA-A HSPCs resembled the pathways characteristic of healthy HSPCs.

Since previous studies have shown that p21 participates in the transcriptional repression of different genes of the FA/BRCA pathway[5,6], changes in the expression of FA/BRCA genes were also comparatively investigated in FANCA$^+$ and FANCA$^-$ HSPCs from GT-treated patients. As shown in **Fig 3C**, in addition to *FANCA*, several other genes participating in the FA/BRCA pathway, including UBE2T (FANCT), *USP1*, *FANCI*, *MHF*, *BRCA1 (FANCS)*, *BRCA2 (FANCD1)* and *RAD51C (FANCO)* and *RPA* were upregulated in FANCA$^+$ vs. FANCA$^-$ HSPCs. Again, many of these genes were also upregulated when HD HSPCs were compared with uncorrected FANCA$^-$ HSPCs.

Overall, the results obtained in this study demonstrate for the first time that GT - in particular lentiviral-mediated GT in FA patients - reverts the transcriptional program of diseased

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

HSPCs, which then display gene expression signatures and molecular pathways characteristic of healthy HSPCs.

DISCUSSION

Gene therapy has emerged as a safe and efficient therapeutic option for diverse monogenic disorders affecting the hematopoietic system (See review in[2]). In a previous study we showed that in addition to monogenic diseases affecting one or more lympho-hematopoietic lineages, GT can also correct the phenotype of severely affected HSPCs in patients with more complex diseases, such as FA[1]. In that clinical study, we demonstrated that the ectopic expression of FANCA reproducibly conferred engraftment and proliferative advantage of corrected HSCs in non-conditioned FA-A patients. Nonetheless, the question of whether corrected FA HSPCs acquire the transcriptional profile of healthy HSPCs was not addressed in that study, nor in any other monogenic disease GT trial.

By leveraging scRNAseq profiling, we have extended the conclusions from our previous study regarding the phenotypic correction of HSPCs in FA patients receiving GT. By comparing the transcriptional program of BM $CD34^+$ cells that ectopically express FANCA versus coexisting uncorrected HSPCs within the same patient, we are now able to demonstrate for the first time that GT not only corrects the expression of the defective gene (*FANCA* in this case) but also reprograms the molecular circuits of diseased FA-A HSPCs to acquire a normalized gene expression pattern.

Importantly, we show that several defective pathways implied in the FA cellular phenotype were significantly upregulated as a result of the ectopic expression of *FANCA* in FA-A HSPCs. Moreover, the restored expression of *FANCA* in these cells resulted in a decreased expression of two genes with particular implications in FA, *TGF-β* and *p21*. Increased mRNA levels of these two genes have been observed in HSPCs from FA patients and murine FA models, and are believed to contribute to the BMF characteristic of the disease[7,8]. Notably, previous studies have shown that p21 downregulates USP1 upon exposure to DNA damage, disrupting FANCD2/L mono-ubiquitination and nuclear foci formation[6]. Additional studies have also shown that several genes involved in the telomere biology and in the FA/BRCA pathway are downregulated via the p21/E2F4 pathway[5]. Our study now reveals that GT results in the upregulation of several genes associated with telomere biology and FA-associated pathways. Also of interest is the downregulated expression of HLA-A, HLA-C and HLA-E genes observed in corrected FA HSPCs. This can be accounted by documented

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

interferon-gamma overexpression in FA HSPCs[9,10], and this cytokine's transactivation of classical and non-classical HLA class I genes, such as the HLA-A, HLA-C and HLA-E genes[11,12]. Although it might be expected that p53 or Myc - which have been shown to be upregulated in FA patients[7,13] – should be also downregulated in corrected FA HSPCs, our data do not support this hypothesis. Reasons accounting for these apparent unexpected observations could be related either to the enhanced expression of cell cycle genes in corrected FA HSPCs, or to post-transcriptional regulatory mechanisms involved in p53 expression[7], and also to the moderate BMF status of the FA patients enrolled in the FANCOLEN I trial[1].

Taken together, our study demonstrates for the first time that GT reverts the transcriptional program and defective molecular pathways in corrected HSPCs. This observation has a particular impact in FA, and possibly in other HSC diseases also characterized by DNA repair defects, since here we demonstrate that GT conducted prior to BMF restores the molecular program of FA HSPCs towards a signature characteristic of healthy HSPCs.

METHODS

Patients and Healthy donors

FA patients included in this study, FA-02004, FA-02006 and FA-02008, were FA-A patients enrolled in the FANCOLEN-1 gene therapy trial (FANCOLEN-1; ClinicalTrials.gov, NCT03157804; European Clinical Trials Database, 2011-006100-12) and complied with all relevant ethical regulations and approved by the ethics committees at Hospital Vall d'Hebron in Barcelona and Hospital del Niño Jesús in Madrid.

