Gene therapy restores the transcriptional program of hematopoietic stem cells in Fanconi anemia

Miren Lasaga,[1*] Paula Río,[2,3,4*] Amaia Vilas-Zornoza,[5,6*] Nuria Planell,[1] Susana Navarro,[2,3,4] Diego Alignani,[7] Beatriz Fernández-Varas,[3,8] Daniel Mouzo,[1] Josune Zubicaray,[9] Roser M. Pujol,[3,10,11] Eileen Nicoletti,[12] Jonathan D. Schwartz,[12] Julián Sevilla,[3,9] Marina Ainciburi,[5,6] Asier Ullate-Agote,[5,6] Jordi Surrallés,[3,10,11] Rosario Perona,[3,8,13] Leandro Sastre,[3,8] Felipe Prosper,[5,6#] David Gomez-Cabrero,[1,14,15#] and Juan A. Bueren[2,3,4#]

[1]Translational Bioinformatics Unit, Navarrabiomed, Universidad Pública de Navarra (UPNA), IdiSNA, Pamplona, Spain; [2]Hematopoietic Innovative Therapies Division, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain; [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), and Servicio de Hematologia y Terapia Celular, Clínica Universidad de Navarra, IDISNA, Pamplona, Spain; [6]Centro de Investigación Biomédica en Red de Cáncer, CIBERONC, Madrid, Spain; [7]Flow Cytometry Core, CIMA, Universidad de Navarra, Pamplona, Spain; [8]Instituto de Investigaciones Biomédicas Alberto Sols, CSIC/UAM, Madrid, Spain; [9]Hemoterapia y Hematología Pediátrica, Fundación para la Investigación Biomédica, Hospital Infantil Universitario Niño Jesús, Madrid, Spain; [10]Departmento de Genética y Microbiología, Universitat Autónoma de Barcelona, Barcelona, Spain; [11]Fundación Instituto de Investigación del Hospital de la Santa Creu y Sant Pau, Barcelona, Spain; [12]Rocket Pharmaceuticals Inc., New York, NY, USA; [13]Instituto de Salud Carlos III, Madrid, Spain; [14]Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia and [15]Bioscience Program, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

[*]ML, PR and AV-Z contributed equally as first authors.

[#]JAB, DG-C and FP contributed equally as senior authors.

Correspondence: Juan A. Bueren
juan.bueren@ciemat.es

David Gomez-Cabrero
david.gomezcabrero@kaust.edu.sa

Felipe Prosper
fprosper@unav.es

Received: November 16, 2022.
Accepted: March 29, 2023.
Early view: April 6, 2023.

https://doi.org/10.3324/haematol.2022.282418

Supplementary Materials for

Gene Therapy Restores the Transcriptional Program of Hematopoietic Stem Cells in Fanconi Anemia

Miren Lasaga[1,†], Paula Río[2,3,4,†], Amaia Vilas-Zornoza[5,6,†], Nuria Planell[1] , Susana Navarro [2,3,4], Diego Alignani[7], Beatriz Fernández-Varas[3,8], Daniel Mouzo[1], Josune Zubicaray[9], Roser M. Pujol[3,10,11], Eileen Nicoletti[12], Jonathan D. Schwartz[12], Julián Sevilla[3,9], Marina Ainciburi[5,6], Asier Ullate-Agote[5,6], Jordi Surrallés[3,10,11], Rosario Perona[3,8,13], Leandro Sastre[3,8], Felipe Prosper[5,6,*], David Gomez-Cabrero[1,14,15,*] and Juan A. Bueren[2,3,4,*].

†Equal first author contribution.

*Corresponding authors with equal contribution. Juan A. Bueren. Email: juan.bueren@ciemat.es; David Gomez-Cabrero. Email: David.gomez.cabrero@navarra.es; Felipe Prosper Email:fprosper@unav.es.

This PDF file includes:

- Supplementary Methods.
- Supplementary Figures S1, S2, S3, S4
- Supplementary Tables S1, S2, S3, S4

Supplementary Methods

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

The number of proviral copies per cell (VCN/cell) was analyzed after genomic extraction of the DNA using the DNA easy blood and tissue kit (Qiagen) or by proteinase K lysis as previously described (2). 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 (2). Percentages of gene correction in PB and BM were directly deduced from the number of copies of the therapeutic provirus in these cells, (eg. 20% corrected cells is deduced from the presence of 0.2 copies per PB or BM cells), given that the VCN per transduced CD34$^+$ cell was consistently lower than 1.0 (2).

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 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. According to these analyses $CD34^+$ cell populations were classified as *corrected* ($FANCA^+$) and *uncorrected* ($FANCA^-$) cells, considering that $FANCA^-$ is enriched with cells that only express the endogenous mutated *FANCA* mRNA, while $FANCA^+$, that includes cells with higher FANCA expression, is enriched with cells that express both the endogenous mutated *FANCA* plus the ectopic functional *FANCA* mRNA.

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 14,208 (FA-02002), 720 (FA-02004), 1,438 (FA-02006), 1,995 (FA-02008), and 12,549 (HD) 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. Uniform Manifold Approximation and Projection (UMAP) was performed to plot the data of each sample. 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 4 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 (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 Biosciences) as previously shown. The data-analysis processing of those

samples was conducted as the protocol described for FA samples. Next, we performed unsupervised clustering with the Louvain algorithm as implemented in Seurat[16]. 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 in 3YI 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 four 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).

Analysis of the sensitivity of hematopoietic colony forming cells to the genotoxic agent mitomycin C

To assess the influence that gene therapy had in the response of FA hematopoietic progenitors to mitomycin C (MMC), the number of colonies generated in the absence and the presence of this agent was assessed. In these experiments a total number of $2.5x10^5$ nucleated BM cells fractionated with Hydroxyethylstarches (HES; Grifols) were plated in plates containing 1 mL methylcellulose medium (MethocultTM #H4434) supplemented with 10 µg/mL anti-TNFα and 1 mM N-

acetylcysteine, in the absence and the presence of 10 nM MMC (Sigma-Aldrich). Cells were then cultured for 14 days at 37ºC, 5% CO_2 and 5% O_2, and colonies were then scored under an inverted microscope.

