haematologica

Journal of The Ferrata Storti Foundation

De novo UBE2A mutations are recurrently acquired during chronic myeloid leukemia progression and interfere with myeloid differentiation pathways

by Vera Magistroni, Mario Mauri, Deborah D'Aliberti, Caterina Mezzatesta, Ilaria Crespiatico, Miriam Nava, Diletta Fontana, Nitesh Sharma, Wendy Parker, Andreas Schreiber, David Yeung, Alessandra Pirola, Sara Redaelli, Luca Massimino, Paul Wang, Praveen Khandelwal, Stefania Citterio, Michela Viltadi, Silvia Bombelli, Roberta Rigolio, Roberto Perego, Jacqueline Boultwood, Alessandro Morotti, Giuseppe Saglio, Dong-Wook Kim, Susan Branford, Carlo Gambacorti-Passerini, and Rocco Piazza

Haematologica 2019 [Epub ahead of print]

Citation: Vera Magistroni, Mario Mauri, Deborah D'Aliberti, Caterina Mezzatesta, Ilaria Crespiatico, Miriam Nava, Diletta Fontana, Nitesh Sharma, Wendy Parker, Andreas Schreiber, David Yeung, Alessandra Pirola, Sara Redaelli, Luca Massimino, Paul Wang, Praveen Khandelwal, Stefania Citterio, Michela Viltadi, Silvia Bombelli, Roberta Rigolio, Roberto Perego, Jacqueline Boultwood, Alessandro Morotti, Giuseppe Saglio, Dong-Wook Kim, Susan Branford, Carlo Gambacorti-Passerini, and Rocco Piazza. De novo UBE2A mutations are recurrently acquired during chronic myeloid leukemia progression and interfere with myeloid differentiation pathways.
Haematologica. 2019; 104:xxx
doi:10.3324/haematol.2017.179937

De novo UBE2A mutations are recurrently acquired during chonic myeloid leukemia progression and interfere with myeloid differentiation pathways

Vera Magistroni [1*], Mario Mauri [1*], Deborah D'Aliberti [1], Caterina Mezzatesta [1], Ilaria Crespiatico [1], Miriam Nava [1], Diletta Fontana [1], Nitesh Sharma [1], Wendy Parker [2], Andreas Schreiber [2], David Yeung [2,3], Alessandra Pirola [4], Sara Readelli [1], Luca Massimino [1], Paul Wang [2], Praveen Khandelwal [1], Stefania Citterio [5], Michela Viltadi [1], Silvia Bombelli [1], Roberta Rigolio [1], Roberto Perego [1], Jacqueline Boultwood [6,7], Alessandro Morotti [8], Giuseppe Saglio [8], Dong-Wook Kim [9], Susan Branford [2,3,10], Carlo Gambacorti-Passerini [1,11#] and Rocco Piazza [1#]

1 Department of Medicine and Surgery, University of Milano Bicocca, Monza, Italy
2 Center for Cancer Biology, SA Pathology, Adelaide, Australia
3 University of Adelaide, Australia
4 GalSeq s.r.l., viale Italia 46, Monza, 20900, Italy
5 Department of Bioscience and Biotechnology, University of Milano Bicocca, Milano, Italy
6 Bloodwise Molecular Haematology Unit, John Radcliffe Hospital, University of Oxford, UK
7 NIHR Biomedical Research Centre, Oxford, UK
8 Department of Clinical and Biological Sciences, San Luigi Hospital, University of Turin, Turin, Italy
9 Department of Hematology, Catholic University, Seoul, Korea
10 University of South Australia, Australia
11 Hematology and Clinical Research Unit, San Gerardo Hospital, Monza, Italy
*VM and MM contributed equally to this work
#RP and CGP contributed equally to this work

Running Title
UBE2A somatic variants in CML progression

Corresponding Authors
Vera Magistroni, vera.magistroni@unimib.it
Rocco Piazza, rocco.piazza@unimib.it

Abstract word count: 162
Main text word count: 3310
Tables: 2; Figures: 4; Supplemental files: 1

Acknowledgments
The authors would like to thank Manuela Carrera and Giuliana Laurenza for technical assistance. This work was supported by Associazione Italiana Ricerca sul Cancro (IG-14249 to CGP, IG-17727 to RP, IG-22082 to RP), by the European Union's Horizon 2020 Marie Skłodowska-Curie Innovative Training Networks (ITN-ETN) with grant agreement No.: 675712CGP and by Giovani Ricercatori #GR-2011-02351167 to A.M. CGP is a member of the European Research Initiative for ALK-Related Malignancies (www.erialcl.net). JB acknowledges support from Bloodwise-UK.

ABSTRACT

Despite the advent of tyrosine kinase inhibitors, a proportion of chronic myeloid leukemia patients in chronic phase fails to respond to Imatinib or to second generation inhibitors and progress to blast crisis. Limited improvements in the understanding of the molecular mechanisms responsible for chronic myeloid leukemia transformation from chronic phase to the aggressive blast crisis were achieved until now. We present here a massive parallel sequencing analysis of 10 blast crisis samples and of the corresponding autologous chronic phase controls which reveals, for the first time, recurrent mutations affecting the ubiquitin-conjugating enzyme E2A gene (*UBE2A*, formerly *RAD6A*). Additional analyses on a cohort of 24 blast crisis, 41 chronic phase as well as 40 acute myeloid leukemia and 38 atypical chronic myeloid leukemia patients at onset confirmed that *UBE2A* mutations are specifically acquired during chronic myeloid leukemia progression with a frequency of 16.7% in advanced phases. *In vitro* studies show that the mutations here described cause a decrease in UBE2A activity, leading to an impairment of myeloid differentiation in chronic myeloid leukemia cells.

INTRODUCTION

Chronic Myeloid Leukemia (CML) is a myeloproliferative disorder with an incidence of 1-2 cases per 100000/year and is characterized by the presence of the *BCR-ABL1* fusion gene, the product of the reciprocal translocation between chromosomes 9 and 22[1]. After the translocation, the coding regions of *BCR* and *ABL1* genes are juxtaposed, leading to an enhanced ABL1 tyrosine kinase activity[2]. CML is a multi-step disease, evolving from a mild and easy to control form, called chronic phase (CP) to a very aggressive and incurable acute phase called blast crisis (BC). The majority of CML-CP patients are successfully treated with drugs able to impair BCR-ABL1 kinase activity (Tyrosine Kinase inhibitors-TKIs), thus confirming the central role of the oncogenic fusion protein in CML pathogenesis[3, 4]. However, a fraction of these patients fails to respond to the treatment (primary resistance) or becomes resistant after an initial response[4-6]. The persistence of BCR-ABL1 activity typically drives the progression to the advanced phase of the disease within 3-5 years. One of the open issues in CML concerns the dissection of the molecular mechanisms driving the transformation to BC, commonly considered as a heterogeneous disease at molecular level[7-9]. BC is mainly characterized by the rapid expansion of the differentiation-arrested BCR-ABL1-positive blast cells[10], therefore resembling an acute leukemia. In the major part of BC cases (~70%), blasts maintain myeloid features, while in 20-30% blast lineage is lymphoid. BCR-ABL1 expression, which increases during CML progression in conjunction with BCR transcription, seems to have a prominent role in this process, hyperactivating proliferative and anti-apoptotic signals and inducing genetic instability[5, 11, 12]. Previous reports showed the existence of a heterogeneous molecular signature among distinct BC patients[5, 7, 8]. However, these data were limited by the scarcity of matched CP/BC samples, due to the infrequent progression to BC after the advent of TKIs. Here we analyzed 10 paired CP/BC samples through a whole-exome sequencing (WES) approach, identifying somatic variants specific for BC progression since not present in the

autologous CP controls. Along with several mutations previously identified as BC driver events[5, 7, 13], we detected, for the first time, recurrent BC-specific mutations occurring on *UBE2A* gene. These data suggest that the appearance of UBE2A variants in CML cells could contribute to BC progression through the impairment of myeloid differentiation.

METHODS

Cell lines

The BA/F3-BCR-ABL1 and 32Dcl3-BCR/ABL1 cell lines were generated and maintained as described in[14, 15]. K562 and 293FT were purchased from DSMZ (Braunschweig, Germany) and Thermo-Fisher-Scientific (Waltham, MA, USA) respectively, and were maintained following manufacturer instructions.

Patients

Diagnosis and staging were performed according to the World Health Organization WHO-2008 classification[16]. Peripheral Blood (PB) or Bone Marrow (BM) of 10 matched CML chronic phase/blast crisis samples, 31 CP-CML, 14 AP/BC-CML, 38 atypical-CML (aCML) and 40 AML were collected at diagnosis and after obtaining written informed consent approved by the institutional ethics committee. The study was conducted in accordance with the Declaration of Helsinki. Samples were prepared as described in[17].

Whole Exome Sequencing (WES)

Genomic DNA (gDNA) was extracted from purified cells with PureLink Genomic DNA kit (Thermo-Fisher-Scientific). 1 µg of gDNA from each sample was fragmented (500bp) with a Diagenode-Bioruptor sonicator system (Diagenode, Belgium) and processed according to the standard Illumina protocol. The Illumina TruSeq Exome Enrichment kit (Illumina Inc., San Diego, USA) was used to enrich the genomic libraries for the exonic regions and samples were sequenced as described in Supplemental Material.