For inclusion, patients had to have at least one hematological parameter above. Patients were infused with autologous $CD34^+$ cells after transduction with the therapeutic PGK-*FANCA*.Wpre* LV[14]. Previous data has shown that although no conditioning was given to these patients prior to cell infusion, a progressive engraftment of corrected cells took place over time, implying the presence of a chimeric population of corrected and uncorrected cells in their hematopoietic tissues[1]. Prior to BM aspiration, HDs (median age = 20 y/o) provided informed consent for their participation in this study, which was approved by the clinical research ethics committee of Clínica Universidad de Navarra.

Bone marrow cells

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Bone marrow cells from GT treated patients were obtained in the course of routine follow-up studies of the GT trial and as part of additional exploratory studies. Samples used in these studies were obtained 4 years (FA-02004), 3 years (FA-02006) and 2 years (FA-02008) after infusion of transduced $CD34^+$ cells, respectively. Patients FA-02004 and FA-02008 had been treated with eltrombopag to stimulate hematopoiesis 12 and 6 months prior to the evaluation of BM cells, respectively. In all instances samples were processed immediately after BM aspiration.

For the purification of $CD34^+$ cells, erythrocytes were lysed with ammonium chloride lysis solution (0.155 mmol/L NH_4Cl + 0.01 mmol/L $KHCO_3$ + 10^{-4} mmol/L EDTA), washed using PBS (Gibco) + 0.2%BSA (10%) + 2% PenStrep (Gibco) and stained using CD45 APC (clone 2D1; Biolegend) and CD34 PECY7 (clone 4H11; eBiosciences) for 30 min at 4ºC. DAPI was used at a concentration of 1 µg/mL as a viability marker. $CD34^+$ cells were then sorted in a BD INFLUX™ (BD Biosciences) or BD FACSAria II (BD Biosciencs) as previously shown[1]. Purified $CD34^+$ cells were directly used for scRNA seq analysis and small aliquots stored at -80ºC for vector copy number analysis.

Analysis of lentiviral vector copy numbers in BM $CD34^+$ cells

The number of proviral copies per $CD34^+$ cell (VCN/cell) was performed after genomic extraction of the DNA using the DNA easy blood and tissue kit (Qiagen) or by proteinase K lysis as previously described [1]. Duplex qPCR was conducted to detect the *Psi* sequence of the provirus and the *Albumin,* as a control gene. To amplify *Psi* sequence: *Psi* forward (*Psi*.F): 5' CAGGACTCGGCTTGCTGAAG 3' and *Psi* reverse (Psi.R): 5' TCCCCCGCTTAATACTGACG 3' primers were used and detected with the Taqman probe *Psi*.P FAM: 5'CGCACGGCAAGAGGCGAGG3'. To normalize to endogenous *Albumin*, specific primers for *Albumin* were used: *Alb* forward (*Alb*.F): 5´ GCTGTCATCTCTTGTGGGCTG 3´ and *Alb* reverse (*Alb*.R.): 5´ ACTCATGGGAGCTGCTGGTTC 3´ together with a Taqman probe *Alb*.P VIC: 5' CCTGTCATGCCCACACAAATCTCTCC 3'. qPCR was conducted in an Applied 7500 Fast Real Time PCR system (Thermo Fisher Scientific), as previously described[1].

Single-cell RNA-sequencing (scRNA-seq)

The transcriptome of BM $CD34^+$ cells was investigated using NEXTGEM Single Cell 3' Reagent Kits v3.1 (10X Genomics) according to the manufacturer's instructions. Between 2.000 and 6.000 $CD34^+$ cells were loaded at a concentration of 700-1,000 cells/µL on a

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Chromium Controller instrument (10X 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 used to sequence a cell-identifying 16 bp 10X barcode for each cell and a 12 bp Unique Molecular Identifier (UMI) for each transcript. Upon cell lysis, reverse transcription yielded full-length, barcoded cDNA. This cDNA was then released from the GEMs, PCR-amplified and purified with magnetic beads (SPRIselect, Beckman Coulter). Enzymatic Fragmentation and Size Selection was used to optimize cDNA size prior to library construction. Fragmented cDNA was then end-repaired, A-tailed and ligated to Illumina adaptors. A final PCR-amplification with barcoded primers allowed sample indexing. Library quality control and quantification was performed using Qubit 3.0 Fluorometer (Life Technologies) and Agilent's 4200 TapeStation System (Agilent), respectively. Sequencing was performed in a NextSeq500 (Illumina) (Read1: 28 cycles; Read 55 cycles; i7 index: 8 cycles) at an average depth of 20,000 reads/cell.

scRNA-seq: bioinformatics.