Telomere length studies

DNA was extracted from patient's blood samples and the telomere length was determined by quantitative PCR as previously described(20). In this method the amount of telomere DNA (T) and of the single copy *36B4* reference gene (S) were determined by quantitative PCR for each blood sample. The ratio between these two parameters (T/S) was a measure of the relative telomere length. A control DNA isolated from the cultured cell line MCF-7 was used as an internal control in each experiment to normalize the T/S ratio obtained for the experimental samples. The telomere length of each sample was calculated from the normalized T/S ratios using the formula: telomere length in Kbp = T/S x 3.86 + 1.89. Three independent experiments with triplicates were conducted for each sample.

Basic statistical tests

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.

Anova 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 $CD34^+$ 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(15) was used. In all the cases the multiple testing was addressed using Bonferroni; and for any analysis the null hypothesis was rejected if adjusted p-value <0.05.

References

1. Ferrari G, Thrasher AJ, Aiuti A. Gene therapy using haematopoietic stem and progenitor cells. Nat Rev Genet. 2021 Apr;22(4):216-34.
2. Rio P, Navarro S, Wang W, Sanchez-Dominguez R, Pujol RM, Segovia JC, et al. Successful engraftment of gene-corrected hematopoietic stem cells in non-conditioned patients with Fanconi anemia. Nat Med. 2019 Sep;25(9):1396-401.
3. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol. 2017 Apr;19(4):271-81.
4. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nature communications. 2017 Jan 16;8:14049.
5. Jaber S, Toufektchan E, Lejour V, Bardot B, Toledo F. p53 downregulates the Fanconi anaemia DNA repair pathway. Nature communications. 2016 Apr 1;7:11091.
6. Rego MA, Harney JA, Mauro M, Shen M, Howlett NG. Regulation of the activation of the Fanconi anemia pathway by the p21 cyclin-dependent kinase inhibitor. Oncogene. 2012 Jan 19;31(3):366-75.
7. Ceccaldi R, Parmar K, Mouly E, Delord M, Kim JM, Regairaz M, 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. 2012 Jul 06;11(1):36-49.
8. Zhang H, Kozono DE, O'Connor KW, Vidal-Cardenas S, Rousseau A, Hamilton A, et al. TGF-β Inhibition Rescues Hematopoietic Stem Cell Defects and Bone Marrow Failure in Fanconi Anemia. Cell stem cell. 2016 May 5;18(5):668-81.
9. Rodríguez A, Zhang K, Färkkilä A, Filiatrault J, Yang C, Velázquez M, et al. MYC Promotes Bone Marrow Stem Cell Dysfunction in Fanconi Anemia. Cell stem cell. 2021 Jan 7;28(1):33-47.e8.
10. Flach J, Bakker ST, Mohrin M, Conroy PC, Pietras EM, Reynaud D, et al. Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. Nature. 2014 Aug 14;512(7513):198-202.

11. Brosh RM, Jr., Bellani M, Liu Y, Seidman MM. Fanconi Anemia: A DNA repair disorder characterized by accelerated decline of the hematopoietic stem cell compartment and other features of aging. Ageing Res Rev. 2017 Jan;33:67-75.
12. Leteurtre F, Li X, Guardiola P, Le Roux G, Sergère JC, Richard P, et al. Accelerated telomere shortening and telomerase activation in Fanconi's anaemia. Br J Haematol. 1999 Jun;105(4):883-93.
13. Hanson H, Mathew CG, Docherty Z, Mackie Ogilvie C. Telomere shortening in Fanconi anaemia demonstrated by a direct FISH approach. Cytogenetics and cell genetics. 2001;93(3-4):203-6.
14. Gonzalez-Murillo A, Lozano ML, Alvarez L, Jacome A, Almarza E, Navarro S, et al. Development of lentiviral vectors with optimized transcriptional activity for the gene therapy of patients with Fanconi anemia. Hum Gene Ther. 2010 May;21(5):623-30.
15. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888-902.e21.
16. Team RC. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019.
17. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008 9 October 2008 P10008.
18. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504.
19. Castella M, Pujol R, Callen E, Ramirez MJ, Casado JA, Talavera M, et al. Chromosome fragility in patients with Fanconi anaemia: diagnostic implications and clinical impact. J Med Genet. 2011 Apr;48(4):242-50.
20. Cawthon RM. Telomere measurement by quantitative PCR. Nucleic Acids Res. 2002 May 15;30(10):e47.