Plasmids, transfections and lentiviral infections

BA/F3_BCR-ABL1-positive cells were transfected with pMIGR1_UBE2A vectors (Supplemental Material) as in [14] and were analyzed for GFP positivity with a FACSAria flow cytometer (BD Bioscience, San Jose, USA) and FACS-sorted when transfection efficiency was lower than 85 %.

32Dcl3-BCR/ABL1 cells were electroporated by using a Gene Pulser® II Electroporation System (BIORAD) with pMIGR1_UBE2A WT and I33M vectors as described in[14] .

To obtain stable UBE2A WT or I33M cell lines, GFP positive population was FACS-sorted with a MoFlo Astrios cell sorter equipped with Summit 6.3 software (both from Beckman Coulter, Miami, FL, USA).

For UBE2A silencing, K562 cells were infected with lentivirus obtained from MISSION-shRNA pLKO.1-based vectors (TRCN0000320625) (Sigma-Aldrich, Missouri, USA) and packaged using 293FT cell line. As a control, a pLKO.1MISSION non-target control vector (SHC002) (Sigma-Aldrich) was used. After infection K562 cells were maintained in 2µg/mL puromycin for selection of silenced (K562_shUBE2A) and control cells (K562_shNC).

Quantitative Real-Time PCR (RT-qPCR)

Total RNA was extracted using Trizol (Thermo-Fisher-Scientific) following manufacturer instructions. 1µg of total RNA was used to synthesize cDNA using Reverse Transcription Reagents (Thermo-Fisher-Scientific) after pre-treatment with DNAseI (Thermo-Fisher-Scientific) to avoid contamination from genomic DNA. Real-Time-Quantitative PCR (RT-qPCR) was performed using TaqMan® Brilliant II QPCR Master Mix (Agilent technologies, CA, USA) on a Stratagene-MX3005P (Agilent-technologies) under standard conditions. The housekeeping gene glucoronidase beta gene (GUSB) was used as an internal reference[12]. TaqMan® Gene Expression Assays (Thermo-Fisher-Scientific) were used (Supplementary Table S1).

***In vitro*-translation and ubiquitination assay**

In-vitro translation of UBE2A proteins was performed with 1-Step Human Coupled IVT Kit-DNA (Thermo-Fischer-Scientific) following manufacturer instructions. For ubiquitination assay we incubated 15µg of UBE2A proteins with 1µg of GST-Ubiquitin (Enzo-Life-Sciences, NY, USA), 0.2 ng of Ubiquitin Activating Enzyme (E1) (Enzo-Life-Science), 2mM ATP, Energy Regeneration Solution

(BostonBiochem, MA, USA), 2mM MgCl2, 2mM KCl, 16µg of BA/F3_BCR-ABL1 whole cell lysate in 50mM TrisHCl (ph7.5). The reactions were incubated for 20 min at 37°C. The products were analyzed by Western Blotting.

For the enzymatic activity of WT and mutated UBE2A 15µg of UBE2A in vitro synthetized protein were used. The AMP-Glo Assay (Promega catalog v5011) was used in order to quantify the amount of AMP generated by the ubiquitin conjugation machinery, composed by 170ng/µl Ubiquitin Protein, 15ng/µl UBA1 and 50 µM ATP (SignalChem).

The production of AMP from ATP is directly proportional to the enzymatic activity of the ubiquitination machinery and therefore it was used to measure the ubiquitination in presence of WT and mutated UBE2A. The AMP signal was detected using the AMP Detection Solution (Promega) and a TECAN reading plate (Infinite F200Pro TECAN).

Neutrophilic differentiation

For induction of neutrophilic differentiation 32Dcl3-BCR/ABL1 UBE2A WT and I33M cells were treated as previously described[18].

32Dcl3-BCR/ABL1 cells expressing UBE2A WT or I33M were seeded at a density of $2x10^5$ cells per milliliter and cultured in presence of imatinib mesylate (1 µM final concentration) in combination with human recombinant GCSF (10ng/ml) or IL3 (0,5ng/ML).

At day 3 and 6 cells were analyzed using FACS for CD11b surface expression and imaged with confocal microscopy (Supplementary material and methods).

RESULTS

Single nucleotide variants (SNVs) acquired during CML progression

Genomic DNA (gDNA) from matched CP/BC samples was obtained for each patient at diagnosis (CP) and after progression to BC. Initially whole exome sequencing (WES) data from 10 patients

were analyzed. CP samples were used as baseline controls for each patient to identify somatic variants selectively occurring in BC (Table 1), thus allowing the recognition of molecular events occurring exclusively upon CML progression. By using this approach we identified mutations on genes already associated with BC, such as *RUNX1*, *IKZF1*, *NRAS*, *ASXL1* and *ABL1*[7, 13]. A total of 41 non-synonymous single nucleotide variants (SNVs) and small indels were identified, with a mean of 4.1 mutations/patient acquired upon BC progression. Of these events, 31 were transitions, 7 transversions and 3 indels, with the C>T substitution being the most frequent (63.4%) (Supplementary Figure 1). In one patient (pt #7) no acquired exonic SNVs could be detected during CML progression.

Analysis of SNVs data showed the presence of two recurrently mutated genes in this cohort: ABL1, with mutations F486S, E255V and T315I occurring on the *BCR-ABL1* fusion gene and leading to TKI resistance (30%, C.I. 95% 0.574, 0.026), and *UBE2A* (Xq24), an E2-ubiquitin conjugating enzyme required for post-replicative DNA damage repair [15] (20%, C.I. 95% 0.447, 0.000), which has never been previously reported as mutated in CML patients. UBE2A mutations occurred on two non-contiguous residues: D114V and I33M (Table 1, Table 2). Patient #3 (male) showed an *UBE2A* variant frequency of 93%, as expected given that the gene is localized on the X chromosome. Patient #8 (female) carried a heterozygous UBE2A mutation (mutation ratio: 39%). The high mutation ratio observed in both patients suggests that UBE2A is present in the dominant BC clone (Table 1).

UBE2A mutations are recurrent and acquired in late CML

The evidence of recurrent, somatic *UBE2A* mutations has never been reported in BC cases, however they were previously found in other clonal disorders both of solid and hematopoietic origin, confirming their potential role in tumor progression[19]. To further characterize the pattern

and the frequency of *UBE2A* mutations in a larger cohort of patients, we sequenced 31 additional CML CP samples at onset, 14 AP/BC, 40 Acute Myeloid Leukemia and 38 aCML samples. No evidence of *UBE2A* mutations could be found in CP, AML and aCML samples, while in two AP/BC samples somatic *UBE2A* variants D114Y and M34fs were detected. Globally, acquired *UBE2A* mutations could be detected in a total of 16.7% (4/24) advanced (AP/BC) CML cases (95% C.I. 1.78-31.62) (Table 2).

UBE2A mutations affect protein activity

Polyphen-2 (http://genetics.bwh.harvard.edu/pph/)[20], DANN[11] and FATHMM-MKL[21] analyses revealed that all the UBE2A variants identified were potentially damaging, as also suggested by the presence of a N-terminal frameshift variant (M34fs) in one of the patients (Table 2). To gain insight into the functional role of UBE2A mutations, we stably transfected the BA/F3_BCR-ABL1 cell line[14] with the wild-type (WT) and the mutated UBE2A variants, I33M and D114V. The level of UBE2A expression in stable transfectants was verified both at protein (Figure 1A) and mRNA levels (Figure 1B).

The analysis of the levels of ubiquitin-conjugated H2A, a known UBE2A substrate[22], in total cell lysate revealed a decreased H2A ubiquitination for both UBE2A variants compared to WT (Figure 1C), with the effect of I33M being more prominent. In line with these findings, suggesting a decreased UBE2A activity for both variants, ubiquitination assay performed with in vitro translated WT and mutated UBE2A proteins confirmed a decrease in ubiquitin-conjugating activity for mutants compared to the WT form (Figure 1D). To further support this indication, we developed a new in vitro assay based on the measurement of the AMP concentration as a proxy to assess the overall level of ubiquitination. This test was performed in presence of GST-Ubiquitin and of the E1 Ubiquitin activating enzyme UBA1 together with WT or mutated UBE2A. This analysis revealed a

significant decrease in ATP consumption, and therefore in ubiquitin-conjugating activity, for UBE2A mutants compared to the WT form (Figure 1E: enzyme specific activity assay, 1.55 ($p<0.01$) and 1.53 ($p<0.01$) fold decrease in UBE2A D114V and I33M AMP concentration compared to UBE2A WT).