Data filtering and normalization: Sequenced libraries were demultiplexed, aligned to human transcriptome (hg38) and quantified using Cell Ranger (v_3.0.1). Ongoing analysis was conducted using Seurat (V_3.2.0)[15] in R (V_3.5.2)[16]. Quality control filters based on the number of detected genes, number of UMIs and percentage of mitochondrial UMIs were performed to each one of the samples. The thresholds were defined based on the distribution of the previously mentioned parameters and visual inspection of quality control scatter plots. After filtering of low quality cells, a total number of 720 (FA-02004), 1,438 (FA-02006), 1,995 (FA-02008), and 12,549 (Healthy donor) cells were retained.

Each single cell dataset was individually normalized, using the Normalize Seurat function. Feature counts for each cell were divided by the total counts for that cell and multiplied by the scale factor. This was then natural-log transformed. The data was regressed out by cell cycle stadium, number of features and number of counts. Non-linear dimensional reduction was performed (UMAP) to plot the data of each samples. PCA was defined as dimensional reduction to use in the UMAP graph. Each of the FA samples was integrated with the healthy donor sample.

Cell annotation**:** For cell annotation, we use the annotation conducted in three additional human samples of healthy young individuals (3YI). The isolation protocol of 3YI includes the cell types in the 3 FA patients and the HD: Ficoll-Paque Plus (GE healthcare) density gradient centrifugation and stained using CD34 (clone 8G12; BD bioscience) CD64 (clone 10.1; Biolegend) CD19 (clone SJ25C1; Biolegend) CD10 (clone HI10A; Biolegend) CD3

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(clone OKT3; BIolegend) CD36 (clone CLB-IVC7; Sanquin Plesmanlaan) CD61 (clone RUU-PL7F12; BD bioscience) for 15 min at RT. And finally, CD34+ CD64- CD19- CD10- CD3- CD36+CD61+ cells were then sorted in a BD FACSAria II (BD Biosciencs) as previously shown. The data-analysis processing of those samples was conducted as the protocol describe for FA samples. Next, we performed unsupervised clustering with the Louvain algorithm as implemented in Seurat[17]. We tested several resolution values and assessed the results by calculating the average silhouette for each cluster. We determined the cluster markers using the Seurat function FindAllMarkers, with the MAST method. Finally, we annotated the clusters by manually inspecting the most specific markers and looking for curated markers in the literature. Using the robust annotation conducted in 3YI, the "label transfer function" from Seurat was used to annotate the three FA and the HD samples[15]. It is important to note that while the annotation of the 3YI is valid for the annotation of FA and HD samples, we decided not to include the 3YI samples in the analysis as the cell proportions may be different based on the isolation protocol.

Differential expression analysis: The differential expression analysis was conducted using FindMarkers function in the Seurat package. Genes were considered differentially expressed if |logFC|>0.25 and adjusted p-value<0.05.

GSEA analysis: Gene set enrichment analyses were conducted using ClusterProfiller (version 3.10.1)[17] in R[16]. The normalized data of each sample and cell type was ranked by the logFC value and the analysis was run comparing our data with GO biological processes. A gene set was considered significantly enriched if GO adjusted p-value<0.05.

Pathway visualization

After GSEA analysis two core pathways were selected (Cell cycle and FA/BRCA) for visualization purpose using Cytoscape (version 3.8.2.)[18]. The values of logFC were for each one of the samples and different contrast using omic visualizer package. The pathways are shown as imported in the Cytoscape package; for two genes an alternative gene symbol is shown (MHF, RPA).

Basic statistical tests

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Proportion test: In each FA sample and cell type, a two-proportion statistical test was conducted to investigate significant differences in cell type proportion between FANCA$^+$ and FANCA$^-$ cells in each CD34$^+$ subpopulation.

Annova test: We conducted a two-tail ANOVA test to investigate the differences of FANCA expression between therapy treated patients among the FANCA$^+$ set for each each CD34$^+$ cell subpopulation.

Wilcoxon test: The comparison of the expression of the FANCA gene between FANCA$^+$ cells in FA integrated samples and HDs was performed using a two-sided Wilcoxon test for each CD34$^+$ cell subpopulation. In HD, only cells with >0 FANCA gene expression value were considered.

Binomial Test: We conducted a binomial test to investigate if the shared directionality of changes for two contrasts, "FANCA$^+$ vs FANCA$^-$" and "HD vs FANCA$^+$", was significantly over-represented. To this end, the genes sharing the same directionality for both contrasts were classified as 1, and 0 otherwise; the binomial was conducted considering a probability of 0.5, number of experiments of 382 and a one-tail p-value associated to values larger than the observed. The analysis was conducted separately for each cell type and sample.

Binary Correlation: We conducted a binary correlation analysis between the directionality of the same genes in the two contrasts, "FANCA$^+$ vs FANCA$^-$" and "HD vs FANCA$^+$". All the upregulated genes were classified as 1 and the downregulated as 0. A binary correlation test was conducted using R.