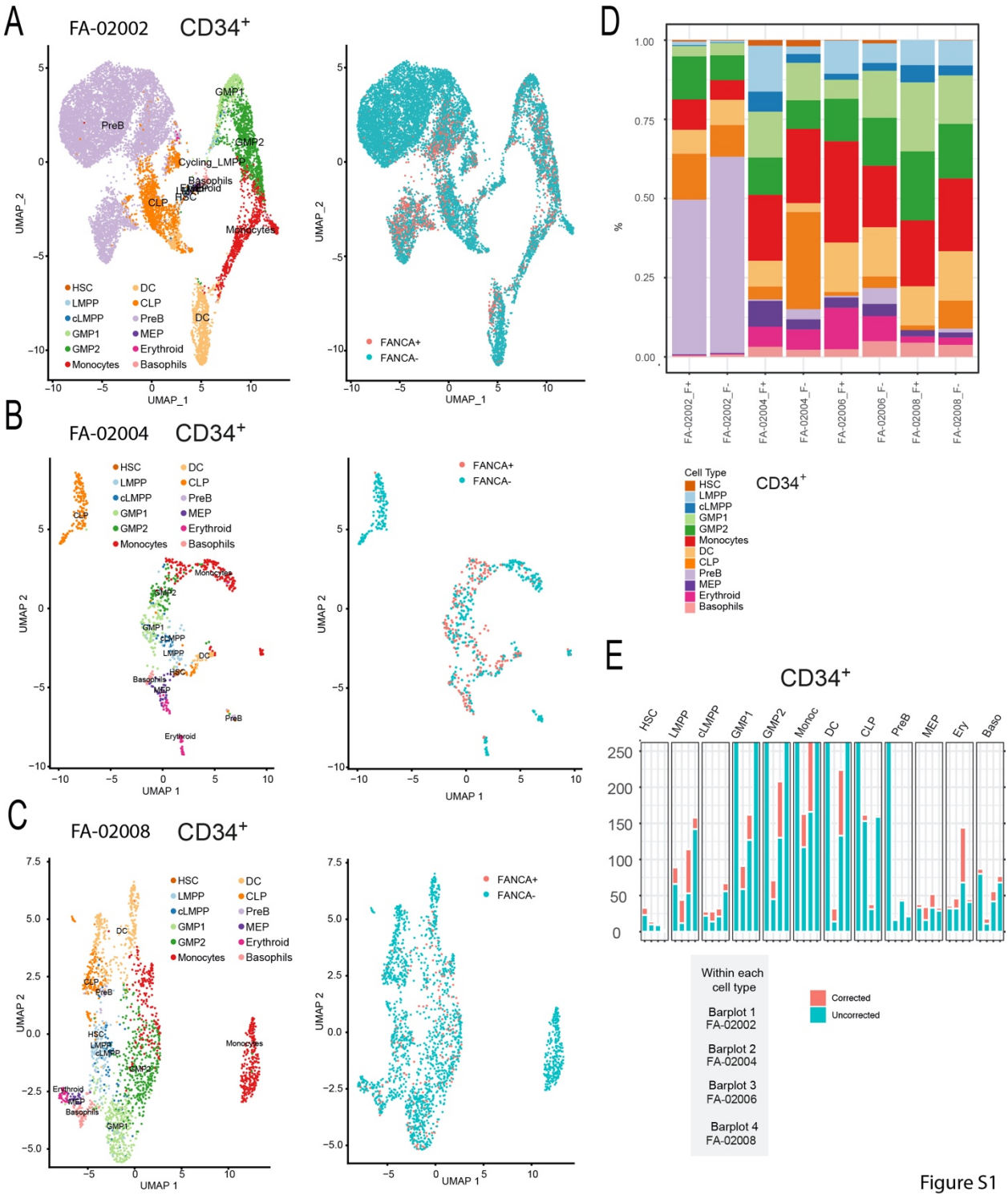

Figure S1. (A) Left panel: UMAP plot showing the clustering analysis for CD34$^+$ BM cells from the FA-02002 patient undergoing FANCA gene therapy. A total of 12 clusters were identified, spanning the different HSPC subpopulations. Identified clusters include an HSC cluster

(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). Right panel: Distribution of FANCA positive cells (corrected cells; red) versus FANCA negative cells (uncorrected cells; blue). **(B,C)** Same as (A), including the analysis of the FA-02004 and FA-02008 respectively patient. **(D)** Cell type proportions for each individual separating in each case $FANCA^+$ and $FANCA^-$ cells. **(E)** Barplot showing the total number of cells in the different HSPC populations corresponding to the four gene therapy treated patients. In each case, the number of $FANCA^+$ (red) and $FANCA^-$ (blue) cells is shown. Panel E is a zoom over low prevalent cells that complements Fig. 2C.

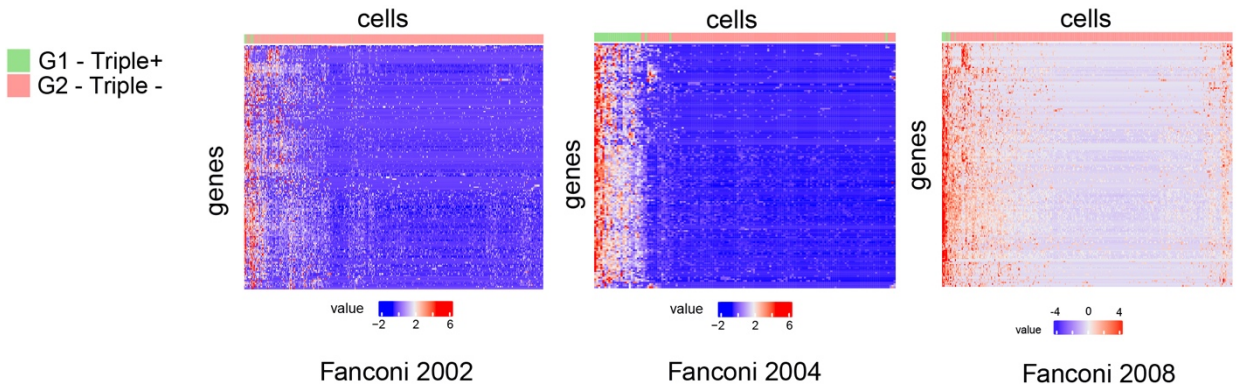

Figure S2. Fingerprint validation. Comparison of gene expression of fingerprint between G1 (cells with reads mapping to FANCA, poly-A sequence of the viral vector, and PGK1) and G2 (cells with no reads mapping to FANCA, neither to viral vector's poly-A sequence, neither to PGK1). For each sample, cells have been selected, and expression plotted separately. In addition, values have been z-scored within each of the samples. In the case of patient FA-02006, no cell in G1 was identified.