Transcriptome analysis of UBE2A cellular models shows significant perturbation of downstream pathways related to myeloid development

To identify the gene networks perturbed by the UBE2A knock-out, stable lentiviral UBE2A silencing models were generated (Figure 2A-B) in the human myeloid K562 cell line (K562-shUBE2A and K562-shNC cells for UBE2A silencing and scrambled control, respectively; Supplementary Figure 2). Whole-Transcriptome analysis (RNA-Seq) highlighted the presence of 168 differentially expressed genes, with 117 of them being downregulated and 51 upregulated (Figure 2C). Gene Set Enrichment Analysis (GSEA) showed significant enrichment for ontologies related to myeloid differentiation (Figure 2D) and neural development (Supplementary Figure 3). RT-qPCR on K562shNC/shUBE2A cell lines on a set of five differentially expressed genes (*ITGB4, RDH10, CLEC11A, CSF3R, RAP1GAP*) confirmed RNA-Seq data (Figure 2E). Interestingly, the colony stimulating factor 3 receptor (CSF3R) was potently downregulated in shUBE2A both at mRNA (12.5-fold down-regulation; Figure 2E) and protein levels (Figure 2F), hence suggesting that its down-modulation may play a role in the differentiation block that is ultimately responsible for the onset of the BC. Immunoblot analysis on CP/BC mononuclear cells from patient #3, which acquired the D114V-UBE2A mutation in BC phase, confirmed CSF3R down-modulation (Figure 2G).

To confirm the expression signature identified in the UBE2A silencing models, we stably overexpressed UBE2A WT and I33M in the 32Dcl3-BCR/ABL1 murine myeloid cell line (Supplementary Figure 4). In line with the expression profile shown in K562 UBE2A silenced cells,

also in these cell lines we observed a comparable modulation in the previous analyzed set of five differentially expressed genes (Figure 2H; ITGB4 1.52 $p<0.01$, RDH10 1.30 $p<0.05$, CLEC11A 2.84 $p<0.01$, CSF3R 0.25 $p<0.05$, RAP1GAP 0.27 $p<0.01$; data are reported as fold-change in UBE2A I33M compared to UBE2A WT), therefore supporting the hypotheses that 1) UBE2A mutations modulate the activity of the target protein in a loss of function manner and 2) UBE2A mutations may probably act as dominant negative variants. Comparison of our signature with known BC data (GEO _GSE47927 - HSC data were used for BC vs CP calculation) indicated the presence of a moderate positive linear correlation (Supplementary Figure 5; $R^2 = 0.234$). Notably, CSF3R expression level resulted to be markedly decreased also in the reference BC database, with a Log2 fold-change of -2.19. Globally, these data indicate that UBE2A mutations are directly responsible for the modulation of CSF3R, ITGB4, RDH10, CLEC11A and RAP1GAP expression. This hypothesis is also corroborated by the 32Dcl3 cell models.

UBE2A activity is involved in myeloid differentiation

Erythrocytes and megakaryocyte differentiation can be induced in K562 cells by treating with hydroxyurea or phorbol 12-myristate 13-acetate (PMA), respectively[23, 24]. Treatment of UBE2A-silenced K562 cells with hydroxyurea showed a significant delay in the ability to differentiate into erythrocytes, as assessed by Glycophorin A (GYPA-CD235a) expression levels when compared with the scrambled control (Figure 3A-C: relative CD235a expression compared to shNC Fold Change at day 0: 0.54 ± 0.13, $p<0.001$; day 1: 0.70 ± 0.12, $p<0.05$; day 3: 0.61 ± 0.15, $p<0.05$). Fluorescence-activated cell sorting analysis (FACS) showed a 45% decrease of GYPA surface expression in silenced cells compared to controls after 24hrs of treatment (Figure 3B). In line with these findings, induction of hemoglobin (HBB) production was almost completely suppressed in shUBE2A cells (6.6 fold relative decrease of HBB mRNA level at 24hrs of treatment: $p<0.001$)

further confirming the negative effect of UBE2A silencing on erythroid differentiation (Figure 3C). Similarly, treatment of K562 cells with the megakaryocytic-inducing agent PMA showed significant impairment of megakaryocyte differentiation in shUBE2A cells, as assessed by the expression levels of CD41 (33% downregulation of surface protein expression, $p< 0.001$) and CD44 (38% decrease at mRNA level; $p< 0.01$) after PMA treatment (Supplementary Figure 6).

Neutrophilic differentiation was similarly tested in the 32Dcl3 BCR/ABL1 cell lines overexpressing UBE2A WT or I33M. Treatment of the UBE2A I33M cell line with GCSF + IL-3 showed a delay in neutrophilic differentiation, as assessed by CD11b expression levels when compared with both UBE2A WT or control (Figure 4A). Cells treated with IL-3 alone were used as an internal control. Fluorescence-activated cell sorting analysis (FACS) showed no difference in CD11b surface expression in UBE2A I33M cell line compared to controls after 3 days of treatment but showed a 37% decrease at day 6 which was also confirmed by confocal microscopy analysis (Figure 4B).

DISCUSSION

In line with previous results[5, 7, 9], our analysis done on matched CP/BC CML samples showed considerable somatic heterogeneity in BC phase. In all the samples we detected a very low number of acquired SNVs, corresponding to an average of 4.1 non-synonymous mutations per patient, a frequency far below the average reported for other hematopoietic neoplasms, such as Acute Myeloid Leukemia (AML: 7.8) and Chronic Lymphocytic Leukemia (CLL: 11.9)[25]. This can be in part explained by the characteristics of our analysis, where somatic variants occurring in BC were filtered against those in CP, therefore filtering-out all the driver and passenger variants pre-existing the evolution to BC. All BC samples showed the prevalence of transition events and, in particular, of C:G>T:A substitutions, accounting for 66.7% of all the SNVs (Supplementary Figure 1). Approximately 85% of the C:G>T:A transitions were part of a CpG dinucleotide. Cytosines in CpG sites are known to be affected by a high mutation rate, caused by a spontaneous deamination of methylated cytosines [26]. This mutation pattern is also in accordance with a BCR-ABL1 dependent mutation signature, characterized by inhibition of the mismatch repair system (MMR) and by accumulation of reactive oxygen species (ROS), as previously reported[27]. Mutations in *RUNX1* and *IKZF1*, both involved in hematopoietic differentiation, have already been detected in the advanced stages of CML[7] and are confirmed here as specific markers for BC progression. Along with this, the *XPO1* gene (exportin-1) mutated here in a single patient with the E571K substitution, is also frequently mutated in clonal hematological disorders, with the E571K mutation widely represented in chronic lymphocytic leukemia[28]. SNVs analysis showed the presence of a recurrent mutation affecting the *UBE2A* gene (Xq24) (pt#3 and pt#8). UBE2A is an E2-ubiquitin conjugating enzyme which has never been found mutated in CML. Interestingly the two patients harboring *UBE2A* mutations lacked any recognizable CNAs (Table 2). WES and targeted resequencing of a broader cohort showed that somatic *UBE2A* mutations are found in a significant fraction (16.7%)

of advanced CML phases, thus confirming the initial exome analysis and suggesting a driver role for UBE2A loss of function during disease progression.

The *Saccaromyces Cerevisiae* UBE2A homolog Rad6 participates in DNA repair, sporulation and cell cycle regulation[29]; in mammals a role for UBE2A in the regulation of transcription and chromatin reorganization through post-translational histone modifications has been recently hypothesized[30]. Germline mutations of the *UBE2A* gene in humans have been associated with the X-linked Nascimento-type intellectual disability syndrome[31-33]. In order to understand the effect of UBE2A mutations in a BCR-ABL1-positive model, we tested the activity of exogenous UBE2A both in the WT or mutated forms (D114V and I33M) in BA/F3 BCR-ABL1-positive cell lines. We observed a reduced amount of mono-ubiquitinated histone H2A, a known UBE2A substrate, after over-expression of mutated UBE2A compared to the WT (Figure 1C), which indicates that the UBE2A mutations analyzed in this study decrease the activity of the enzyme. This result has been further confirmed by *in-vitro* assays for Ubiquitination and enzymatic activity on total cell lysates (Figure 1D and E), thus providing evidence of a damaging effect of the two mutations on UBE2A function. Accordingly, one of the four variants identified in our cohort is a N-terminal frameshift mutation, thus supporting this hypothesis. This evidence is further strengthened by the distribution of UBE2A mutations throughout the entire protein, a pattern that is more common for genes undergoing inactivation. Mutations in the UBE2A paralog UBE2B were not detected in this study, which suggests a specific role for UBE2A in chronic myeloid leukemia. Stable silencing of UBE2A in the BCR-ABL-positive K562 cell line or overexpression of the I33M mutated form in a BCR-ABL1-positive 32Dcl3cl3 myeloid cell line showed profound down-modulation of CSF3R, a critical regulator of myeloid lineage differentiation and development[34, 35]. CSF3R, also known as granulocyte colony-stimulating factor receptor (GCSFR), is a member of the hematopoietin receptor superfamily[35] and plays a key role in promoting neutrophilic differentiation but may also

support the development of different types of hematopoietic progenitors[34]. This suggests a potential role for CSF3R modulation in the suppression of myeloid differentiation in BC. Although the precise mechanism by which UBE2A controls CSF3R expression is at present unknown, our data suggest that UBE2A-mediated CSF3R regulation occurs at transcriptional level. Alteration of CSF3R transcription could occur either by a direct activity of UBE2A on CSF3R promoter through epigenetic mechanisms[36] or indirectly by UBE2A-mediated ubiquitination of specific transcription factors. Further studies will be necessary to clarify this process and to establish the relevance of CSF3R deregulation in the impairment of CML cells differentiation. In line with these findings we showed that impairment of UBE2A function induces a delay in the differentiation of K562 and 32Dcl3-BCR/ABL1 cells after PMA, hydroxyurea or GCSF treatment, suggesting an important role for UBE2A as a modulator of myeloid differentiation.