General considerations: To perform all the statistical tests R[12] was used. In all the cases the multiple testing was addressed Bonferroni[15], and significant results were considered if adjusted p-value <0.05.

ACKNOWLEDGEMENTS

The authors thank Ramón García-Escudero for careful reading of the manuscript and helpful discussions. The authors also thank A. de la Cal for coordinating the delivery of BM samples from patients with FA. The authors are also indebted to the patients with FA, their families and clinicians from the Fundación Anemia de Fanconi. This work was supported by grants from the European Commission's Seventh Framework Program (HEALTH-F5-2012-305421) to the EUROFANCOLEN Consortium J.B. and J.Se; from Ministerio de Economía, Comercio y Competitividad and Fondo Europeo de Desarrollo Regional (RTI2018-097125-B-I00 to

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

P.R. and S.N.); from Gobierno de Navarra, Departamento de Desarrollo Económico y Empresarial (AGATA 0011-1411-2020-000010); from Instituto de Salud Carlos III (ISCIII) and co-financed by FEDER CIBERONC (CB16/12/00489 and CB16/12/00225); Redes de Investigación Cooperativa (TERCEL RD16/0011/0005) and from Consejería de Educación, Juventud y Deporte de la Comunidad de Madrid (AvanCell Project; B2017/BMD3692).. M.A. was supported by a PhD Fellowship from Ministerio de Ciencia, Innovación y Universidades (FPU18/05488). CIBERER and CIBERONC are initiatives of the Instituto de Salud Carlos III and Fondo Europeo de Desarrollo Regional.

Author contributions

M.L., P.R., A.V-Z., N.P., S.N., D.A, M.A, A.U-A., D-G.C performed the experimental studies and analyzed the data. J.S. and J.Z. provided critical materials. M.L., P.R., D-G.C, F.P. J.A.B wrote the manuscript. J.D.S. reviewed the manuscript. P.R., S.N., D-G.C, F.P. J.A.B designed the study. All authors discussed the results and contributed to the preparation of the manuscript.

Competing interests

The Hematopoietic Innovative Therapies Division at CIEMAT receives funding and has licensed the PGK-FANCA-WPRE* lentiviral vector to Rocket Pharmaceuticals. J.A.B. and J.S. are consultants for Rocket Pharmaceuticals. J.D.S. is an employee of Rocket Pharmaceuticals. P.R., S.N., J.S and J.A.B. are inventors on patents filed by CIEMAT, CIBERER and Fundación Jiménez Díaz, and may be entitled to receive financial benefits from the licensing of such patents.

Data availability

The authors declare that the data supporting the findings of this study will be made available before the manuscript publication.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

FIGURE LEGENDS

Figure 1. scRNAseq analysis of corrected and uncorrected hematopoietic stem and progenitor cells from FA-A patients 2-4 years after lentiviral-mediated gene therapy.

(A) Experimental design: Three FA-A patients that had been treated by ex vivo lentiviral-mediated gene therapy in the absence of conditioning were included in this study. At 2-4 years post-gene therapy, these patients harbored a chimeric population of corrected and uncorrected hematopoietic stem and progenitor cells (HSPCs) in their bone marrow (BM). Aliquots of BM $CD34^+$ cells from these patients were sorted and processed for single-cell RNA-seq. Bioinformatic analyses were then conducted to comparatively investigate changes in the transcriptional program of corrected and uncorrected HSPCs coexisting in each of the three gene therapy treated patients. **(B)** UMAP plot showing the clustering analysis for BM $CD34^+$ cells from a FA-A patient previously treated by gene therapy (FA-02006 patient as an example; see **Fig. S1** for patients FA-02004 and FA-02008). A total of 12 clusters were identified, spanning the different HSPC subpopulations. Identified clusters include a HSC cluster (hematopoietic stem cell; brown). Clusters with megakaryocytic-erythroid identity include MEP (erythroid-megakaryocyte progenitor; purple), Erythroid (erythroid progenitor; pink), and Basophils (basophil progenitor; light pink). Clusters with lympho-myeloid identity include LMPP (lymphoid-primed multipotent progenitor; light blue), Cycling-LMPP (blue), CLP (common lymphoid progenitor; orange), GMP1 and GMP2 (granulocyte-monocyte progenitor; light green and green), Monocytes (monocyte progenitor; red), DC (dendritic cell progenitor; nude), and PreB (B cells progenitor; light purple). **(C)** The same UMAP as shown in panel B but highlighting the distribution of $FANCA^+$ cells (FANCA mRNA detectable; red) versus $FANCA^-$ cells (FANCA mRNA levels are below detection limit; blue). **(D)** Barplot showing the total number of cells in the different HSPCs corresponding to the three gene therapy treated patients. In each case, the number of $FANCA^+$ (red) and $FANCA^-$ (blue) cells is shown and the percentage represented by $FANCA^+$ is written in the top of the bar. **(E)** Boxplot of integrated and normalized FANCA gene expression in $FANCA^+$, depicted by cell type and FA individual (n=3). **(F)** Barplot representation of the differential expression analysis between $FANCA^+$ and $FANCA^-$ for each FA patient (n=3) and for each of the HSPC populations. In each case, the number of upregulated and downregulated genes is shown. Upregulated genes were defined as logFC>0.25 (orange), upregulated and significant genes were defined as a logFC>0.25 and adjusted p-value<0.05 (red), downregulated genes were defined as a logFC<-0.25 (light blue) and downregulated and significant genes were define as logFC<-0.25 and adjusted p-value<0.05 (dark blue). The number of genes with the same

