

## Epigenetically induced ectopic expression of UNCX impairs the proliferation and differentiation of myeloid cells

Giulia Daniele,<sup>1</sup> Giorgia Simonetti,<sup>2</sup> Caterina Fusilli,<sup>3</sup> Ilaria Iacobucci,<sup>2</sup> Angelo Lonoce,<sup>1</sup> Antonio Palazzo,<sup>1</sup> Mariana Lomiento,<sup>4</sup> Fabiana Mammoli,<sup>4</sup> Renè Massimiliano Marsano,<sup>1</sup> Elena Marasco,<sup>2</sup> Vilma Mantovani,<sup>5,6</sup> Hilmar Quentmeier,<sup>7</sup> Hans G Drexler,<sup>7</sup> Jie Ding,<sup>7</sup> Orazio Palumbo,<sup>8</sup> Massimo Carella,<sup>8</sup> Niroshan Nadarajah,<sup>9</sup> Margherita Perricone,<sup>2</sup> Emanuela Ottaviani,<sup>2</sup> Carmen Baldazzi,<sup>2</sup> Nicoletta Testoni,<sup>2</sup> Cristina Papayannidis,<sup>2</sup> Sergio Ferrari,<sup>4</sup> Tommaso Mazza,<sup>3</sup> Giovanni Martinelli<sup>2</sup> and Clelia Tiziana Storlazzi\*<sup>1</sup>

<sup>1</sup>Department of Biology, University of Bari "A. Moro", Italy; <sup>2</sup>Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Italy; <sup>3</sup>IRCCS Casa Sollievo della Sofferenza, Bioinformatics Unit, San Giovanni Rotondo, Italy; <sup>4</sup>Department of Life Science, University of Modena and Reggio Emilia, Modena, Italy; <sup>5</sup>Center for Applied Biomedical Research (CRBA), S. Orsola-Malpighi Hospital, Bologna, Italy; <sup>6</sup>Unit of Medical Genetics, Department of Medical and Surgical Sciences, S. Orsola-Malpighi Hospital University of Bologna, Italy; <sup>7</sup>Leibniz-Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Department of Human and Animal Cell Lines, Braunschweig, Germany; <sup>8</sup>Medical Genetics Unit, IRCCS "Casa Sollievo della Sofferenza (CSS)" Hospital, San Giovanni Rotondo, Italy; <sup>9</sup>MLL Münchner Leukämielabor GmbH, München, Germany and <sup>10</sup>Institute of Hematology "L. e A. Seragnoli" S.Orsola-Malpighi Hospital, Bologna, Italy

©2017 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2016.163022

Received: December 27, 2016.

Accepted: April 12, 2017.

Pre-published: April 14, 2017.

Correspondence: cleliatiziana.storlazzi@uniba.it

---

## Supplementary Methods

### Patients, cell lines, and normal tissues

All patients who were enrolled at the Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna (Bologna, Italy) provided informed consent for blood collection and biological analyses, in agreement with the Declaration of Helsinki.

AML cell lines were held by the DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH; www.dsmz.de). Detailed references and cultivation protocols related to the cell lines have been previously described<sup>1</sup>.

As PCR controls, we used peripheral blood (PB), cord blood (CB) and bone marrow (BM) stem cell-derived CD34+ RNA from healthy donors, as well as commercial BM RNA (Cat. No. 636591; Clontech, Jesi, Italy), total and fetal brain RNA (Cat. Nos. 636102 and 636526), Y79 (retinoblastoma), and HEK293T (embryonic kidney) cell lines (Supplementary Table S1B). Human CD34+ cells were obtained as previously described<sup>2</sup>.

### Fluorescence in situ hybridization (FISH)

FISH was performed on BM metaphases from 3:1 methanol:acetic acid-fixed chromosome suspensions, as previously described<sup>3</sup>, in Case 1 at the time of diagnosis (1-Dx), in AML patients (no. 9, 16, and 61) showing the highest *UNCX* expression, and in all of the investigated cell lines, to exclude the occurrence of any cryptic rearrangement. Appropriate bacterial artificial chromosome (BAC) (Roswell Park Cancer Institute [RPCI]-11 Human Male Bac Library, Buffalo, NY) and fosmid (WIBR2 Human Fosmid Library) clones were selected according to the GRCh37/hg19 sequence assembly and obtained from the BACPAC Resource Center (<http://bacpac.chori.org>). Each clone was preliminarily tested on normal human metaphase cells.

In these four AML patients, we also investigated the mutational status of the *NPM1*, *FLT3*, and *WT1* genes.

### RT-qPCR

The TaqMan *UNCX* Gene expression assay (ID Hs01394890\_g1, Applied Biosystems, Milan, Italy) was also used to validate the effectiveness of *UNCX* infection assays in human purified cord blood (CB) CD34+ cells (Online Supplementary Figure S6D).

RT-qPCR analyses of genes encoding specific surface markers and transcription factors implicated in hematopoiesis were performed on CD34+ cells infected with the LUNCXIΔN vector by using the FastStart Universal Probe Master Mix (Roche Diagnostics). CD34+ cells transduced with the empty vector (LXIΔN) served as the calibrator. RNA extraction was performed at different time points to investigate. The genes analyzed were *CD34* (ID Hs00990732\_m1), *HOXA10* (ID Hs00172012\_m1), *KLF4* (ID Hs00358836\_m1), and *MAFB* (ID Hs00534343\_s1). *GAPDH* (ID Hs002758991\_g1) was used as the endogenous control. All reagents used in TaqMan gene-expression assays were supplied by Applied Biosystems (Milan, Italy).

The expression profile of the most interesting differentially expressed genes derived matching our exon-array and TCGA (The Cancer Genome Atlas) data was tested by using RT-qPCR experiments with SYBR Green (SYBR Green PCR Master Mix, Applied Biosystems) (primer sequences are available upon request). The experiments were performed in a subset of AML patients whose RNA material was available (23/23 of *UNCX*+ and 22/39 of *UNCX*-). The arithmetic mean and the pooled standard deviation of the two groups of patients were compared and the *UNCX*- cases were used as the calibrator and *GAPDH* as a reference. Statistical significance was analyzed in comparisons of *UNCX*+ and *UNCX*- cases by using the relative expression software tool REST<sup>4</sup>.

### *UNCX* transcription pattern analysis

To study the size of the *UNCX* 5'UTR region, two primer pairs (*UNCX* 5'UTR I and *UNCX* 5'UTR II; sequences are available upon request) were designed based on the human genome sequence corresponding to the 5'UTR sequence of the *Rattus norvegicus* gene (Accession No. D87748.1), because the rat gene sequence was the largest one found in the UCSC browser (GRCh37/hg19). These two primer pairs were used in RT-PCR assays on human total brain RNA. The *UNCX* 5'UTR I-F oligo was then used in combination with a reverse primer designed to target a region within *UNCX* exon 3 (*UNCX* ex3.1R) in assays on NB-4 cell line RNA.