In conclusion, in this work we identified recurrent, somatic *UBE2A* mutations occurring in a significant fraction of advanced CML cases. We propose that the acquisition of somatic UBE2A mutations affects myeloid developmental pathways, promoting a differentiation blockade. Further studies will be required to thoroughly dissect the molecular mechanisms responsible for these effects and to define possible therapeutic strategies for UBE2A-mutated BC-CML cases.

References

1. Heisterkamp N, Stam K, Groffen J, de Klein A, Grosveld G. Structural organization of the bcr gene and its role in the Ph' translocation. Nature. 1985;315(6022):758-761.
2. Daley GQ, Van Etten RA, Baltimore D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science. 1990;247(4944):824-830.
3. Gambacorti-Passerini C, Antolini L, Mahon FX, et al. Multicenter independent assessment of outcomes in chronic myeloid leukemia patients treated with imatinib. J Natl Cancer Inst. 2011;103(7):553-561.
4. Hochhaus A, Larson RA, Guilhot F, et al. Long-Term Outcomes of Imatinib Treatment for Chronic Myeloid Leukemia. N Engl J Med. 2017;376(10):917-927.
5. Perrotti D, Jamieson C, Goldman J, Skorski T. Chronic myeloid leukemia: mechanisms of blastic transformation. J Clin Invest. 2010;120(7):2254-2264.
6. Gambacorti-Passerini CB, Gunby RH, Piazza R, Galietta A, Rostagno R, Scapozza L. Molecular mechanisms of resistance to imatinib in Philadelphia-chromosome-positive leukaemias. Lancet Oncol. 2003;4(2):75-85.
7. Grossmann V, Kohlmann A, Zenger M, et al. A deep-sequencing study of chronic myeloid leukemia patients in blast crisis (BC-CML) detects mutations in 76.9% of cases. Leukemia. 2011;25(3):557-560.
8. Boultwood J, Perry J, Zaman R, et al. High-density single nucleotide polymorphism array analysis and ASXL1 gene mutation screening in chronic myeloid leukemia during disease progression. Leukemia. 2010;24(6):1139-1145.
9. Johansson B, Fioretos T, Mitelman F. Cytogenetic and molecular genetic evolution of chronic myeloid leukemia. Acta Haematol. 2002;107(2):76-94.
10. Kantarjian HM, Keating MJ, Talpaz M, et al. Chronic myelogenous leukemia in blast crisis. Analysis of 242 patients. Am J Med. 1987;83(3):445-454.
11. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics. 2015;31(5):761-763.
12. Marega M, Piazza RG, Pirola A, et al. BCR and BCR-ABL regulation during myeloid differentiation in healthy donors and in chronic phase/blast crisis CML patients. Leukemia. 2010;24(8):1445-1449.
13. Mullighan CG, Miller CB, Radtke I, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature. 2008;453(7191):110-114.
14. Puttini M, Coluccia AM, Boschelli F, et al. In vitro and in vivo activity of SKI-606, a novel Src-Abl inhibitor, against imatinib-resistant Bcr-Abl+ neoplastic cells. Cancer Res. 2006;66(23):11314-11322.
15. Piazza RG, Magistroni V, Gasser M, et al. Evidence for D276G and L364I Bcr-Abl mutations in Ph+ leukaemic cells obtained from patients resistant to Imatinib. Leukemia. 2005;19(1):132-134.
16. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood. 2009;114(5):937-951.
17. Piazza R, Valletta S, Winkelmann N, et al. Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. Nat Genet. 2013;45(1):18-24.
18. Schuster C, Forster K, Dierks H, et al. The effects of Bcr-Abl on C/EBP transcription-factor regulation and neutrophilic differentiation are reversed by the Abl kinase inhibitor imatinib mesylate. Blood. 2003;101(2):655-663.
19. de Miranda NF, Georgiou K, Chen L, et al. Exome sequencing reveals novel mutation targets in diffuse large B-cell lymphomas derived from Chinese patients. Blood. 2014;124(16):2544-2553.

20. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 2002;30(17):3894-3900.
21. Shihab HA, Rogers MF, Gough J, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics. 2015;31(10):1536-1543.
22. Sung P, Prakash S, Prakash L. The RAD6 protein of Saccharomyces cerevisiae polyubiquitinates histones, and its acidic domain mediates this activity. Genes Dev. 1988;2(11):1476-1485.
23. Kim KW, Kim SH, Lee EY, et al. Extracellular signal-regulated kinase/90-KDA ribosomal S6 kinase/nuclear factor-kappa B pathway mediates phorbol 12-myristate 13-acetate-induced megakaryocytic differentiation of K562 cells. J Biol Chem. 2001;276(16):13186-13191.
24. Park JI, Choi HS, Jeong JS, Han JY, Kim IH. Involvement of p38 kinase in hydroxyurea-induced differentiation of K562 cells. Cell Growth Differ. 2001;12(9):481-486.
25. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. Science. 2013;339(6127):1546-1558.
26. Duncan BK, Miller JH. Mutagenic deamination of cytosine residues in DNA. Nature. 1980;287(5782):560-561.
27. Stoklosa T, Poplawski T, Koptyra M, et al. BCR/ABL inhibits mismatch repair to protect from apoptosis and induce point mutations. Cancer Res. 2008;68(8):2576-2580.
28. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011;475(7354):101-105.
29. Shekhar MP, Lyakhovich A, Visscher DW, Heng H, Kondrat N. Rad6 overexpression induces multinucleation, centrosome amplification, abnormal mitosis, aneuploidy, and transformation. Cancer Res. 2002;62(7):2115-2124.
30. Roest HP, Baarends WM, de Wit J, et al. The ubiquitin-conjugating DNA repair enzyme HR6A is a maternal factor essential for early embryonic development in mice. Mol Cell Biol. 2004;24(12):5485-5495.
31. Haddad DM, Vilain S, Vos M, et al. Mutations in the intellectual disability gene Ube2a cause neuronal dysfunction and impair parkin-dependent mitophagy. Mol Cell. 2013;50(6):831-843.
32. Budny B, Badura-Stronka M, Materna-Kiryluk A, et al. Novel missense mutations in the ubiquitination-related gene UBE2A cause a recognizable X-linked mental retardation syndrome. Clin Genet. 2010;77(6):541-551.
33. Nascimento RM, Otto PA, de Brouwer AP, Vianna-Morgante AM. UBE2A, which encodes a ubiquitin-conjugating enzyme, is mutated in a novel X-linked mental retardation syndrome. Am J Hum Genet. 2006;79(3):549-555.
34. Yang FC, Tsuji K, Oda A, et al. Differential effects of human granulocyte colony-stimulating factor (hG-CSF) and thrombopoietin on megakaryopoiesis and platelet function in hG-CSF receptor-transgenic mice. Blood. 1999;94(3):950-958.
35. Cosman D. The hematopoietin receptor superfamily. Cytokine. 1993;5(2):95-106.
36. Kim J, Guermah M, McGinty RK, et al. RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell. 2009;137(3):459-471.
37. Wu J, Huen MS, Lu LY, et al. Histone ubiquitination associates with BRCA1-dependent DNA damage response. Mol Cell Biol. 2009;29(3):849-860.
38. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671-675.
39. Piazza R, Ramazzotti D, Spinelli R, et al. OncoScore: a novel, Internet-based tool to assess the oncogenic potential of genes. Sci Rep. 2017;7(7):46290.

Table 1: Single Nucleotide Variants (SNVs) and indels identified by exome sequencing in BC samples and absent in the paired CP control (*: stop codon; OncoScore is a text-mining tool that scores genes according to their association with cancer based on available biomedical literature; higher scores correspond to a stronger association with cancer)

Patient		Disease Progression	Time to BC from CP at diagnosis (months)	Gene name (aa substitution)	Mutation ratio	OncoScore[39]
1	F	Lymphoid	1	RTP2(A190V)	53%	10.22
				KCNH3(A314V)	59%	5.61
2	F	Myeloid	59	FAT4(R1698W)	57%	57.19
				FUT3(R354C)	48%	30.21
				RUNX1(K194N)	68%	73
3	M	Myeloid	74	SMARCA4(A945T)	55%	48.69
				UBE2A(D114V)	93%	25.66
				ABL1(F486S)	50%	85.8
4	F	Myeloid	41	PTPN11(G503V)	41%	41.25
				AMER3(R709H)	41%	34.16
				LAMA2(P1025S)	49%	10.79
				GRIN3A(R1024*)	41%	5.95
				SMC5(L1102*)	27%	24.13
				MESDC2(E130del)	45%	28.07
				CCDC40(S17L)	34%	8.89
5	F	nd	14	NRAS(Q61R)	32%	81.65
				DEFB119(R42H)	46%	0
				IKZF1(N159S)	44%	72.76
				AK8(R125H)	67%	18.39
6	M	Lymphoid	52	-	-	
7	M	Lymphoid	-	PPT1(V168A)	39%	8.58
				MDH1B(A272T)	47%	0
				GPR98(R1745C)	45%	5.54
				CEL(E216Q)	46%	18.97
				LRP4(D449N)	54%	12.28
				CYP2B6(R145W)	40%	15.77
				BCR(F615W)	44%	81.02
8	F	Myeloid	59	ASXL1(G641_fs)	37%	77.48
				EPB41L3(P963L)	37%	76.11
				FGFR4(V262M)	39%	46.29
				UBE2A(I33M)	39%	25.66
9	F	Lymphoid	4	ABL1(E255V)	28%	85.8
				BARD1(G527_fs)	28%	83.5
				BSN(R3264H)	26%	3.32
				EFCAB4B(V643M)	31%	20.44
				KRT7(R339W)	41%	70.25
				AP5M1(D289N)	38%	0
				XPO1(E571K)	32%	47.33
10	F	Myeloid	10	HMCN1(S1371L)	53%	16.1
				ABCA13(T2019M)	53%	28.17
				ABL1(T315I)	45%	85.8