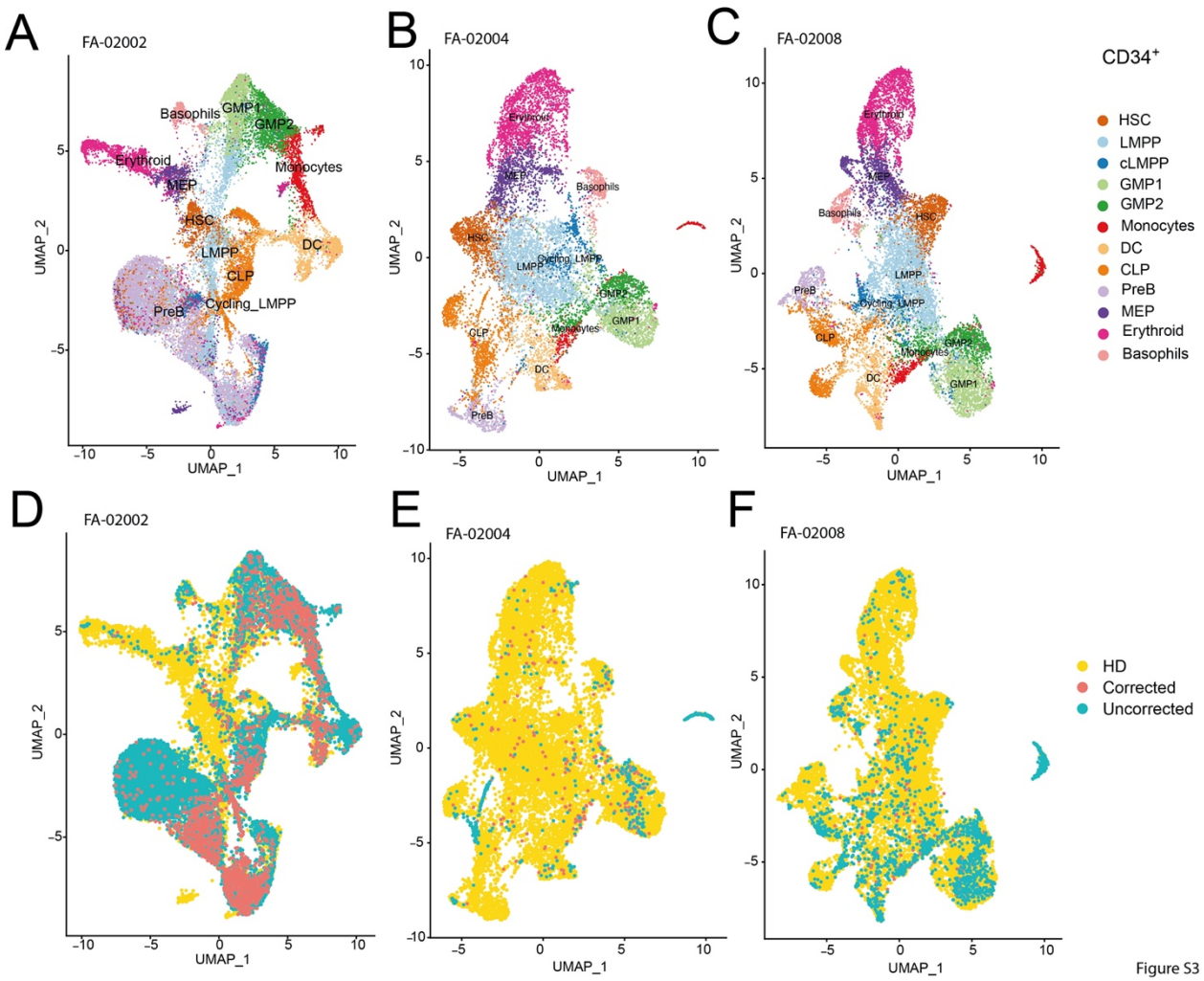

Figure S3. (A) UMAP plot showing the clustering analysis for CD34^{+} BM cells from the FA-02002 patient undergoing FANCA gene therapy and the healthy donor. A total of 12 clusters were identified, spanning the different HSPC subpopulations. Identified clusters include an HSC cluster (hematopoietic stem 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). **(B, C)** Same as (A), including the analysis of the FA-02004 and FA-02008 respectively patient. **(D)** Distribution of cells classified as cells derived from the healthy

donor (yellow), FANCA^{+} cells (red) and FANCA^{-} cells (blue) for FA-02002 patient. (E, F) Same as (D) including the analysis of the FA-02004 and FA-02008 respectively patient.