Moreover, to detect the presence of *UNCX* alternative transcripts, apart from the “canonical” *UNCX* transcript (Accession No. NM\_001080461.1), we designed specific forward and reverse primers (BU176043-F, R, and AW148819-F, R; sequences are available upon request) to target the overlapping region of the 3 ESTs (BU176043, AW148819, and HY330599) mapped within *UNCX* intron II. Appropriate combinations of exon- and EST-specific primers were used in RT-PCR experiments performed with KAPA2G Robust HotStart (Kapabiosystems, Boston, MA), according to the manufacturer's instructions. The optimized PCR conditions were the following for all primer pairs used: 2 min at 95°C followed by 35 cycles of 20 s at 95°C, 15 s at 60°C,

and 100 s at 72°C. Subsequently, 1 µL of each PCR product was amplified by performing nested and semi-nested RT-PCR according to the following conditions: 5 s at 95°C followed by 25 cycles of 20 s at 95°C, 15 s at 60°C, and 50 s at 72°C, and then a final extension for 120 s at 72°C. A first run of RT-PCR was performed using the *UNCX* 5'UTR II-F and AW148819-R primer combination. Subsequently, the PCR products were subjected to nested PCR with these primer pairs: *UNCX* 5'UTR II-F + *UNCX* AR, *UNCX* AF + BU176043-R, and BU176043-F + AW148819-R.

All primers were designed using the Primer3 internet-based interface<sup>5</sup>, and were checked for specificity using the BLAT tool of the UCSC Human Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgBlat?hgsid=337267937&command=start>) (sequences are available upon request).

### Western blotting

Patient samples and cells lines were lysed using Passive Lysis Buffer (cod. E1941, Promega, Milan, Italy). Cell lysates were separated using SDS-PAGE and transferred to Hybond-C Extra nitrocellulose membranes (Amersham Biosciences, GE Healthcare Lifescience, Milan, Italy). Blots were probed with anti-*UNCX* antibodies (see main text) at 1:1000 dilution and an alkaline phosphatase-conjugated secondary antibody (Sigma-Aldrich S.r.l., Milan, Italy) at 1:5000 dilution. Antibody reactivity was visualized with nitroblue tetrazolium chloride (NBT) plus 5-bromo-4-chloro-3'-indolylphosphate p-toluidine salt (BCIP) in alkaline phosphatase buffer (Bio-Rad, Milan, Italy). We were unable to identify *UNCX* in the samples under study because of a lack of specific bands corresponding to the predicted molecular weight (53,690 Da, corresponding to 531 amino acids) of the protein (UniProtKB Accession No. A6NJT0) encoded by the transcript with GenBank Accession No. NM\_001080461 (data not shown). Identical results were obtained despite the use of two different anti-*UNCX* antibodies specific for the N-terminus and the C-terminus of the protein (ab105966, Abcam, Cambridge, UK, and AV47546, Sigma, respectively).

Western blotting experiments were also used to detect *UNCX* protein in U937 transfected cells as previously described<sup>6</sup>. Blotted membranes were pre-blocked with a solution containing 5% nonfat milk (Regilait, Saint-Martin-Belle-Roche, France) and then incubated with the primary antibody (1:1000, Anti-Flag M2 Antibody; cod. F3165 Sigma-Aldrich S.r.l., Milan, Italy) and a common secondary antibody conjugated to horseradish peroxidase (1:10000). Immunoreactive bands were detected using BM Chemiluminescence Blotting Substrate (Roche Diagnostics, Mannheim, Germany). To normalize the protein samples, vinculin expression was analyzed using a mouse anti-human vinculin monoclonal antibody (1:1000) (Merck Millipore, Darmstadt, Germany) (Online Supplementary Figure S6E).

### Methylation analysis of the *UNCX* canyon

We subjected 1 µg of genomic DNA to bisulfite conversion by using the EZ DNA Methylation Kit (Zymo Research, Orange, CA) according to the manufacturer's protocol, except that the conversion step consisted of 21 cycles at 95°C for 30 s and 50°C for 15 min. Universal unmethylated and methylated DNAs (Millipore, Billerica, MA) were used as internal controls. Moreover, a 477-bp amplicon for the *IGF2* region was included in the assay for quality assessment of bisulfite-treated DNA<sup>7</sup>. This protocol employs T7-promoter-tagged PCR amplification of bisulfite-converted DNA, followed by the generation of single-stranded RNA and subsequent base-specific cleavage by RNase A. Here, 20 nL of the cleavage fragments were nanodispensed onto 384-element silicon chips preloaded with the matrix. Mass spectra were collected using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS, MassARRAY mass spectrometer, Bruker-Sequenom, Billerica, MA), and the methylation ratios (MRs) of the spectra were generated using EpiTYPER software v.1.0 (Sequenom).

A 2.5-Kb genomic segment (chr7:1,270,388-1,272,932 - GRCh37/hg19), including the 5' upstream region and *UNCX* exon 1, was submitted to EpiDesigner BETA software (Sequenom) for prediction of CpG islands; for the analysis, 11 partially overlapping amplicons were selected that included 246 CpGs, 193 of which could be analyzed using this methodology. The analysis was performed in three *UNCX*<sup>+</sup> (no. 1-Dx, 3, and 4) and seven *UNCX*<sup>-</sup> (no. 5, 6, 7, 10, 11, 12, and 13) patients (Table 1), as well as in PB, which was used as a negative control. Moreover, extended regions spanning 10 and 19 Kb upstream (5') and downstream (3') of *UNCX* (chr7:1,261,900–1,295,600), respectively, were investigated to assess the methylation status of *UNCX* canyon borders; 20 amplicons containing 381 CpGs were selected for the analysis, and 318 CpGs could be analyzed using our method. We investigated a subset of five *UNCX*<sup>+</sup> (no. 1-Dx, 9, 16, 30, and 33) and seven *UNCX*<sup>-</sup> (no. 1-Rem, 6, 13, 21, 22, 32, and 49) AML patients (Table 1), along with a pool of CD34<sup>+</sup> CB cells as a negative control.

For statistical analysis, differences between groups were analyzed using conventional Student's *t* test or one-way analysis of variance (ANOVA). For comparisons, the MR values were submitted to angular transformation. Data

are displayed using Box and Whisker plots. Statistical calculations were performed using Statistics, version 8.0 (www.statistics.com). Amplicon details and primer sequences are available upon request.

### ***DNMT3A* mutational analysis**

An amplicon-sequencing assay (454 Life Sciences, Branford, CT) was used to assess *DNMT3A* mutational status, as previously described<sup>8</sup>, (exons 7–23 in 27/62 patients, according to sample availability). NGS data were analyzed using GS Variant Analyzer Software 2.5.3 (454 Life Sciences) and Sequence Pilot version 3.5.2 (JSI Medical Systems, Kippenheim, Germany). Known single-nucleotide polymorphisms were filtered out using NCBI dbSNP version 135<sup>9</sup>. Deleterious effects of mutations at the DNA level were predicted using MutationTaster software<sup>10</sup>.