Table 2: UBE2A SNVs and indels identified in BC samples and absent in the paired CP control

Chr	Position	Ref	Var	Codon	AA Change	Polyphen 2 HDIV	DANN Score	Fathmm MKL	phastCons7 Vertebrate
chrX	119574955	A	G	ATA->ATG	Ile33Met	D	0.992	N	0.999
chrX	119583137	A	T	GAT->GTT	Asp114Val	D	0.993	D	1
chrX	119583136	G	T	GAT->TAT	Asp114Tyr	D	0.996	D	1
chrX	119574957		Ins AT	ATG->ATATG	M34fs	n.a.	n.a.	n.a.	n.a.

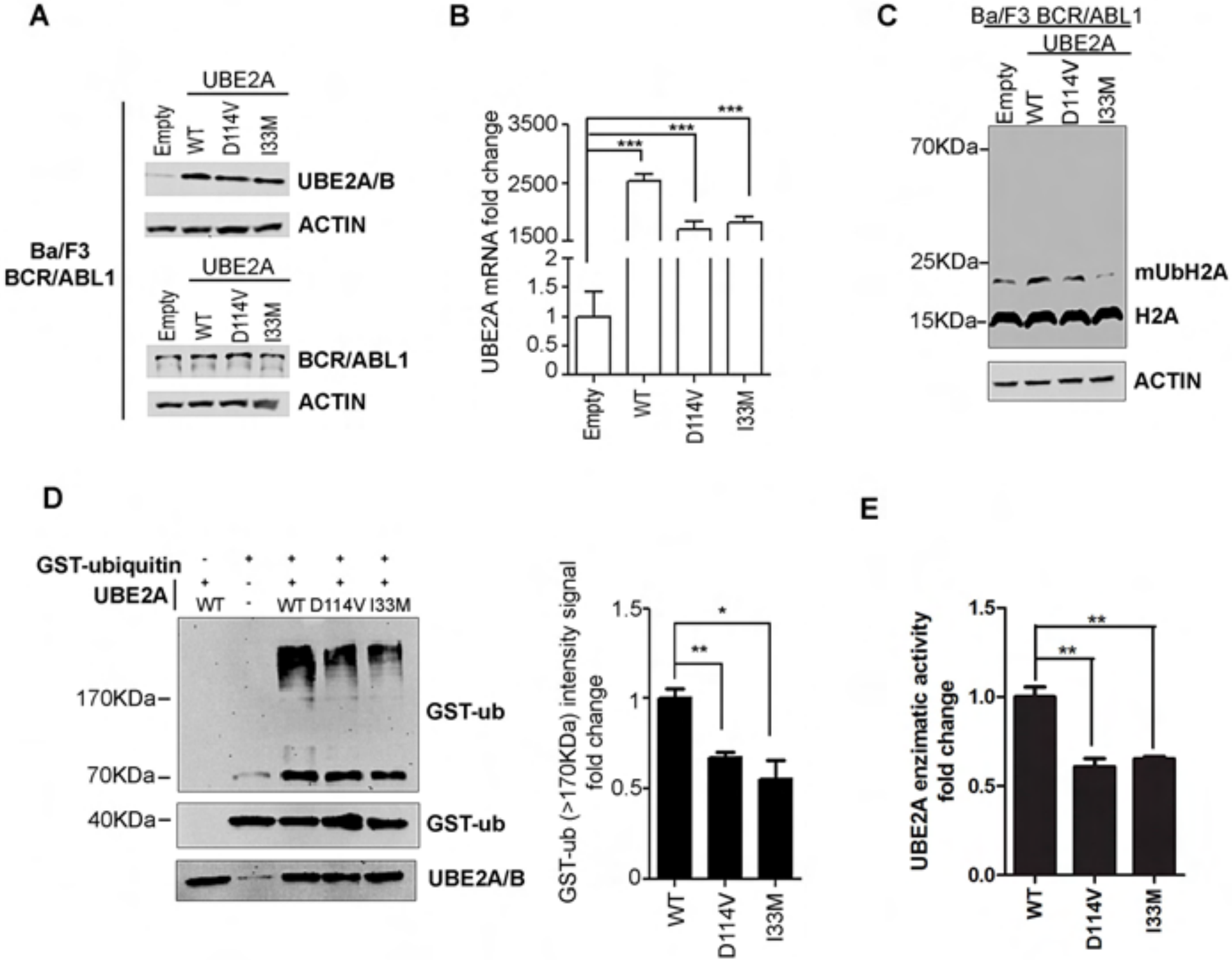
Figure legends

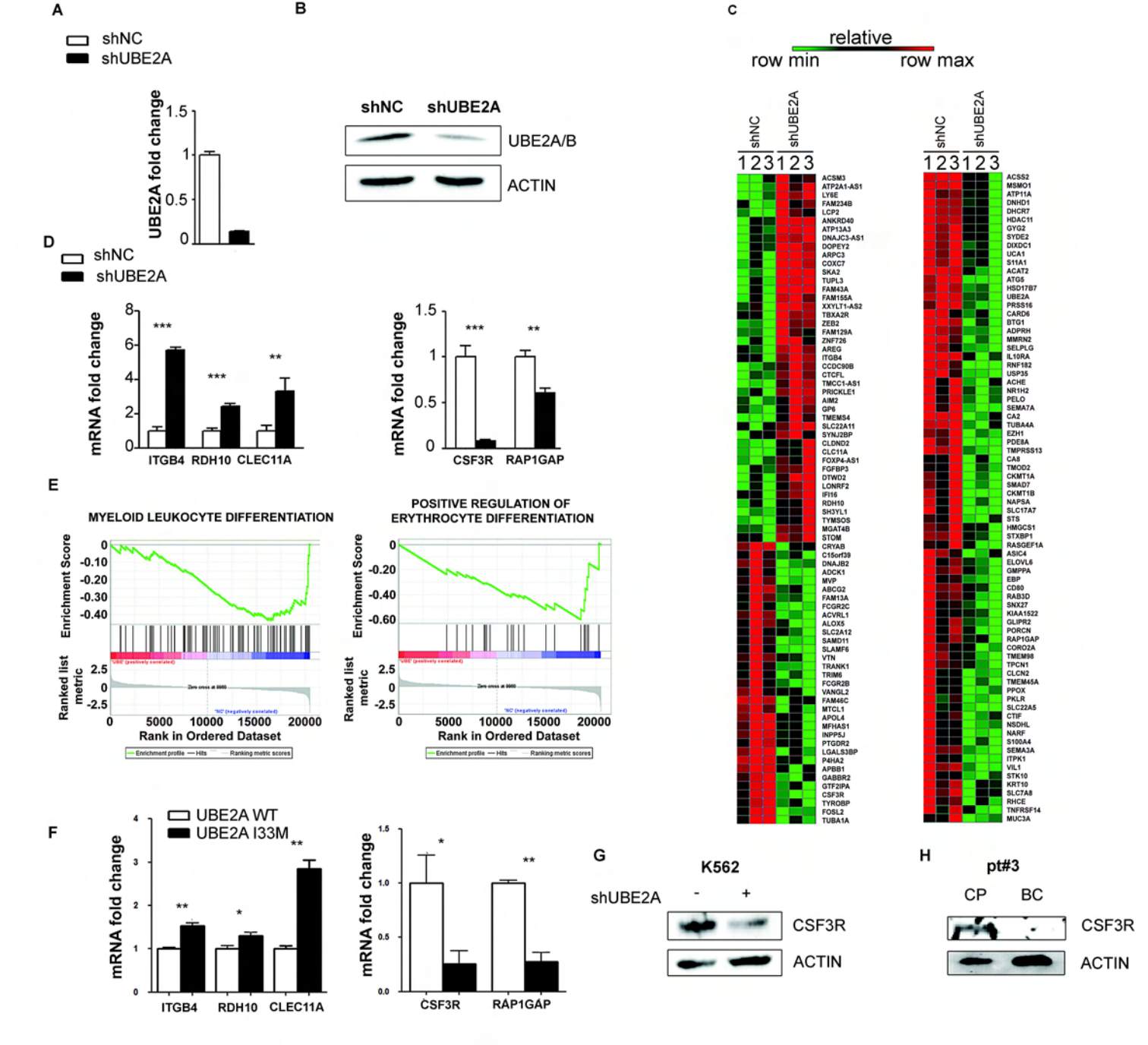
Figure 1. Activity of UBE2A mutants. **A**) Western Blot analysis of total cell lysates from BA/F3_BCR-ABL cell lines stably transfected with pMIGR-UBE2A vectors encoding for wild-type (WT) or mutated (D114V or I33M) UBE2A. Empty vector has been used as negative control. **B**) RT-qPCR of total RNA extracted from BA/F3_BCR-ABL_pMIGR/UBE2A cell lines. The values are normalized on the EMPTY cells (*** p<0.001). **C**) Western Blot of total cell lysates from BA/F3_BCR-ABL_pMIGR/UBE2A cell lines. The signal at 14KDa corresponds to histone H2A. The signal at ~23KDa corresponds to monoubiquitinated histone H2A (mUbH2A) according to[37]. **D**) Western Blot analysis of the *in-vitro* ubiquitination reaction performed with *in-vitro* translated UBE2A (WT and mutated forms) and GST-ubiquitin on total BA/F3_BCR-ABL lysate. The graph on the right shows the densitometric analysis of GST-ub signal (>170KDa) from three independent experiments obtained with ImageJ software[38]. The fold change is obtained normalizing the signal on the WT sample (WT vs D114V * p=0.022; WT vs I33M ** p=0.0069). **E**) Histogram showing the enzymatic activity of in vitro expressed UBE2A using the AMP Glow assay (WT vs D114V * p=0,0056; WT vs I33M ** p=0.0036).