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

behavior (upregulated or downregulated) in the three individuals is shown below each HSPCs (labeled as "*shared*").

Figure 2. Comparison of the gene expression signature corresponding to corrected and uncorrected hematopoietic stem and progenitor cells coexisting in gene therapy treated FA patients and healthy individuals.

(A) Left larger panel: UMAP plot showing the clustering analysis of $CD34^+$ BM cells after integration of data from a gene therapy treated FA-A patient and a healthy donor (FA-02006 is included as a representative example; see Fig. S2 for the rest of individuals). **Right smaller panel**: the same UMAP as shown in left panel but highlighting the distribution of HD HSPCs (yellow), $FANCA^+$ HSPCs (red) and $FANCA^-$ HSPCs (blue). Clusters identification as in Figure 1. **(B)** Boxplot representation of normalized single-cell FANCA expression of $FANCA^+$ cells (red) and HD cells (yellow) by HSPC cluster. For each cell type, differences in the expression levels between the ectopic expression of FANCA from corrected FA $CD34^+$ cells and expression levels corresponding to HD $CD34^+$ cells are shown (*adjusted p-value <0.05; **adjusted p-value < 0.01) **(C)** The figure shows the results associated to three differential expression contrasts: $FANCA^+$ vs $FANCA^-$ HSPCs from GT-treated FA patients; HD HSPCs vs $FANCA^-$ HSPCs from GT-treated patients; and HD HSPCs vs $FANCA^+$ HSPCs from GT-treated patients. The second row shows the twelve different $CD34^+$ cell types, and the third one the sample identification from each of the three GT-treated patients. Up-regulated genes (logFC>0) are shown in orange and those with significant upregulation in red (logFC>0.25 and adjusted p-value<0.05). Down-regulated genes (logFC<0) are shown in light blue and those with significant downregulation in dark blue (logFC<-0.25 and adjusted p-value<0.05). Unsupervised hierarchical clustering using Pearson distance and average linkage method was applied for gene classification. The genes included in the heatmap are those that for at least one cell type are identified as differentially expressed (abs(logFC)>0.25 and adjusted p-value<0.05) in "at least two patients", and "showing the same direction of the change for the three patients", when considering the contrast $FANCA^+$ vs $FANCA^-$ (n=382; see the entire list in Table S2). FANCA was excluded from the list.

Figure 3. Restored gene expression pathways associated to the ectopic expression of FANCA in HSPCs from gene therapy treated FA patients.

(A) Comparative density plots showing the distribution of differential expression derived logFC for a selected set of pathways. For each patient, paired comparisons between $FANCA^+$ vs $FANCA^-$ HSPCs from GT-treated FA patients and HD HSPCs vs $FANCA^-$ HSPCs are shown. The complete name of the pathways with numbers are: (1) DNA damage