### **Exon array experiments**

From 100 ng of total RNA per sample, cDNA was synthesized using the WT Expression Kit (Ambion Inc./Applied Biosystems, Santa Clara, CA). The cDNA was fragmented and labeled with biotin by using the GeneChip WT Terminal Labeling Kit (Affymetrix, Santa Clara, CA). Biotinylated targets were hybridized onto an Affymetrix GeneChip Human Exon 1.0 ST Array according to the manufacturer's instructions. Each array was washed and stained in an Affymetrix Fluidics Station 450 and scanned to generate a .CEL file by using an Affymetrix GeneChip Scanner 3000 7G with Command Console (AGCC) Software (Affymetrix). Affymetrix Expression Console Software (version 1.0) was used to perform quality assessment.

### **Exon array and TCGA data analysis**

To detect Differentially Expressed Genes (DEGs) between *UNCX*-positive and *UNCX*-negative sample groups, two-ways ANOVA with Benjamini-Hochberg False Discovery Rate (FDR) <0.05 was applied to both exon array and TCGA data. Fold changes (FC) and p-values (p) were used as filtering conditions for creating differentially expressed gene lists ( $|FC| > 1.5$  and  $p < 0.05$ ). Groups' correspondence between the patients of our cohort and the TCGA samples was assessed through sparse Partial Least Square–Discriminant Analysis (sPLS-DA). This technique belongs to the latent variables methods (e.g. PLS, Principal Component Regression) and faces the classification problem by Linear Discriminant Analysis (LDA). To test whether the overlap between the differentially expressed genes, obtained by the analysis of our exon array and of TCGA data, was not due by chance, the analysis was performed in a sub-cohort of TCGA AML samples, made of all 126 *UNCX*-TCGA-negative cases and 12 patients expressing *UNCX* over the 75<sup>th</sup> percentile. Starting from known class membership (i.e., *UNCX*-TCGA-negative and *UNCX*-TCGA-positive), sPLS-DA looked for a gene set signature that best discriminated *UNCX*-negative from *UNCX*-positive TCGA samples. Once obtained it, we performed Principal Component Analysis (PCA) on the samples of our internal cohort using only the gene fingerprint found in the previous step. Transcriptional separation between *UNCX*-positive and *UNCX*-negative for both datasets was represented graphically by means of Partek® Genomics Suite®, version 6.6, build 6.16.0812, 2016 Partek Inc., St. Louis, MO, USA. Both sPLS-DA and PCA were performed by the R package mixOmics version 6.1.1 (<http://mixomics.org/>).

### **Purification of CD34 stem/progenitor and precursor cells**

Mononuclear cells were isolated using Ficoll-Hypaque (Lymphoprep; Nycomed Pharma, Oslo, Norway) gradient separation. CD34<sup>+</sup> cells were separated using a magnetic cell-sorting procedure (EasySep Human CD34 positive selection kit; StemCell Technologies, Vancouver, BC, Canada). CD34<sup>+</sup> cell purity, assessed using flow cytometry, was  $\geq 95\%$  in all preparations (data not shown). After immunomagnetic separation, CD34<sup>+</sup> cells were seeded in 24-well plates at  $5 \times 10^5$  cells/mL in Iscove's modified Dulbecco's medium (GIBCO, Grand Island, NY) containing 20% human serum (Bio-Whittaker, Walkersville, MD) and stem cell factor (SCF) (50 ng/mL), Flt3-ligand (50 ng/mL), recombinant human thrombopoietin (TPO) (20 ng/mL), IL-6 (10 ng/mL), and IL-3 (10 ng/mL) (all from R&D Systems, Minneapolis, MN)<sup>2,11</sup>.

### **Retroviral vector construction and packaging**

We constructed the LUNCXIΔN retroviral vector expressing a *UNCX* full-length cDNA in the context of a bicistronic transcript driven by the viral LTR<sup>12</sup>. This vector encoded a truncated version of the low-affinity nerve growth factor receptor (ΔLNGFR), which was used as a reporter gene for estimating transduction efficiency in hematopoietic cells. The *UNCX* gene was excised from a Myc-DDK-tagged ORF clone of *Homo sapiens* UNC HB (Origene, Bologna, Italy) by using the *Bgl*III and *Pme*I restriction sites and cloned in the same sites of the LXIΔN vector<sup>12</sup>, which resulted in construction of the LUNCXIΔN retroviral vector. Packaging lines for LUNCXIΔN were generated through trans-infection in the ecotropic Phoenix cells and amphotropic GP+envAm12 cells, as previously described<sup>13</sup>. The U937 cell line was transduced with the LUNCXIΔN and

LXIAN retroviral vectors to test their efficiency. *UNCX* transduction efficiency was validated by flow cytometry (see flow cytometry analysis subsection) performed using an antibody against *NGFR* and by RT-qPCR (see above). In vitro-translated *UNCX* was detected through Western Blotting (performed as described above; Online Supplementary Figure S6E). In LUNCXIAN we detected a product of 75 KDa rather than 57 KDa as expected. This was probably due to a post-translational modification of the protein<sup>14</sup>.

#### **Hematopoietic cell transduction and purification**

Human cord blood (CB) CD34<sup>+</sup> progenitor cells, pre-activated for 48 h under previously described liquid-culture conditions (see Purification of CD34 stem/progenitor and precursor cells subsection), were transduced for 2–3 cycles of infection (6–8 h each) with viral supernatant on retronectin-coated plates (10 g/cm<sup>2</sup>). Transduced CD34<sup>+</sup> cells were purified by an immunomagnetic procedure in which mouse anti-human p75-NGFR monoclonal antibodies and tiny FACS-compatible magnetic nanoparticles were used in a column-free magnetic system (EasySep “Do-It-Your Self” Selection Kit, StemCell Technologies, Voden Medical Instruments spa, Milan, Italy) per the manufacturer’s guidelines. *UNCX* transduction efficiency of CD34<sup>+</sup> cells was monitored, in a set of three independent experiments, through flow cytometry analysis (Coulter Epics XL, Coulter Electronics Inc., Hialeah, FL) of  $\Delta$ LNGFR expression (positive range: 15–30%) performed using a biotinylated mouse anti-human p75-NGFR monoclonal antibody (BD Biosciences, San Diego, CA) (Online Supplementary Figure S6A-C). Antibody incubations were performed as previously described<sup>11</sup>.

#### **Colony-Forming Cell (CFC) assays**

LUNCXIAN- and LXIAN-transduced CD34<sup>+</sup> cells were NGFR-purified at Day 5 of liquid culture at 37°C in a humidified atmosphere with 5% CO<sub>2</sub>; these cells and NT cells were cultured in methylcellulose-based media, as previously described<sup>15</sup>. LXIAN cells were used as a control to exclude any  $\Delta$ LNGFR reporter-gene interference with the analyzed processes. Proliferation and differentiation rates of myeloid hematopoietic progenitors were determined at 14 days after seeding (approximately 400 CD34<sup>+</sup> infected cells). Results are the mean  $\pm$  standard error mean (SEM) from three independent experiments in triplicate.