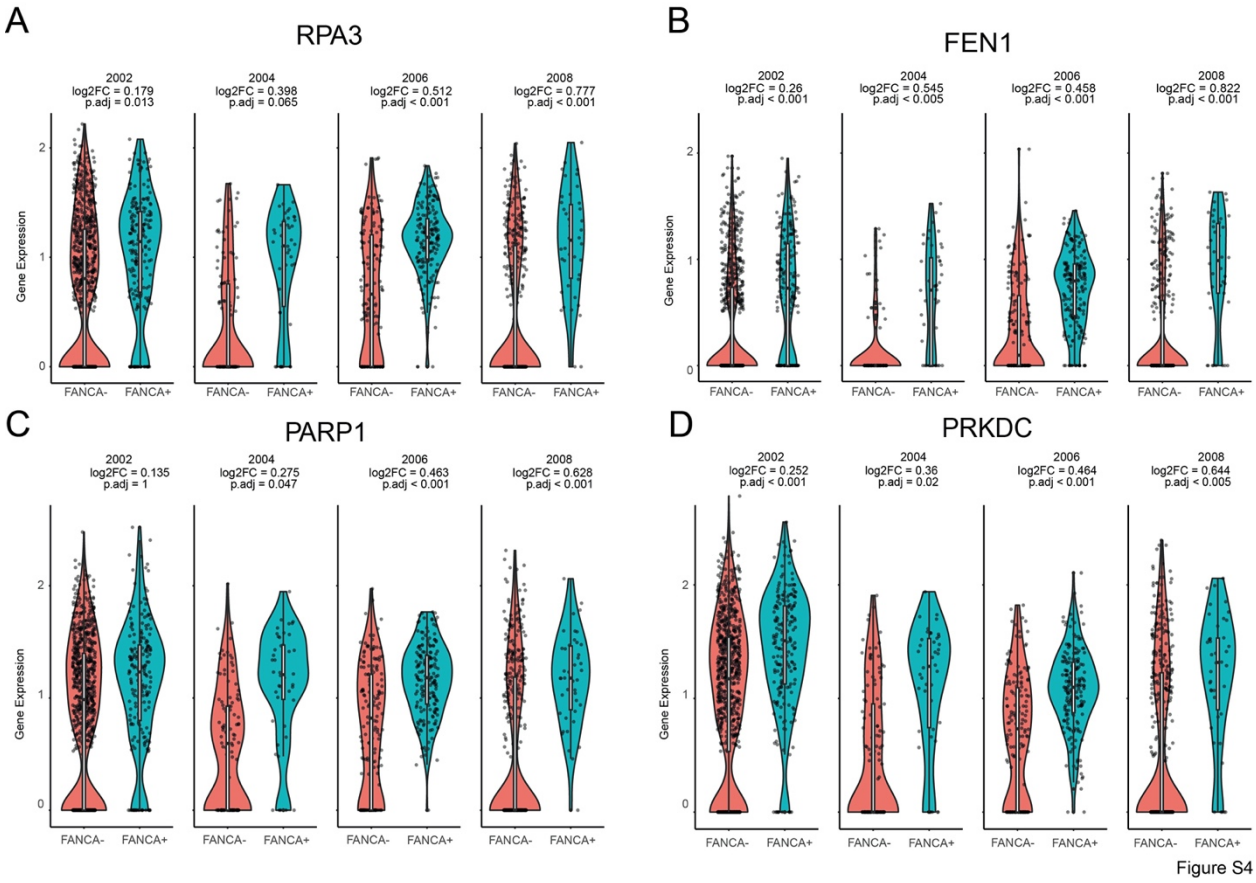

Figure S4. Gene expression differences between FANCA+ and FANCA- for selected Telomere related genes in CD34+ progenitor monocytes. The violin plots depict normalized gene expression for each patient separately in CD34+ progenitor monocytes for FANCA- (red, left) and FANCA+ (light-blue, right). Log2FC denotes logarithmic fold-change, and p.adj denotes adjusted p-value for each FANCA+ vs. FANCA- contrast. Panels A, B, C, and D show the information for RPA3, FEN1, PARP1, and PRKDC respectively.

Cell type	FA-02002	FA-02004	FA-02006	FA-02008
	p.adjust	p.adjust	p.adjust	p.adjust
HSC	1.000	1.000	1.000	1.000
LMPP	0.211	0.000	0.206	1.000
Cycling_LMPP	1.000	1.000	1.000	1.000
GMP1	1.000	1.000	0.000	1.000
GMP2	0.000	1.000	1.000	1.000
Monocytes	0.000	1.000	0.000	1.000
DC	1.000	0.125	1.000	1.000
CLP	0.000	0.000	0.399	0.022
PreB	0.000	1.000	0.000	1.000
MEP	1.000	0.342	1.000	1.000
Erythroid	1.000	1.000	0.099	1.000
Basophils	1.000	1.000	1.000	1.000

Table S1. Results of the statistical test comparing the proportion of $FANCA^{+}$ vs $FANCA^{-}$ HSPCs. For each HSPC type and each patient's sample the adjusted p-value is shown. Bonferroni was used to correct for multiple testing.