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

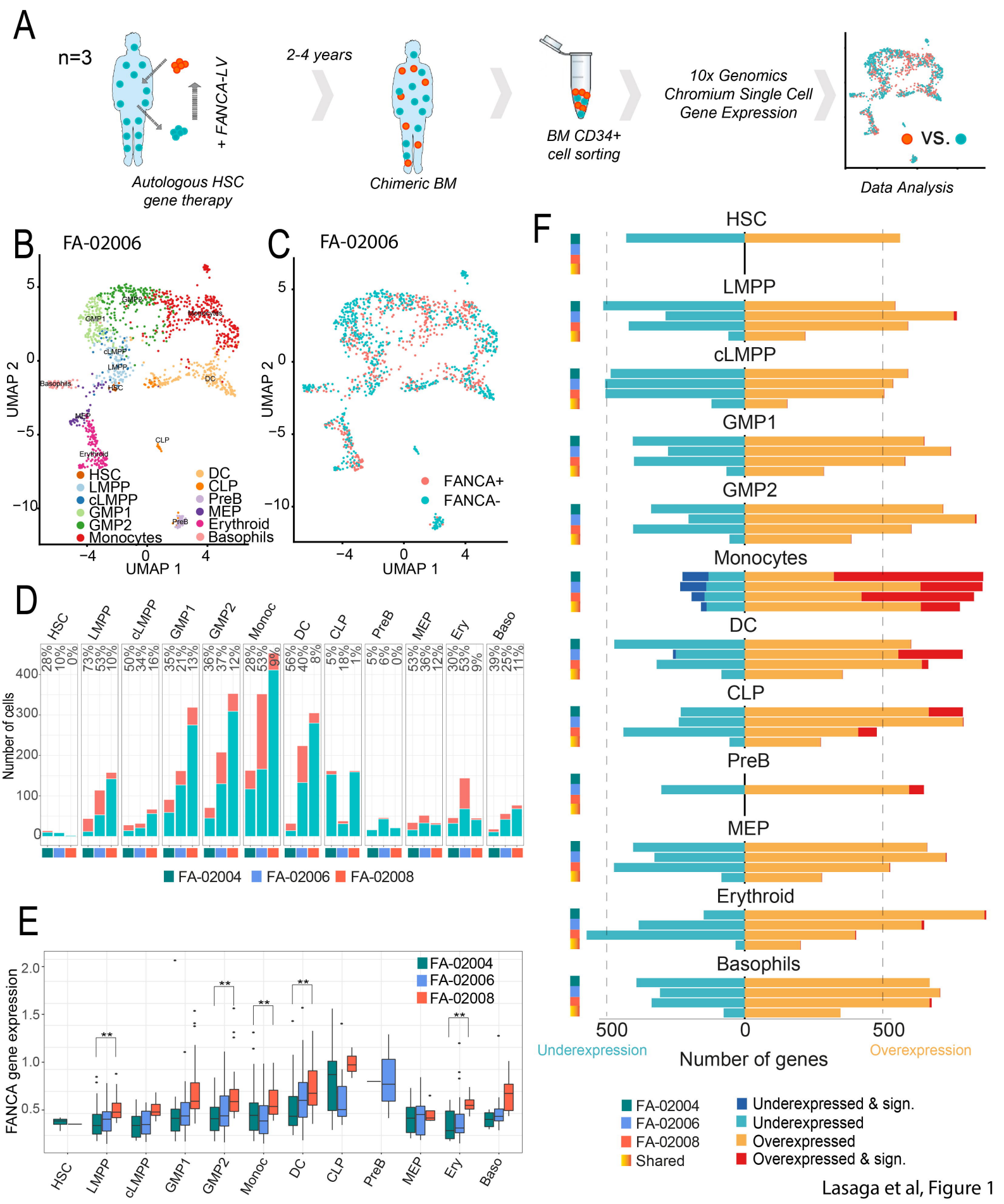
response, detection of DNA damage, (2) double−strand break repair via homologous recombination, (3) telomere maintenance via semi−conservative replication, (4) double−strand break repair via nonhomologous end joining, (5) positive regulation of G2/M transition of mitotic cell cycle, and (6) negative regulation of G2/M transition of mitotic cell cycle. **(B)** Upregulated and downregulated genes implicated in the control of the cell cycle. The figure represents the logFC associated to two contrasts: $FANCA^+$ vs $FANCA^-$ HSPCs from GT-treated FA patients (internal crowns), and HD HSPCs vs $FANCA^-$ HSPCs from GT-treated patients (external crowns). Each external and internal crown is divided in three parts, each of them representing one FA patient. Upregulated genes are shown in red and downregulated genes in blue. **(C)** Same representation as in B showing changes in the expression of genes participating in the FA/BRCA pathway.