#### **Growth curves**

LUNCXIAN, LXIAN, and NT cells were seeded at a density of  $\sim 200 \times 10^3$  cells/mL and grown under the same medium conditions as described above. The cells were counted on a Neubauer hemocytometer and the mean  $\pm$  SEM values were calculated from three independent experiments. For subsequent time points (Days 3, 5, and 8 PI), the entire cell population was transferred to plates featuring progressively larger surface areas to maintain the initial seeding density, and then counted as above. The relative cell number (in relation to the seeded density) was graphed to obtain the growth curves (Figure 4B).

#### **Flow cytometry analysis**

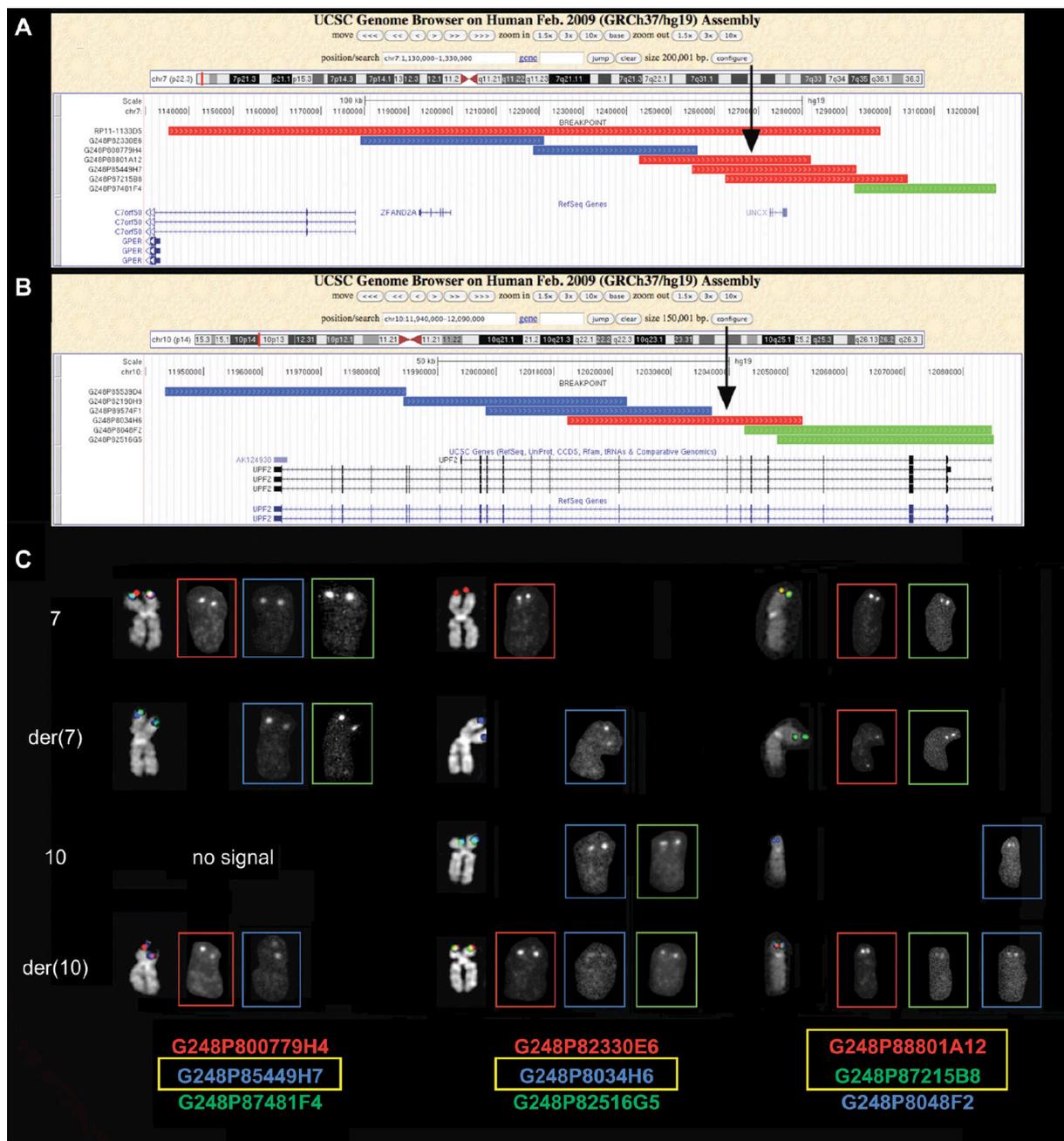
Surface antigen expression was detected using a phycoerythrin (PE)-conjugated CD34 monoclonal antibody (Miltenyi Biotec, Bergisch Gladbach, Germany). Antibody incubations were performed as previously described<sup>13</sup>. Samples were analyzed in terms of the positivity percentage by using a Coulter Epics XL-MCL flow cytometer. Results are the mean  $\pm$  SEM values from two independent experiments.

## Supplementary Results

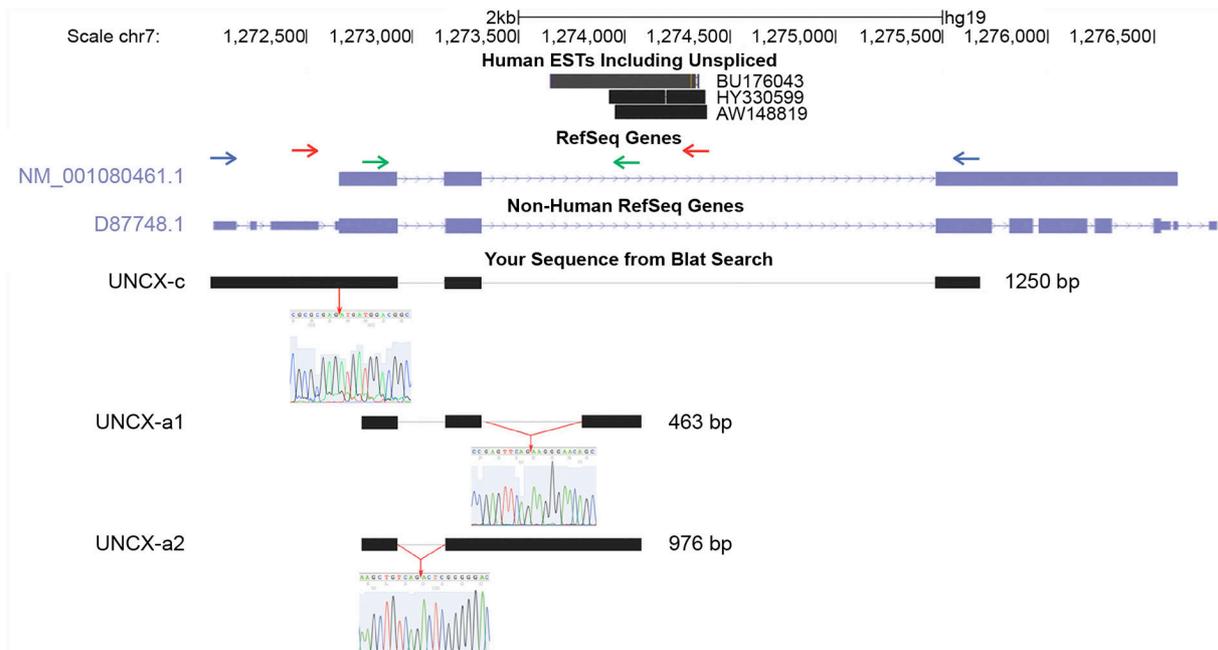
### ***UNCX* exhibits a complex expression pattern**