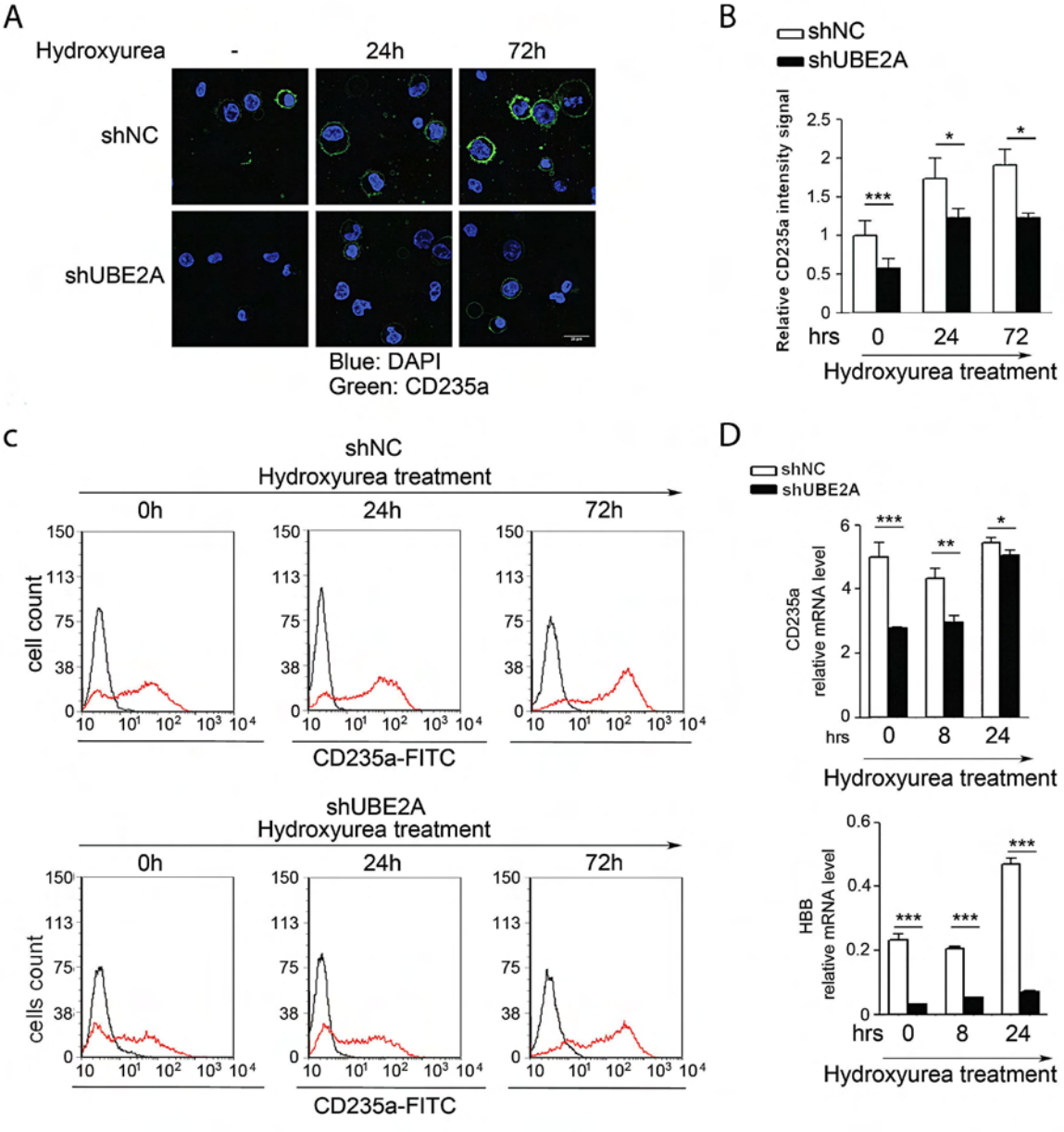
Figure 2. UBE2A silencing in K562 cells. **A**) RT-qPCR analysis of total RNA extracted from K562 cell lines infected with a lentiviral based system for UBE2A silencing (shNC: scrambled negative control; shUBE2A: UBE2A silenced cells). Values are normalized on shNC cells (*** p<0.0001). **B**) Western Blot analysis of total cell lysates from K562_shNC and K562_shUBE2A cells. **C**) Heat map of RNA-sequencing data showing color-coded expression levels of differentially expressed genes in three distinct populations of K562-shUBE2A compared to control (shNC). **D**) RT-qPCR analysis in K562 cell lines of a subset of differentially expressed genes identified by RNA-sequencing. **E**) Gene set enrichment analysis (GSEA) of the shUBE2A transcriptome. **F**) RT-qPCR analysis in the 32Dcl3 cell line of a subset of differentially expressed genes identified by RNA-sequencing. **G and H**) CSF3R protein levels in total cell lysate of K562 cells (G) and of BC/CP samples from patient #3, carrying *UBE2A* mutation in the BC phase (H).

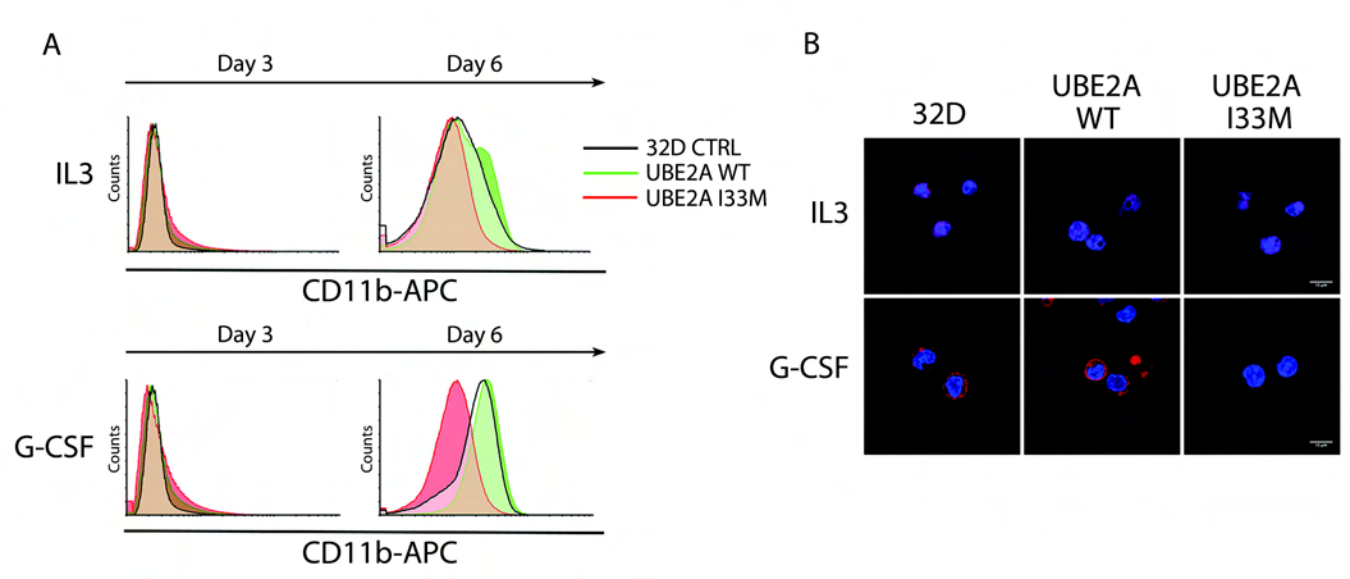
Figure 3. Induction of erythroid differentiation in UBE2A-silenced K562 cell line. K562 cells were treated with 400 μM hydroxyurea. **A**) CD235a immunofluorescence staining for UBE2A-silenced K562 (shUBE2A) and control (shNC) cells after hydroxyurea or mock (-) treatments for the indicated times. (scale bar: 25μm). **B**) Average intensity of CD235a signal obtained acquiring 10 fields from two independent experiments for each sample (approximately 80 cells each). **C**) FACS analysis of CD235a levels in K562 cells in presence (red line) or absence (black line) of hydroxyurea. **D**) Quantification of CD235a and hemoglobin mRNA relative levels (HBB: Hemoglobin-subunit-beta) through qRT-PCR after hydroxyurea treatment.