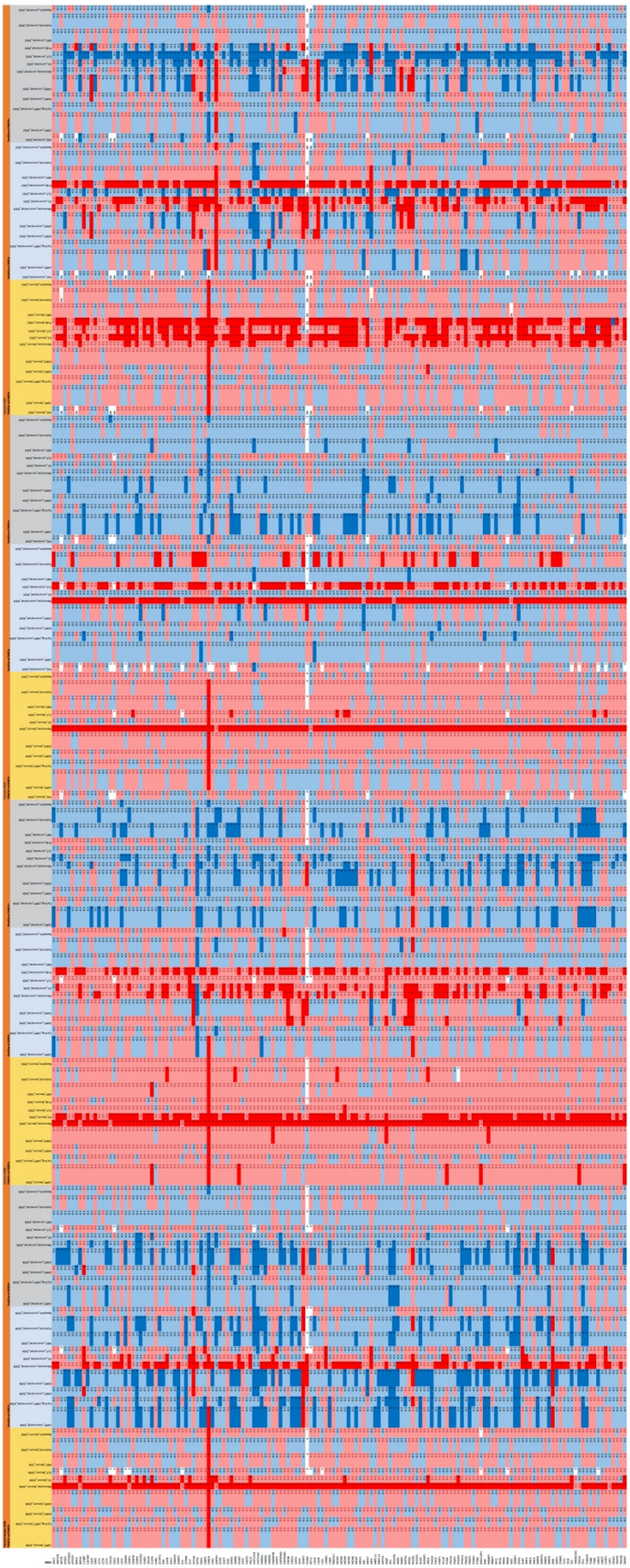

Table S2. Differential $FANCA^+$ signature derived from Fig.1F. The genes included in the list are those that for at least one cell type are identified as differentially expressed (abs(logFC)>0.25 and adjusted p-value<0.05) in "at least three patients", and "showing the same direction of the change for the three patients", when considering the contrast $FANCA^+$ vs $FANCA^-$ HSPCs (n=152). FANCA was excluded from the analysis. For each gene the information is provided for each individual, for each contrast ($FANCA^+$ vs $FANCA^-$ and Healthy vs FANCA+) conducted for each cell type. There is a color/number code that denotes: 1/red upregulated and statistically significant; 0.5/light-red upregulated and not statistically significant; 0 no changes; -1/blue downregulated and statistically significant; 0.5/light-blue downregulated and not statistically significant.

Cell type	Binomial Test								Correlation			
	FA-02002		FA-02004		FA-02006		FA-02008		FA-02002	FA-02004	FA-02006	FA-02008
	p.adjust	# of genes	p.adjust	# of genes	p.adjust	# of genes	p.adjust	# of genes	p.adjust	p.adjust	p.adjust	p.adjust
HSC	1.0000	73	0.0000	107	NA	NA	NA	NA	0.8340	0.0000	NA	NA
LMPP	1.0000	76	0.1410	90	1.0000	77	1.0000	50	1.0000	0.0006	0.0453	0.3469
Cycling_LMPP	0.0007	100	1.0000	74	1.0000	73	1.0000	66	0.0005	1.0000	1.0000	1.0000
GMP1	0.0434	93	1.0000	58	0.0000	127	1.0000	60	0.1230	1.0000	0.0059	1.0000
GMP2	1.0000	64	1.0000	78	1.0000	68	1.0000	48	1.0000	0.0001	1.0000	1.0000
Monocytes	0.0000	130	0.0000	152	0.0000	141	0.0000	145	1.0000	0.0839	1.0000	1.0000
DC	0.0000	143	1.0000	79	0.0000	144	0.0001	103	0.0000	0.0029	0.0015	1.0000
CLP	1.0000	21	0.0000	138	0.0000	137	1.0000	73	0.0005	0.4071	1.0000	1.0000
PreB	0.0000	148	NA	NA	0.0000	135	NA	NA	0.0000	NA	0.1257	NA
MEP	0.0004	101	1.0000	41	1.0000	45	1.0000	53	1.0000	1.0000	1.0000	1.0000
Erythroid	1.0000	81	0.0000	127	1.0000	73	0.0022	98	1.0000	0.0000	0.0479	0.0947
Basophils	0.0154	95	1.0000	80	0.0000	123	0.8386	85	0.2829	0.0092	0.0059	0.4477

Table S3. Comparison of the directionality of the following two contrasts: "FANCA$^+$ vs. FANCA$^-$", and "Healthy vs. FANCA$^-$". "# genes tot", denotes the total number of genes per individual and cell-type considered for the analysis. For the two analysis the genes sharing the same directionality for both contrasts were classified as 1, and 0 otherwise (see Methods). "Binomial test": the analysis was executed independently by cell type and sample. The adjusted p-value after Bonferroni multiple testing correction is provided. NA denotes that it was not possible to compute the p-value due to the limited number of cells. "# of genes" denotes the number of genes with the same directionality in both contrasts. Correlation: binary correlation.