*UNCX* expression analysis showed that its transcription start site (TSS) mapped upstream to that reported by GRCh37/hg19 in RT-PCR analysis conducted using human RNA from brain cells with both *UNCX* 5'UTR I and *UNCX* 5'UTR II primer pairs (mapping 600 and 466 bp upstream of the *UNCX* TSS in the UCSC browser, respectively; data not shown), and with the *UNCX* 5'UTR I-F + *UNCX* ex3.1R primer combination [PCR product 1,250 bp in size, named *UNCX*-canonical (*UNCX*-c), GenBank KM587717] (Online Supplementary Figure S2). Moreover, *UNCX* encoded two previously unrecognized alternative transcript isoforms identified by RT-PCR products of 463 bp [*UNCX*-alternative 1 (*UNCX*-a1), GenBank KM587719] and 976 bp [*UNCX*-alternative 2 (*UNCX*-a2), GenBank KM587718] that were generated through the retention of distinct portions of intron II, as observed using appropriate exon-specific and EST-specific primer combinations (Online Supplementary Figure S2). *UNCX*-a1 and -a2 were both amplified in NB-4, OCI-AML3, and Y79 cell lines. *UNCX*-a2 was expressed in 17 AML cases (no. 1, 4, 5, 7, 9, 16, 27, 33, 34, 39, 45, 49, 50, 51, 52, 53, and 54), 15 AML cell lines (CMK, EOL-1, GDM-1, HNT-34, HT-93A, IMS-M1, KG-1, MOLM-16, MUTZ-8, OCI-AML-2, OCI-AML-5, OCI-M2, SKNO-1, TF-1, and THP-1), 2 CML cell lines (MEG-01 and K562), Jurkat cells, HEK293T cells, and normal total brain; by contrast, Cases 48 and 58 showed expression of only *UNCX*-a1 (data not shown). In silico translation of *UNCX*-a1 and *UNCX*-a2 displayed shortened *UNCX* proteins that harbored a truncated HB domain. Intriguingly, the identification of these alternative transcripts also in *UNCX*-patients, supports a potential post-transcriptional regulatory role of the transcripts as long non-coding RNAs (lncRNAs) or a dominant-negative effect of the truncated proteins<sup>16</sup>. Indeed, recent studies have reported that the destabilizing effect of intron retention within mRNAs results in the creation of premature stop codons at translation. Notably, these transcripts have regulatory and potentially oncogenic activity<sup>17,18</sup>. These RNA entities were recently described, in normal and malignant hematopoiesis, to regulate key transcription factors<sup>19-21</sup> and to act as either tumor suppressors [lincRNA-p21<sup>22</sup> and MEG3<sup>23,24</sup>] or oncogenes [ANRIL<sup>25,26</sup>]. However, because of the high GC content of the *UNCX* sequence, the full-length transcript isoforms could not be completely characterized and further in vitro experiments could not be performed to elucidate their function in gene regulation during myelopoiesis.

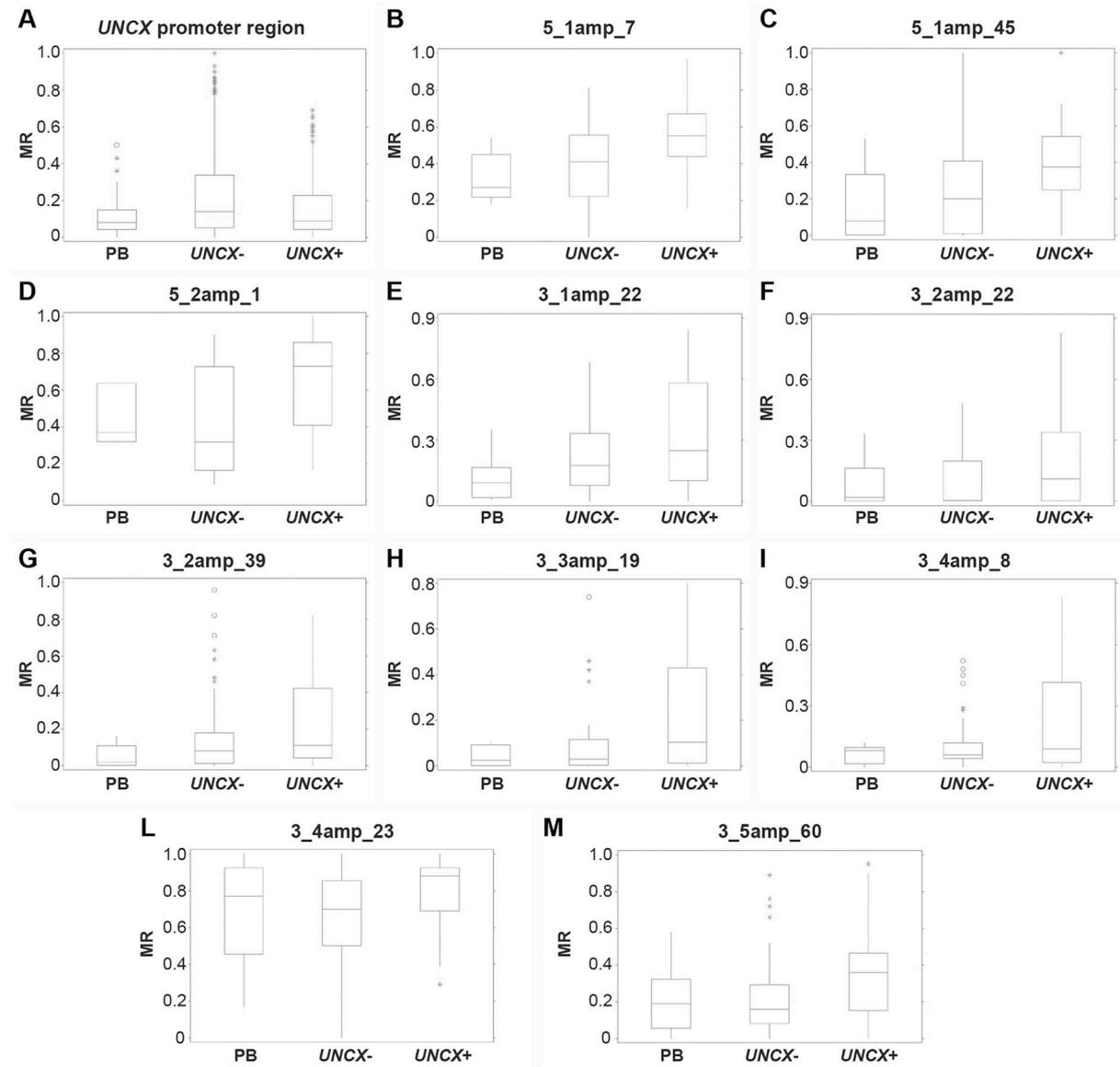
**Supplementary Figure S1. FISH analysis for mapping the breakpoints of t(7;10) translocation in Case 1-Dx.** (A) and (B) Contig maps of BAC clones encompassing the breakpoint regions in chromosome bands 7p22.3 and 10p14, respectively. Blue, red, and green bars correspond to translocated, split, and retained clones, respectively. Black arrows indicate the breakpoints on both chromosome 7 and chromosome 10. (C) Partial karyotype of case 1-Dx showing FISH co-hybridization experiment results. Each column shows merged, pseudocolored and original signals of normal and derivative chromosomes 7 and 10. In cohybridization experiments, probes were directly labeled with fluorescein (green), Cy3 (red), and Cy5 (blue). Chromosomes were counterstained by DAPI. The red, blue, and green rectangles correspond to the results obtained for the probes listed at the bottom of the figure. The yellow rectangles include the name of the fosmid clones showing splitting signals. Digital images were obtained using an automated fluorescence microscope (Leica DMRXA2) equipped with a cooled CCD camera. Cy3, fluorescein, Cy5, and DAPI were detected by separate filters (magnification x100), merged and pseudocolored by Photoshop software.