Figure 4. Induction of neutrophilic differentiation in UBE2A WT or I33M 32Dcl3 cell line. Cells were treated with IL-3 or GCSF. **A**) FACS analysis of CD11b staining after induction of differentiation at day 3 and 6. **B**) CD11b immunofluorescence staining for 32Dcl3 CTRL, UBE2A WT and I33M at day 6 showing a clear reduction in UBE2A I33M CD11b staining (scale bar: 20μm).


A
UBE2A
Empty
WT
D114V
I33M
UBE2A/B
ACTIN
Ba/F3 BCR/ABL1
UBE2A
Empty
WT
D114V
I33M
BCR/ABL1
ACTIN
B
UBE2A mRNA fold change
3500
2500
1500
2
1.5
1
0.5
0
***
***
***
Empty
WT
D114V
I33M
C
Ba/F3 BCR/ABL1
UBE2A
Empty
WT
D114V
I33M
70KDa
25KDa
15KDa
mUbH2A
H2A
ACTIN
D
GST-ubiquitin
UBE2A
WT
WT D114V I33M
170KDa
70KDa
40KDa
GST-ub
GST-ub
UBE2A/B
GST-ub (>170KDa) intensity signal fold change
1.5
1
0.5
0
*
**
WT
D114V
I33M
E
UBE2A enzimatic activity fold change
1.5
1.0
0.5
0.0
**
**
WT
D114V
I33M


A
shNC
shUBE2A
UBE2A fold change
1.5
1
0.5
0
B
shNC
shUBE2A
UBE2A/B
ACTIN
C
relative
row min
row max
shNC
shUBE2A
1 2 3 1 2 3
ACSM3
ATP2A1-AS1
LY6E
FAM234B
LCP2
ANKRD40
ATP13A3
DNAJC3-AS1
DOPEY2
ARPC3
COXC7
SKA2
TUPL3
FAM43A
FAM155A
XXYLT1-AS2
TBXA2R
ZEB2
FAM129A
ZNF726
AREG
ITGB4
CCDC90B
CTCFL
TMCC1-AS1
PRICKLE1
AIM2
GP6
TMEMS4
SLC22A11
SYNJ2BP
CLDND2
CLC11A
FOXP4-AS1
FGFBP3
DTWD2
LONRF2
IFI16
RDH10
SH3YL1
TYMSOS
MGAT4B
STOM
CRYAB
C15orf39
DNAJB2
ADCK1
MVP
ABCG2
FAM13A
FCGR2C
ACVRL1
ALOX5
SLC2A12
SAMD11
SLAMF6
VTN
TRANK1
TRIM6
FCGR2B
VANGL2
FAM46C
MTCL1
APOL4
MFHAS1
INPP5J
PTGDR2
LGALS3BP
P4HA2
APBB1
GABBR2
GTF2IPA
CSF3R
TYROBP
FOSL2
TUBA1A
shNC
shUBE2A
1 2 3 1 2 3
ACSS2
MSMO1
ATP11A
DNHD1
DHCR7
HDAC11
GYG2
SYDE2
DIXDC1
UCA1
S11A1
ACAT2
ATG5
HSD17B7
UBE2A
PRSS16
CARD6
BTG1
ADPRH
MMRN2
SELPLG
IL10RA
RNF182
USP35
ACHE
NR1H2
PELO
SEMA7A
CA2
TUBA4A
EZH1
PDE8A
TMPRSS13
CA8
TMOD2
CKMT1A
SMAD7
CKMT1B
NAPSA
SLC17A7
STS
HMGCS1
STXBP1
RASGEF1A
ASIC4
ELOVL6
GMPPA
EBP
CD80
RAB3D
SNX27
KIAA1522
GLIPR2
PORCN
RAP1GAP
CORO2A
TMEM98
TPCN1
CLCN2
TMEM45A
PPOX
PKLR
SLC22A5
CTIF
NSDHL
NARF
S100A4
SEMA3A
ITPK1
VIL1
STK10
KRT10
SLC7A8
RHCE
TNFRSF14
MUC3A
D
shNC
shUBE2A
mRNA fold change
8
6
4
2
0
***
***
**
ITGB4
RDH10
CLEC11A
mRNA fold change
1.5
1
0.5
0
***
**
CSF3R
RAP1GAP
E
MYELOID LEUKOCYTE DIFFERENTIATION
Enrichment Score
0
-0.10
-0.20
-0.30
-0.40
Ranked list metric
2.5
0
-2.5
0
5000
10000
15000
20000
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
POSITIVE REGULATION OF ERYTHROCYTE DIFFERENTIATION
Enrichment Score
0
-0.20
-0.40
-0.60
Ranked list metric
2.5
0
-2.5
0
5000
10000
15000
20000
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
F
UBE2A WT
UBE2A I33M
mRNA fold change
4
3
2
1
0
**
*
**
ITGB4
RDH10
CLEC11A
mRNA fold change
1.5
1.0
0.5
0.0
*
**
CSF3R
RAP1GAP
G
K562
shUBE2A
-
+
CSF3R
ACTIN
H
pt#3
CP
BC
CSF3R
ACTIN


A
Hydroxyurea
-
24h
72h
shNC
shUBE2A
Blue: DAPI
Green: CD235a
B
shNC
shUBE2A
Relative CD235a intensity signal
2.5
2
1.5
1
0.5
0
***
*
*
hrs
0
24
72
Hydroxyurea treatment
C
shNC
Hydroxyurea treatment
0h
24h
72h
cell count
150
113
75
38
0
10
10¹
10²
10³
10⁴
CD235a-FITC
shUBE2A
Hydroxyurea treatment
0h
24h
72h
cells count
150
113
75
38
0
10
10¹
10²
10³
10⁴
CD235a-FITC
D
shNC
shUBE2A
CD235a
relative mRNA level
6
4
2
0
***
**
*
hrs
0
8
24
Hydroxyurea treatment
HBB
relative mRNA level
0.6
0.4
0.2
0
***
***
***
hrs
0
8
24
Hydroxyurea treatment


A
Day 3
Day 6
IL3
Counts
Counts
32D CTRL
UBE2A WT
UBE2A I33M
CD11b-APC
Day 3
Day 6
G-CSF
Counts
Counts
CD11b-APC
B
32D
UBE2A WT
UBE2A I33M
IL3
G-CSF

Supplemental Methods

Validation of mutations and *UBE2A* target sequencing

All variants detected by Next Generation Sequencing approach were validated by standard Sanger sequencing[1] or Mass Spectrometry method adapted from[2] for genes indicated by * in Table S3. Briefly, 100ng of gDNA was amplified with the High Fidelity Fast Start Taq polymerase (Roche, Indianapolis, IN, USA) following manufacturer instructions. Primers are listed in Table S3. Extension primers used for genes indicated by * were: ASXL1 CTGCCATCGGAGGGG, EPB41L3 TTGGCAGTGTTTCACC, FGF34 CCAACACCACAGCCG, BARD1 TCGACAGGCCGCAGACC, BSN CAAGGCTCCTTCCCTGGA, EFCAB4B GGCGGTGGCTGAGCAGC, KRT7 GGGACATATCCTGCTTGCCCC, MUDENG CCCACTAAATGCAGAGTCAT, UBE2A CCGTCCGAGAACAACAT, XPO1 AAATAGATTTACCATGCATGAATT. The entire coding region of *UBE2A* gene was amplified from gDNA using three different set of primers listed in Table S3.