ID	Description	enrichmentSc	pvalue	p.adjust	Cell_type	Contrast
GO:0009262	deoxyribonucleotide metabolic process	0.71817658	0.00174825	0.01525153	Monocytes	FA-02008 (FAI
GO:0032212	positive regulation of telomere maintenance via telomerase	0.70294651	0.00174825	0.01525153	Monocytes	FA-02008 (FAI
GO:0050686	negative regulation of mRNA processing	0.67199855	0.00174825	0.01525153	Monocytes	FA-02008 (FAI
GO:0051973	positive regulation of telomerase activity	0.67407468	0.00174825	0.01525153	Monocytes	FA-02008 (FAI
GO:0000413	protein peptidyl-prolyl isomerization	0.65410391	0.00175439	0.01525153	Monocytes	FA-02008 (FAI
GO:0000460	maturation of 5.8S rRNA	0.6826268	0.00175439	0.01525153	Monocytes	FA-02008 (FAI
GO:0006270	DNA replication initiation	0.77512633	0.00175439	0.01525153	Monocytes	FA-02008 (FAI
GO:0019692	deoxyribose phosphate metabolic process	0.68523953	0.00175439	0.01525153	Monocytes	FA-02008 (FAI
GO:0033119	negative regulation of RNA splicing	0.71547714	0.00175439	0.01525153	Monocytes	FA-02008 (FAI
GO:0051984	positive regulation of chromosome segregation	0.67395754	0.00175439	0.01525153	Monocytes	FA-02008 (FAI
GO:0070198	protein localization to chromosome, telomeric region	0.8256152	0.00175439	0.01525153	Monocytes	FA-02008 (FAI
GO:1904358	positive regulation of telomere maintenance via telomere lengthening	0.70946365	0.00175747	0.01525153	Monocytes	FA-02008 (FAI
GO:0006195	purine nucleotide catabolic process	0.73517895	0.00176678	0.01525153	Monocytes	FA-02008 (FAI
GO:0006297	nucleotide-excision repair, DNA gap filling	0.67011356	0.00176678	0.01525153	Monocytes	FA-02008 (FAI
GO:0006298	mismatch repair	0.66220311	0.00176678	0.01525153	Monocytes	FA-02008 (FAI
GO:0009394	2'-deoxyribonucleotide metabolic process	0.69503705	0.00176678	0.01525153	Monocytes	FA-02008 (FAI
GO:0030262	apoptotic nuclear changes	0.67789663	0.00176678	0.01525153	Monocytes	FA-02008 (FAI
GO:0051383	kinetochore organization	0.72093834	0.00176678	0.01525153	Monocytes	FA-02008 (FAI
GO:0006921	cellular component disassembly involved in execution phase of apoptosis	0.67510528	0.00177305	0.01525153	Monocytes	FA-02008 (FAI
GO:0032201	telomere maintenance via semi-conservative replication	0.75240532	0.00177305	0.01525153	Monocytes	FA-02008 (FAI
GO:0070199	establishment of protein localization to chromosome	0.84596003	0.00177305	0.01525153	Monocytes	FA-02008 (FAI
GO:0006244	pyrimidine nucleotide catabolic process	0.80755715	0.0017762	0.01525153	Monocytes	FA-02008 (FAI
GO:0007339	binding of sperm to zona pellucida	0.8780535	0.0017762	0.01525153	Monocytes	FA-02008 (FAI
GO:0030261	chromosome condensation	0.71905504	0.0017762	0.01525153	Monocytes	FA-02008 (FAI
GO:1903405	protein localization to nuclear body	0.91541891	0.0017762	0.01525153	Monocytes	FA-02008 (FAI
GO:1904851	positive regulation of establishment of protein localization to telomere	0.91541891	0.0017762	0.01525153	Monocytes	FA-02008 (FAI
GO:1904867	protein localization to Cajal body	0.91541891	0.0017762	0.01525153	Monocytes	FA-02008 (FAI
GO:0042407	cristae formation	0.63080171	0.00177936	0.01525153	Monocytes	FA-02008 (FAI
GO:0009113	purine nucleobase biosynthetic process	0.73084129	0.00178253	0.01525153	Monocytes	FA-02008 (FAI
GO:0009263	deoxyribonucleotide biosynthetic process	0.80297149	0.00178253	0.01525153	Monocytes	FA-02008 (FAI
GO:0070203	regulation of establishment of protein localization to telomere	0.89752693	0.00178253	0.01525153	Monocytes	FA-02008 (FAI
GO:1904031	positive regulation of cyclin-dependent protein kinase activity	0.72451391	0.00178253	0.01525153	Monocytes	FA-02008 (FAI
GO:1904816	positive regulation of protein localization to chromosome, telomeric region	0.91707246	0.00178253	0.01525153	Monocytes	FA-02008 (FAI
GO:0000466	maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	0.73017046	0.00178571	0.01525153	Monocytes	FA-02008 (FAI
GO:0022616	DNA strand elongation	0.72251057	0.00178571	0.01525153	Monocytes	FA-02008 (FAI
GO:0090670	RNA localization to Cajal body	0.86556636	0.00178571	0.01525153	Monocytes	FA-02008 (FAI
GO:0090671	telomerase RNA localization to Cajal body	0.86556636	0.00178571	0.01525153	Monocytes	FA-02008 (FAI
GO:0090672	telomerase RNA localization	0.86556636	0.00178571	0.01525153	Monocytes	FA-02008 (FAI
GO:0090685	RNA localization to nucleus	0.86556636	0.00178571	0.01525153	Monocytes	FA-02008 (FAI
GO:0000469	cleavage involved in rRNA processing	0.68247898	0.00179211	0.01525153	Monocytes	FA-02008 (FAI
GO:0010499	proteasomal ubiquitin-independent protein catabolic process	0.7724766	0.00179211	0.01525153	Monocytes	FA-02008 (FAI
GO:1902751	positive regulation of cell cycle G2/M phase transition	0.69333963	0.00179211	0.01525153	Monocytes	FA-02008 (FAI
GO:0034080	CENP-A containing nucleosome assembly	0.73328797	0.00179533	0.01525153	Monocytes	FA-02008 (FAI
GO:0061641	CENP-A containing chromatin organization	0.73328797	0.00179533	0.01525153	Monocytes	FA-02008 (FAI
GO:0000291	nuclear-transcribed mRNA catabolic process, exonucleolytic	0.6166161	0.00179856	0.01525153	Monocytes	FA-02008 (FAI
GO:0006261	DNA-dependent DNA replication	0.64754021	0.00179856	0.01525153	Monocytes	FA-02008 (FAI
GO:0007088	regulation of mitotic nuclear division	0.47897688	0.00179856	0.01525153	Monocytes	FA-02008 (FAI
GO:0022900	electron transport chain	0.50166408	0.00179856	0.01525153	Monocytes	FA-02008 (FAI
GO:0007076	mitotic chromosome condensation	0.78333745	0.0018018	0.01525153	Monocytes	FA-02008 (FAI
GO:0009148	pyrimidine nucleoside triphosphate biosynthetic process	0.73505206	0.0018018	0.01525153	Monocytes	FA-02008 (FAI