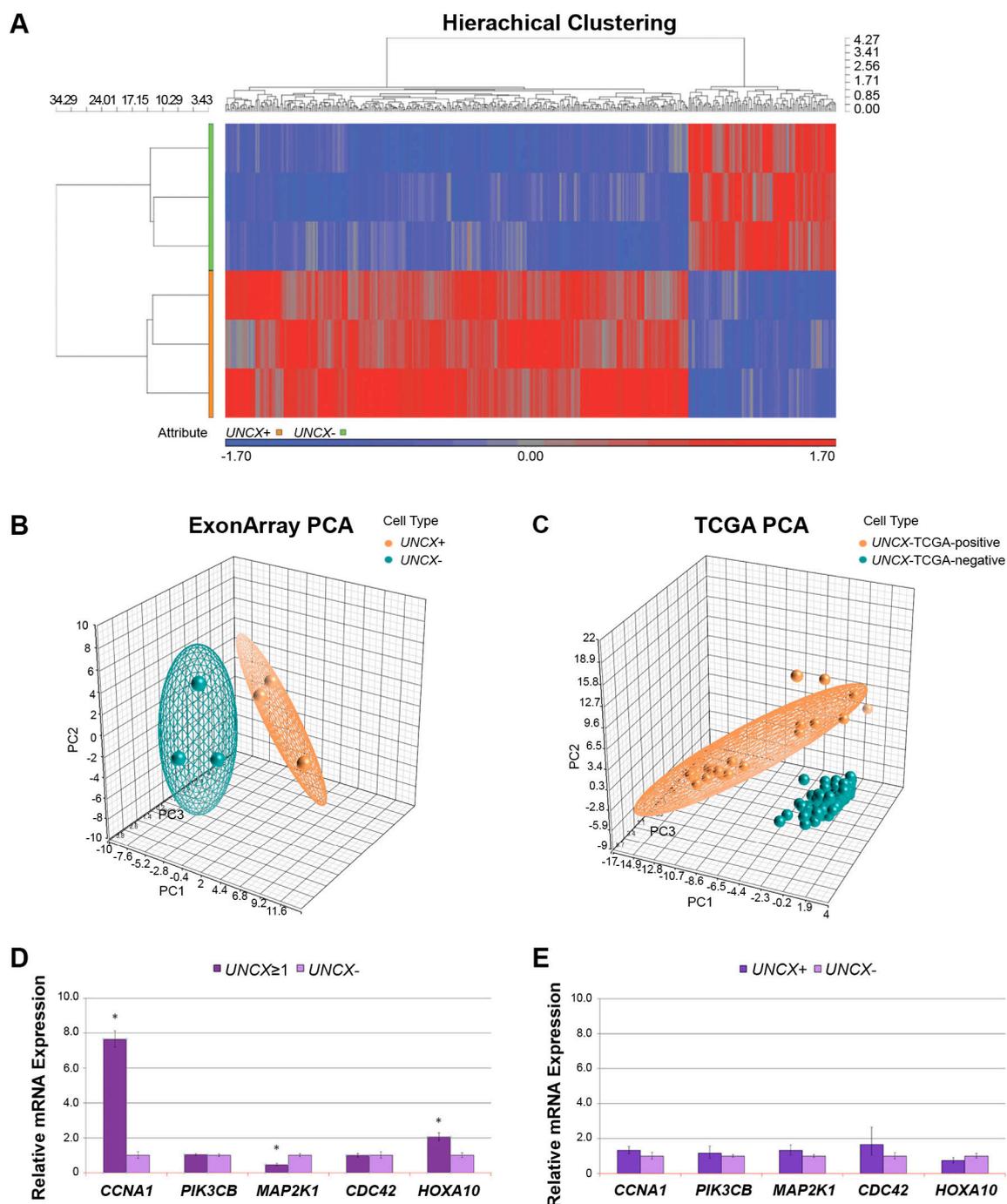
**Supplementary Figure S2. *UNCX* transcription pattern in AML patients and cell lines.** Schematic representation of the obtained RT-PCR products revealing the 5'UTR region (*UNCX*-c) and the alternative transcript isoforms of *UNCX* (*UNCX*-a1 and -a2). The primer pairs used in the study, *UNCX* 5'UTR I-F + *UNCX* ex3.1 R, *UNCX* 5'UTR II-F + AW148819-R, and *UNCX* A F + BU176043-R for *UNCX*-c, *UNCX*-a1, and *UNCX*-a2, are indicated by blue, red, and green arrows, respectively. The partial chromatograms indicate the junction between the newly identified 5'UTR sequence and exon 1 in *UNCX*-c, as well as the newly identified splicing junctions in *UNCX*-a1 and -a2.