REFERENCES

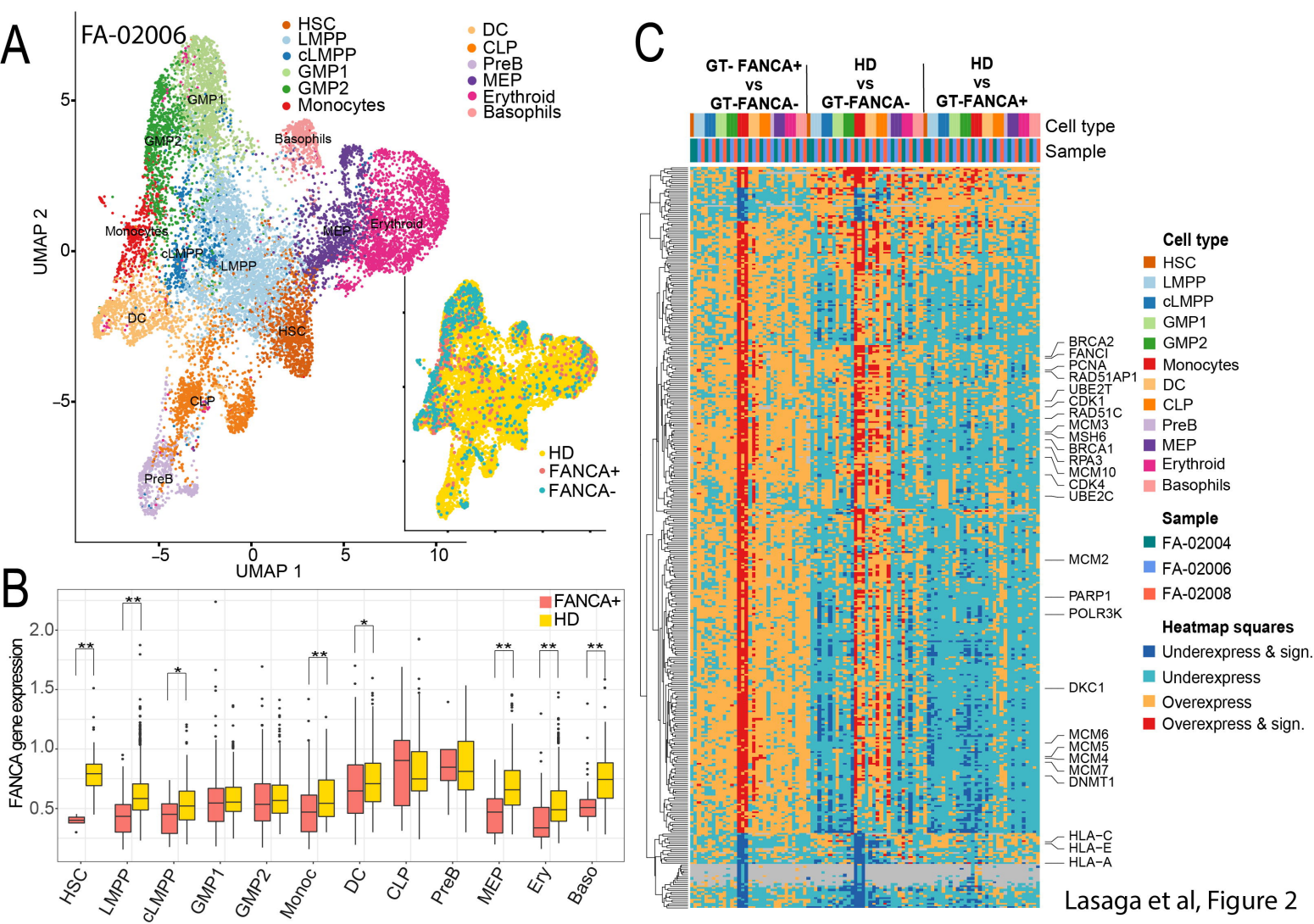
1. Rio, P.*, et al.* Successful engraftment of gene-corrected hematopoietic stem cells in non-conditioned patients with Fanconi anemia. *Nat Med* **25**, 1396-1401 (2019).
2. Ferrari, G., Thrasher, A.J. & Aiuti, A. Gene therapy using haematopoietic stem and progenitor cells. *Nat Rev Genet* **22**, 216-234 (2021).
3. Velten, L.*, et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**, 271-281 (2017).
4. Zheng, G.X.*, et al.* Massively parallel digital transcriptional profiling of single cells. *Nature communications* **8**, 14049 (2017).
5. Jaber, S., Toufektchan, E., Lejour, V., Bardot, B. & Toledo, F. p53 downregulates the Fanconi anaemia DNA repair pathway. *Nature communications* **7**, 11091 (2016).
6. Rego, M.A., Harney, J.A., Mauro, M., Shen, M. & Howlett, N.G. Regulation of the activation of the Fanconi anemia pathway by the p21 cyclin-dependent kinase inhibitor. *Oncogene* **31**, 366-375 (2012).
7. Ceccaldi, R.*, et al.* Bone marrow failure in Fanconi anemia is triggered by an exacerbated p53/p21 DNA damage response that impairs hematopoietic stem and progenitor cells. *Cell stem cell* **11**, 36-49 (2012).
8. Zhang, H.*, et al.* TGF-β Inhibition Rescues Hematopoietic Stem Cell Defects and Bone Marrow Failure in Fanconi Anemia. *Cell stem cell* **18**, 668-681 (2016).
9. Dufour, C.*, et al.* TNF-alpha and IFN-gamma are overexpressed in the bone marrow of Fanconi anemia patients and TNF-alpha suppresses erythropoiesis in vitro. *Blood* **102**, 2053-2059 (2003).
10. Rathbun, R.K.*, et al.* Interferon-gamma-induced apoptotic responses of Fanconi anemia group C hematopoietic progenitor cells involve caspase 8-dependent activation of caspase 3 family members. *Blood* **96**, 4204-4211 (2000).
11. Coupel, S.*, et al.* Expression and release of soluble HLA-E is an immunoregulatory feature of endothelial cell activation. *Blood* **109**, 2806-2814 (2007).
12. Gobin, S.J., van Zutphen, M., Woltman, A.M. & van den Elsen, P.J. Transactivation of classical and nonclassical HLA class I genes through the IFN-stimulated response element. *J Immunol* **163**, 1428-1434 (1999).
13. Rodríguez, A.*, et al.* MYC Promotes Bone Marrow Stem Cell Dysfunction in Fanconi Anemia. *Cell stem cell* **28**, 33-47.e38 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260460; this version posted July 22, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