GO:0009154	purine ribonucleotide catabolic process	0.74979727	0.0018018	0.01525153	Monocytes	FA-02008 (FAI
GO:0035036	sperm-egg recognition	0.74796493	0.0018018	0.01525153	Monocytes	FA-02008 (FAI
GO:0044788	modulation by host of viral process	0.78415702	0.0018018	0.01525153	Monocytes	FA-02008 (FAI
GO:0070202	regulation of establishment of protein localization to chromosome	0.87601141	0.0018018	0.01525153	Monocytes	FA-02008 (FAI
GO:0070987	error-free translesion synthesis	0.70132774	0.0018018	0.01525153	Monocytes	FA-02008 (FAI
GO:1904814	regulation of protein localization to chromosome, telomeric region	0.90333773	0.0018018	0.01525153	Monocytes	FA-02008 (FAI
GO:1990173	protein localization to nucleoplasm	0.82728175	0.0018018	0.01525153	Monocytes	FA-02008 (FAI
GO:0006260	DNA replication	0.59545381	0.00180505	0.01525153	Monocytes	FA-02008 (FAI
GO:0006403	RNA localization	0.49356949	0.00180505	0.01525153	Monocytes	FA-02008 (FAI
GO:0010948	negative regulation of cell cycle process	0.38099218	0.00180505	0.01525153	Monocytes	FA-02008 (FAI
GO:0016072	rRNA metabolic process	0.53085252	0.00180505	0.01525153	Monocytes	FA-02008 (FAI
GO:0051438	regulation of ubiquitin-protein transferase activity	0.57420637	0.00180505	0.01525153	Monocytes	FA-02008 (FAI
GO:0006335	DNA replication-dependent nucleosome assembly	0.76647602	0.00180832	0.01525153	Monocytes	FA-02008 (FAI
GO:0006414	translational elongation	0.63650416	0.00180832	0.01525153	Monocytes	FA-02008 (FAI
GO:0009141	nucleoside triphosphate metabolic process	0.54743047	0.00180832	0.01525153	Monocytes	FA-02008 (FAI
GO:0010389	regulation of G2/M transition of mitotic cell cycle	0.46025685	0.00180832	0.01525153	Monocytes	FA-02008 (FAI
GO:0015985	energy coupled proton transport, down electrochemical gradient	0.75720046	0.00180832	0.01525153	Monocytes	FA-02008 (FAI
GO:0015986	ATP synthesis coupled proton transport	0.75720046	0.00180832	0.01525153	Monocytes	FA-02008 (FAI
GO:0031935	regulation of chromatin silencing	0.76265508	0.00180832	0.01525153	Monocytes	FA-02008 (FAI
GO:0034723	DNA replication-dependent nucleosome organization	0.76647602	0.00180832	0.01525153	Monocytes	FA-02008 (FAI
GO:0048025	negative regulation of mRNA splicing, via spliceosome	0.77236926	0.00180832	0.01525153	Monocytes	FA-02008 (FAI
GO:1904872	regulation of telomerase RNA localization to Cajal body	0.87923737	0.00180832	0.01525153	Monocytes	FA-02008 (FAI
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	0.5223547	0.00181159	0.01525153	Monocytes	FA-02008 (FAI
GO:0000723	telomere maintenance	0.64435832	0.00181159	0.01525153	Monocytes	FA-02008 (FAI
GO:0006278	RNA-dependent DNA biosynthetic process	0.67956821	0.00181159	0.01525153	Monocytes	FA-02008 (FAI
GO:0009124	nucleoside monophosphate biosynthetic process	0.50558654	0.00181159	0.01525153	Monocytes	FA-02008 (FAI
GO:0045047	protein targeting to ER	0.53298755	0.00181159	0.01525153	Monocytes	FA-02008 (FAI
GO:0050684	regulation of mRNA processing	0.500103	0.00181159	0.01525153	Monocytes	FA-02008 (FAI
GO:1903311	regulation of mRNA metabolic process	0.41048013	0.00181159	0.01525153	Monocytes	FA-02008 (FAI
GO:0000387	spliceosomal snRNP assembly	0.74916992	0.00181488	0.01525153	Monocytes	FA-02008 (FAI
GO:0006284	base-excision repair	0.63675473	0.00181488	0.01525153	Monocytes	FA-02008 (FAI
GO:0006333	chromatin assembly or disassembly	0.60308149	0.00181488	0.01525153	Monocytes	FA-02008 (FAI
GO:0009161	ribonucleoside monophosphate metabolic process	0.53045391	0.00181488	0.01525153	Monocytes	FA-02008 (FAI
GO:0019985	translesion synthesis	0.57931035	0.00181488	0.01525153	Monocytes	FA-02008 (FAI
GO:0033683	nucleotide-excision repair, DNA incision	0.6378135	0.00181488	0.01525153	Monocytes	FA-02008 (FAI
GO:0042769	DNA damage response, detection of DNA damage	0.70063136	0.00181488	0.01525153	Monocytes	FA-02008 (FAI
GO:0043628	ncRNA 3'-end processing	0.64536696	0.00181488	0.01525153	Monocytes	FA-02008 (FAI
GO:0051972	regulation of telomerase activity	0.62613013	0.00181488	0.01525153	Monocytes	FA-02008 (FAI
GO:0006120	mitochondrial electron transport, NADH to ubiquinone	0.71676056	0.00181818	0.01525153	Monocytes	FA-02008 (FAI
GO:0007007	inner mitochondrial membrane organization	0.62445786	0.00181818	0.01525153	Monocytes	FA-02008 (FAI
GO:0016458	gene silencing	0.41608258	0.00181818	0.01525153	Monocytes	FA-02008 (FAI
GO:0032206	positive regulation of telomere maintenance	0.66461151	0.00181818	0.01525153	Monocytes	FA-02008 (FAI
GO:0034728	nucleosome organization	0.59992099	0.00181818	0.01525153	Monocytes	FA-02008 (FAI
GO:0071824	protein-DNA complex subunit organization	0.55902571	0.00181818	0.01525153	Monocytes	FA-02008 (FAI
GO:0072599	establishment of protein localization to endoplasmic reticulum	0.50701546	0.00181818	0.01525153	Monocytes	FA-02008 (FAI
GO:2001251	negative regulation of chromosome organization	0.52241721	0.00181818	0.01525153	Monocytes	FA-02008 (FAI
GO:0000086	G2/M transition of mitotic cell cycle	0.42217318	0.00182149	0.01525153	Monocytes	FA-02008 (FAI
GO:0006399	tRNA metabolic process	0.49042598	0.00182149	0.01525153	Monocytes	FA-02008 (FAI
GO:0007338	single fertilization	0.59886016	0.00182149	0.01525153	Monocytes	FA-02008 (FAI
GO:0043484	regulation of RNA splicing	0.47396752	0.00182149	0.01525153	Monocytes	FA-02008 (FAI

Table S4. Gene Set Enrichment Analysis. Results of Gene Set Enrichment Analysis conducted for each FA patient, for cell type and for each of the contrasts of interest. Only top 100 results are provided; such threshold allows the insertion of the table in the supplement document requested (other thresholds may require a tab-supplementary file).