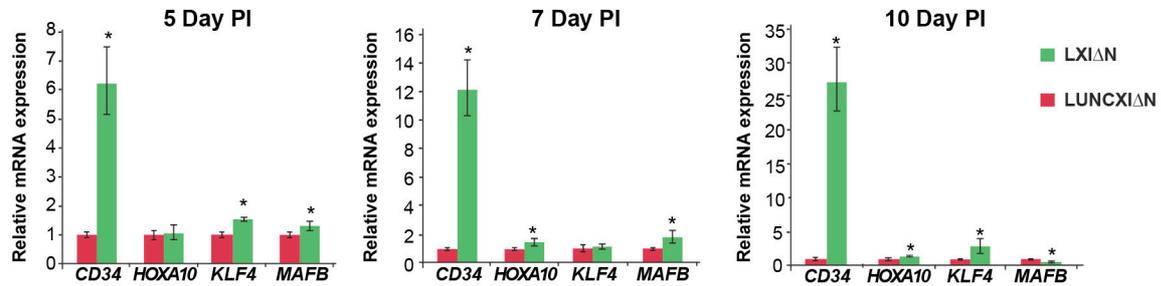
**Supplementary Figure S3. Methylation status evaluated at *UNCX* promoter region and 5' and 3' canyon borders.** (A) Box and Whisker plot illustrating the methylation ratios (MRs) of the promoter region in PB and *UNCX*<sup>+</sup> and *UNCX*<sup>-</sup> patients. Box and Whisker plot illustrating differences in MRs at the 5' (B-D) and 3' (E-M) *UNCX* canyon border among CB CD34<sup>+</sup> cells and *UNCX*<sup>+</sup> and *UNCX*<sup>-</sup> patients. Only the CpGs in the amplicons showing a significant difference between *UNCX*<sup>+</sup> and *UNCX*<sup>-</sup> samples are shown (see Supplementary Table S2). The boxes enclose the middle half of the data and are bisected by a line at the median value. The vertical line at the top and the bottom of each box indicates the range of “typical” data values. Extreme values are displayed as “\*” for possible outliers (values that are outside the box boundaries by >1.5 times the size of the box) and “O” for probable outliers (values that are outside the box boundaries by >3 times the size of the box). The experiments were performed once and only Case1-Dx was analyzed in duplicates.



**Supplementary Figure S4. Gene-expression profiling results obtained by matching exon array and TCGA datasets and RT-qPCR analyses of differentially expressed genes.** (A) Hierarchical clustering analysis of the gene-expression profile obtained for 3 *UNCX*<sup>+</sup> and 3 *UNCX*<sup>-</sup> samples. The columns and rows represent genes and samples, respectively. (B, C) Global views of gene expression obtained by performing PCA between *UNCX*<sup>+</sup> (orange) and *UNCX*<sup>-</sup> (green) samples in exon array (B) and TCGA (C) dataset, respectively. The analysis was performed using Partek Genomic Suite software with the default setting that includes a threshold for removing low background-level intensities. (D, E) RT-qPCR results of *CCNA1*, *PIK3CB*, *MAP2K1*, and *HOXA10* expression in patients with *UNCX* expression values  $\geq 1$  (D) and in the overall cohort of *UNCX*<sup>+</sup> patients (E) versus *UNCX*<sup>-</sup>. \*: Two-tailed  $p \leq 0.05$ . All experiments were performed once, and each sample was analyzed in triplicate.

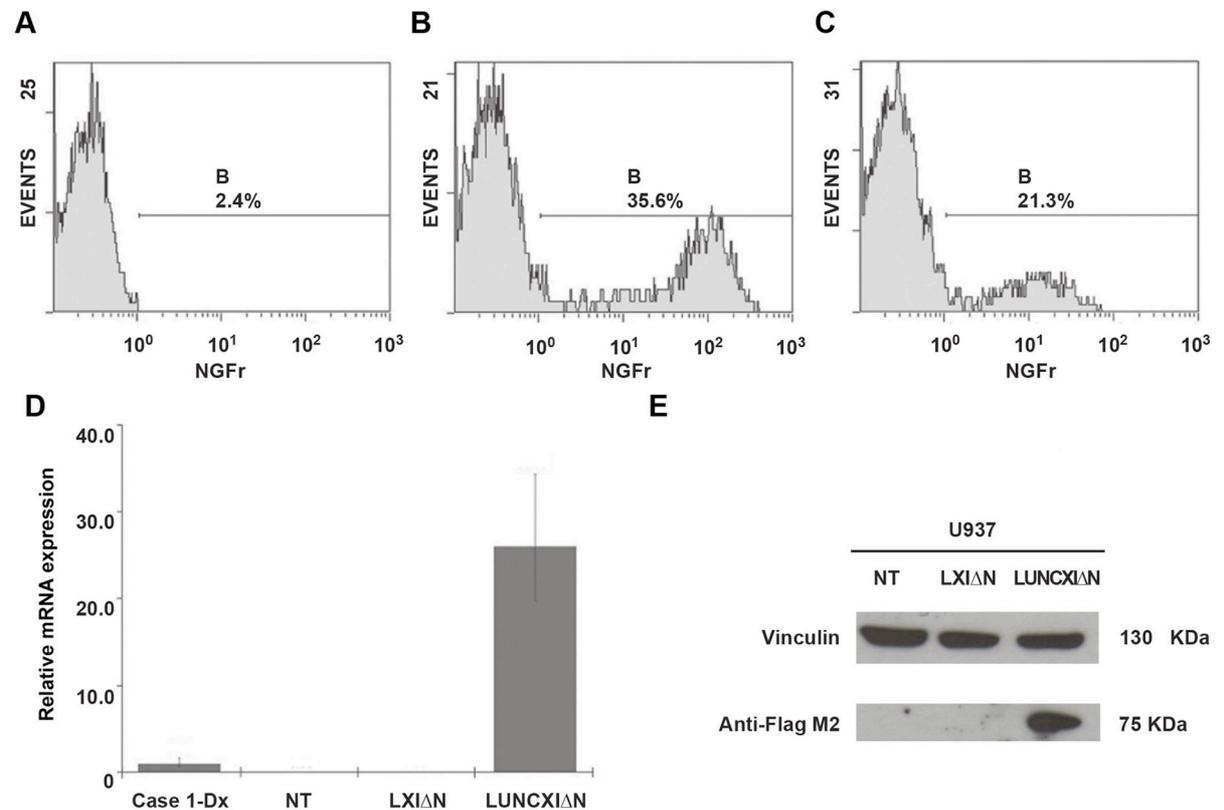


**Supplementary Figure S5. RT-qPCR evaluation of selected differentiation markers, performed at different times post-infection (PI).** RT-qPCR analyses performed on RNA extracted from cells at Days 5, 7, and 10 PI for selected genes encoding specific surface markers and transcription factors implicated in hematopoiesis. The experiments were performed once and each sample was analyzed in triplicate. \*: Two-tailed  $p \leq 0.05$ .  $2^{-\Delta\Delta Ct}$  values are reported in the table below the graphs.