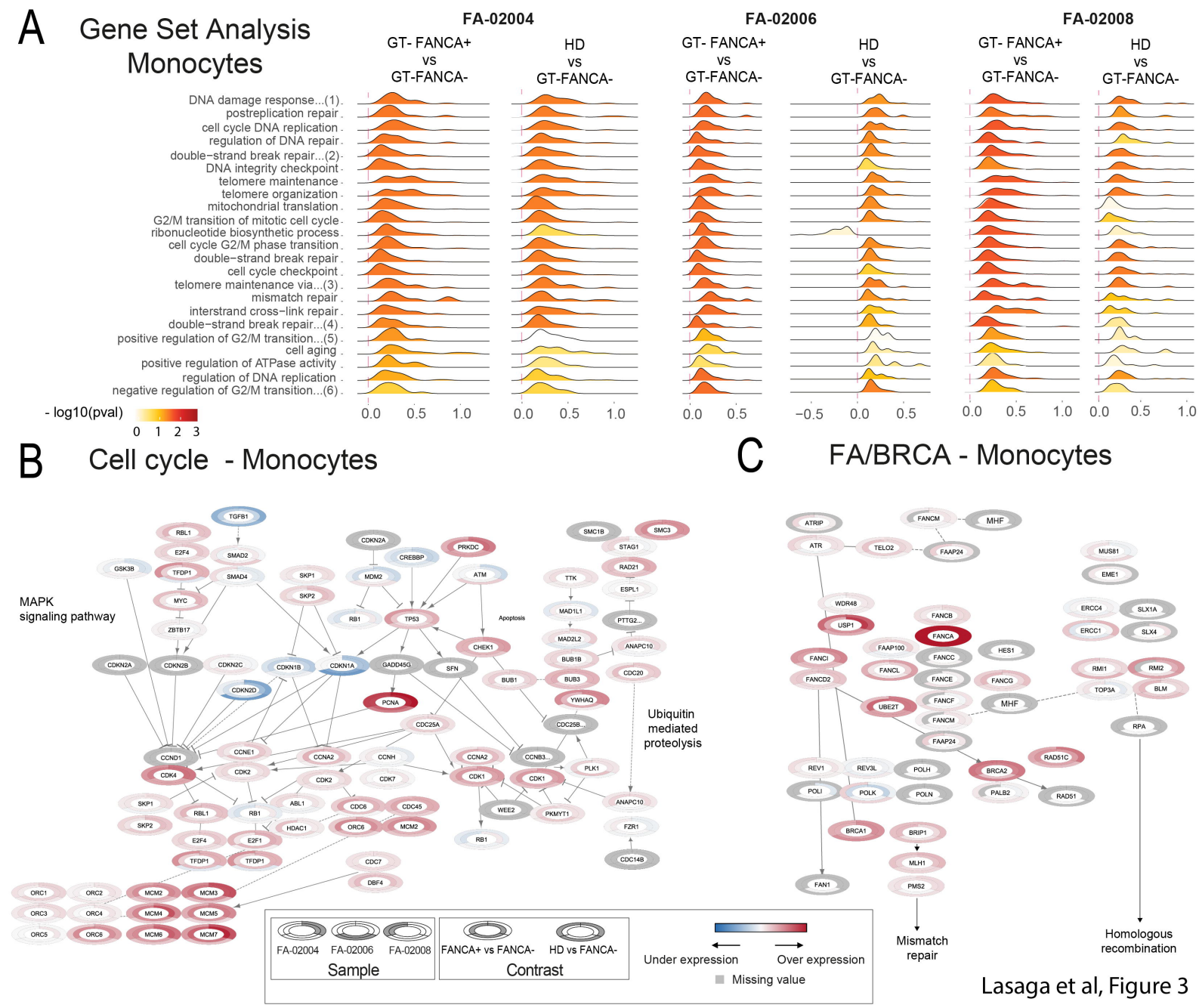
14. Gonzalez-Murillo, A.*, et al.* Development of lentiviral vectors with optimized transcriptional activity for the gene therapy of patients with Fanconi anemia. *Hum Gene Ther* **21**, 623-630 (2010).
15. Stuart, T.*, et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821 (2019).
16. Team, R.C. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2019).
17. Blondel, V.D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **P10008**(2008).
18. Shannon, P.*, et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).



Lasaga et al, Figure 1



Lasaga et al, Figure 2



Lasaga et al, Figure 3