	5 Day PI		7 Day PI		10 Day PI	
	LXI $\Delta$ N	LUNCXI $\Delta$ N	LXI $\Delta$ N	LUNCXI $\Delta$ N	LXI $\Delta$ N	LUNCXI $\Delta$ N
<i>CD34</i>	1.0	6.23	1.0	12.13	1.0	27.10
<i>HOXA10</i>	1.0	1.09	1.0	1.46	1.0	1.28
<i>KLF4</i>	1.0	1.55	1.0	1.16	1.0	2.73
<i>MAFB</i>	1.0	1.32	1.0	1.78	1.0	0.48

**Supplementary Figure S6. Validation of *UNCX* Retrovirus-Mediated Transduction in Human Purified Cord Blood CD34+ cells and in U937 cell line.** (A), (B), and (C) Flow cytometry analysis of  $\Delta$ NGFR expression by using a murine anti-human p75-NGFR mAb conjugated with biotin, performed at Day 2 PI in NT (A), LXI $\Delta$ N (B), and LUNCXI $\Delta$ N (C) CD34+ cells; the line indicates the percentage of positive cells. (D) RT-qPCR results of *UNCX* expression in NT, LXI $\Delta$ N, and LUNCXI $\Delta$ N CD34+ cells, versus Case 1-Dx. (E) Western blotting of the transduced U937 cell line; vinculin was used as a control.



## Supplementary Tables Legend

**Supplementary Table S1.** *UNCX* expression levels in AML cell lines (A), in additional hematological and solid tumor cell lines and normal tissues (B) included in the study, and in normal hematopoietic lineages at different stage of maturation according to Blueprint Epigenome data portal (C).

**Supplementary Table S2.** Clinical characteristics of *UNCX*-TCGA-positive and *UNCX*-TCGA-negative cohorts [including (A) and excluding (B) M3 FAB type]; *FLT3*, *NPM1*, *RAS* and *IDH* mutational status of *UNCX*-TCGA-positive and *UNCX*-TCGA-negative cohorts (C).

**Supplementary Table S3.** Methylation status of CpGs located in the most 5' upstream (A) and 3' downstream (B) regions of *UNCX* gene in our AML patients; intraclass Spearman correlation between FPKM and Beta values of TCGA AML cohorts (C); results of the genome-wide DNA methylation differential analysis among *UNCX*-TCGA-positive and *UNCX*-TCGA-negative AML patients (D); *DNMT3A* NGS results (E).

**Supplementary Table S4.** PCA and Ingenuity analyses: list of differentially expressed genes between *UNCX*+ and *UNCX*- patients (A); Ingenuity analysis results\_Pathways (B); Ingenuity analysis results\_Networks (C); Ingenuity analysis results\_Diseases and Functions (D).

## Supplementary references

1. Drexler HG. *The Guide to Leukemia-Lymphoma Cell Lines*. 2nd ed, 2010.
2. Montanari M, Gemelli C, Tenedini E, et al. Correlation between differentiation plasticity and mRNA expression profiling of CD34+-derived CD14- and CD14+ human normal myeloid precursors. *Cell death and differentiation*. 2005;12(12):1588-1600.
3. Storlazzi CT, Fioretos T, Surace C, et al. MYC-containing double minutes in hematologic malignancies: evidence in favor of the episome model and exclusion of MYC as the target gene. *Hum Mol Genet*. 2006;15(6):933-942.
4. Pfaffl MW, Horgan GW, Dempfle L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic acids research*. 2002;30(9):e36.
5. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*. 2000;132(365-386).
6. Grande A, Manfredini R, Pizzanelli M, et al. Presence of a functional vitamin D receptor does not correlate with vitamin D3 phenotypic effects in myeloid differentiation. *Cell death and differentiation*. 1997;4(6):497-505.
7. Ehrlich M, Zoll S, Sur S, van den Boom D. A new method for accurate assessment of DNA quality after bisulfite treatment. *Nucleic acids research*. 2007;35(5):e29.
8. Roller A, Grossmann V, Bacher U, et al. Landmark analysis of DNMT3A mutations in hematological malignancies. *Leukemia*. 2013;27(7):1573-1578.
9. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic acids research*. 2001;29(1):308-311.
10. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7(8):575-576.
11. Gemelli C, Montanari M, Tenedini E, et al. Virally mediated MafB transduction induces the monocyte commitment of human CD34+ hematopoietic stem/progenitor cells. *Cell death and differentiation*. 2006;13(10):1686-1696.
12. Urbinati F, Lotti F, Facchini G, et al. Competitive engraftment of hematopoietic stem cells genetically modified with a truncated erythropoietin receptor. *Hum Gene Ther*. 2005;16(5):594-608.
13. Grande A, Montanari M, Tagliafico E, et al. Physiological levels of 1alpha, 25 dihydroxyvitamin D3 induce the monocytic commitment of CD34+ hematopoietic progenitors. *J Leukoc Biol*. 2002;71(4):641-651.
14. Dimatteo C, Marucci A, Palazzo A, et al. Role of somatomedin-B-like domains on ENPP1 inhibition of insulin signaling. *Biochimica et biophysica acta*. 2013;1833(3):552-558.
15. Manfredini R, Zini R, Salati S, et al. The kinetic status of hematopoietic stem cell subpopulations underlies a differential expression of genes involved in self-renewal, commitment, and engraftment. *Stem Cells*. 2005;23(4):496-506.
16. Shiote Y, Ouchida M, Jitsumori Y, et al. Multiple splicing variants of Naf1/ABIN-1 transcripts and their alterations in hematopoietic tumors. *Int J Mol Med*. 2006;18(5):917-923.
17. Abdel-Samad R, Zalzali H, Rammah C, et al. MiniSOX9, a dominant-negative variant in colon cancer cells. *Oncogene*. 2011;30(22):2493-2503.
18. Beghini A, Ripamonti CB, Peterlongo P, et al. RNA hyperediting and alternative splicing of hematopoietic cell phosphatase (PTPN6) gene in acute myeloid leukemia. *Hum Mol Genet*. 2000;9(15):2297-2304.
19. Garzon R, Volinia S, Papaioannou D, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proc Natl Acad Sci U S A*. 2014;111(52):18679-18684.
20. Han BW, Chen YQ. Potential pathological and functional links between long noncoding RNAs and hematopoiesis. *Sci Signal*. 2013;6(289):re5.
21. Wang H, Li W, Guo R, et al. An intragenic long noncoding RNA interacts epigenetically with the RUNX1 promoter and enhancer chromatin DNA in hematopoietic malignancies. *Int J Cancer*. 2014;135(12):2783-2794.
22. Huarte M, Guttman M, Feldser D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010;142(3):409-419.
23. Benetatos L, Hatzimichael E, Dasoula A, et al. CpG methylation analysis of the MEG3 and SNRPN imprinted genes in acute myeloid leukemia and myelodysplastic syndromes. *Leuk Res*. 2010;34(2):148-153.
24. Zhang X, Rice K, Wang Y, et al. Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. *Endocrinology*. 2010;151(3):939-947.

25. Yap KL, Li S, Munoz-Cabello AM, et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell*. 2010;38(5):662-674.
26. Yu W, Gius D, Onyango P, et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*. 2008;451(7175):202-206.