Vector constructions and mutagenesis

The pCMV6_AC_GFP plasmid (cod. RG204194) containing the coding sequence (CDS) of UBE2A isoform 1 was obtained from Origene Technologies (Rockville, MD, USA). The UBE2A_CDS was amplified with the High Fidelity Fast Start Taq Polymerase (Roche) following manufacturer instructions with these primers: Fw $^{5'}$AATAAGATCTACCATGTCCACCCCGGCTCGGCG$^{3'}$, Rw $^{5'}$AATACTCGAGCTAACAATCACGCCAGCTTTGTTCTACTATTG$^{3'}$. The amplified product was subsequently digested with BglII and XhoI restriction enzymes (Roche) and cloned into the pMIGR1 vector[3] (a kind gift from G. Cazzaniga, Tettamanti Foundation, San Gerardo Hospital), obtaining the pMIGR1_UBE2A_wt plasmid. Site directed mutagenesis

was applied to create pMIGR1_UBE2A_D114V and pMIGR1_UBE2A_I33M plasmids by using the Pfu Ultra High Fidelity enzyme (Agilent Technologies, Santa Clara, CA, USA). The products were then digested with DpnI (Roche) and 2 µl were used to transform the competent TOP10 bacterial strain (Life Technologies). The presence of the mutations was confirmed by Sanger sequencing. The primers used for mutagenesis reaction were as follows: UBE2A_I33MFw $^{5'}$CGTCCGAGAACAACATGATGGTGTGGAACGC$^{3'}$, UBE2A_I33MRw $^{5'}$ GCGTTCCACACCATCATGTTGTTCTCGGACG$^{3'}$, UBE2A_D114VFw $^{5'}$CATCCATACAGTCTCTGTTGGTTGAACCCAATCCCAATAGTCC$^{3'}$, UBE2A_D114VRw $^{5'}$GGACTATTGGGATTGGGTTCAACCAACAGAGACTGTATGGATG$^{3'}$.

The pT7CFE1-Chis-UBE2A vectors for *in-vitro* protein translation were prepared from pMIGR_UBE2A_wt, pMIGR_UBE2A_D114V and pMIGR_UBE2A_I33M. The UBE2A_CDS was amplified from the plasmids with the Q5^{TM} High-Fidelity DNA Polymerase (New England BioLabs, Ipswich, MA, USA) following manufacturer instructions with these primers: Fw $^{5'}$CCCATATGATGTCCACCCCGGCTC$^{3'}$, Rw $^{5'}$ACTCGAGCTAACAATCACGCCAG$^{3'}$.

The respective amplified products and the pT7CFE1-CHis expression vector (Thermo Fischer SCIENTIFIC, Life Technologies, Waltham, MA, USA) were digested with NdeI and XhoI restriction enzymes (New England BioLabs, Ipswich, MA, USA) and were ligated with T4 DNA Ligase.

Western Blot

Whole cell lysates and western blots were performed as previously described[4]. For histone analysis cells were directly lysed in Laemmli Buffer. The antibody used were: UBE2A(A300-281A)(Bethyl Laboratories Inc., Montgomery, TX, USA), ABL1(K-12) (Santa Cruz Biotechnology Inc, Dallas, TX, USA), H2A(ab18255) (Abcam, Cambridge, UK), Actin (A2066)(Sigma-Aldrich), GST (Sigma-Aldrich), CSF3R(ab126167) (Abcam).

Fluorescence Activated Cell Sorting Analysis (FACS)

K562 or 32Dcl3-BCR/ABL1 cells were stained with CD235a-FITC (130-100-265, Mylteni Biotech, Germany), CD11b-APC or mock controls antibodies after differentiation induction according to manufacturer instructions

At the defined time points cells were washed twice and resuspended in PBS to a final concentration of 1 to $5x10^6$ cells and incubated for 1h a RT with primary antibodies. Stained cells were analyzed on a BD FACSCanto I instrument.

Fluorescent Imaging

K562 cells or 32Dcl3-BCR/ABL1 were washed twice with PBS and fixed for 10 min at 25°C with 4% (w/v) p-formaldehyde in 0.12M sodium phosphate buffer, pH7.4, incubated for 1 h with primary fluorescent conjugated antibodies (CD235a-FITC, 130-100-265, Mylteni Biotech; CD41-FITC, SAB4700372, Sigma-Aldrich and CD11b-APC BD Bioscience) in GDB buffer [0.02M sodium phosphate buffer, pH7.4, containing 0.45M NaCl, 0.2% (w/v) bovine gelatine] and counterstained with DAPI. After two wash with PBS, labelled cells were mounted on glass slides with a 90% (v/v) glycerol/PBS solution. Images were acquired with a LSM710 confocal microscope (Carl Zeiss) and analysed with ImageJ software (https://imagej.nih.gov/ij/).

Whole Exome Sequencing (WES) and Copy Number analysis

Samples were sequenced as described in[5]. WES was performed with a mean coverage of 60X. Image processing and basecall were performed using the Illumina Real Time Analysis Software. Paired Fastq files were aligned to the human reference genome (GRCh37/hg19) using the BWA aln/sampe algorithms[6]. Duplicates were removed using Rmdup. Quality of the aligned reads, somatic variant calling and copy number analysis

were performed using CEQer2, an in-house evolution of CEQer tool[7] as previously described[5,8]. Splicing variants were analyzed using SpliceFinder[9]. Variants were annotated using dbSNP147. All the filtered variants, exported as vcf files, were annotated using Annovar[10] and manually inspected. To specifically identify variants present in the majority of BC cells and therefore likely playing a critical role in the progression, we included variants with a relative frequency ≥25%.

RNA-Sequencing

RNA libraries were generated from 2 µg total RNA extracted with TRIzol reagent (Life Technologies-Thermo Fisher Scientific, Waltham, MA USA) following manufacturer instructions. Libraries were sequenced as described in[5]. A total of three clones was selected for RNA-Seq analysis of both K562shNC and K562shUBE2A cell lines. Padj≤0.05 was used as the main criteria to identify significantly deregulated genes with a Fold change ≥1.5 or ≤ 0.75. Image processing and basecall was performed using the Illumina Real Time Analysis Software. Paired Fastq files were aligned to the human genome (GRCh38/hg38) by using STAR[11], a splice junction mapper for RNA-Seq data, together with the corresponding splice junctions Ensembl GTF annotation, using the following parameters: --runThreadN 8 --outReadsUnmapped Fastx --outFilterType BySJout --outSAMattributes NH HI AS nM MD --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --alignTranscriptsPerReadNmax 100000 --quantMode TranscriptomeSAM GeneCounts --limitBAMsortRAM 16620578182 --outSAMtype BAM SortedByCoordinate --chimSegmentMin 20 --chimJunctionOverhangMin 10.

Chemical compounds

Phorbol 12-myristate 13-acetate (PMA) and Hydroxyurea, Granulocyte colony-stimulating factor (GCSF human) and IL-3 were obtained from Sigma-Aldrich(St.Louis, MA, USA).

Supplemental data

Table S1
List of the Taqman Gene Expression Assays used in the study

	Human	Mouse
UBE2A	Hs_00163308_m1	
ITGB4	Hs_01103156_m1	Mm01266840_m1
RDH10	Hs_00416907_m1	Mm00467150_m1
CSF3R	Hs_01114430_g1	Mm00432735_m1
CLEC11A	Hs_00998294_g1	Mm01206715_m1
RAP1GAP	Hs_00937964_g1	Mm01181224_m1
CD44	Hs_01075864_m1	
CD235a	Hs_00266777_m1	
HBB	Hs_00747223_g1	
ITGA2B	Hs_01116228_m1	

Table S2 List of all the primers used for the analysis (* analyzed by mass spectrometry)

GENE	FORWARD PRIMER	REVERSE PRIMER	Genomic Position, substitution
RTP2	GGCAGGACTGAGGAAGGAGAAC	GCTGGAGGAGGAGGTGACCAC	Chr3: 187416394-187416395, C/T
KCNH3	CAGGTGTCCAGGCAAGAGTG	CTGCTCTCGATCTCCCGCTGGC	Chr12: 49937814-49937815, G/A
FAT4	CATTGGCACAAACGTGATATC	ATGTAACTCTCTGTTAGCCTTTGAC	Chr4: 126242657-126242658, C/T
FUT3	CTAGCAGGCAAGTCTTCTGGAGG	CCCAGCAGAAGCAACTACGAGAG	Chr19: 5843790-5843791, G/A
RUNX1	GCTTTGAGTAGCGAGAGTATTGAC	GGTAACTTGTGCTGAAGGGCTG	Chr21: 36231801-36231802, T/A
SMARCA4	CCCGCAGATCCGTTGGAAGTAC	GCCTACAGCACGCTACAGCCTCTAG	Chr19: 11132616-11132617, G/A
UBE2A	CATGCGGGACTTCAAGAGGTTGC	CAATCACGCCAGCTTTGTTCTAC	ChrX: 118717099-118717100, A/T
UBE2A(I33M)	ACGTTGGATGAGACTCACCCGAAAATGACC	ACGTTGGATGTGTCTTCCCGAAGGTTGCAG	ChrX: 118708918-118708919, A/G
ABL1 (F486S)	CTTGTTGCAGCAAAAGATGGTTAG	GCTGGTCTGTGAACTTTCACCAG	Chr9: 133755487-133755488, T/C
ABL1 (T315I)*	CTTGTTGCAGCAAAAGATGGTTAG	GCTGGTCTGTGAACTTTCACCAG	Chr9: 133748282-133748283, C/T
ABL1 (E255V)	TGACCAACTCGTGTGGAAACTC	TTCGTCTGAGATACTGGATTCCTG	Chr9: 133738364-133738364, A/T
PTPN11	GGTGATTTGTTGGCAAGTGAGGG	GCATGGCAGTTCTTCAATCTGGCAG	Chr12: 112926887-112926888, G/T
FAM123C	CGAGAGGAAGAGACACGAGGTCAC	CACGGAGGTGACACTCTGGATGC	Chr2:131521770-131521771, G/A
LAMA2	GTAGTACCCGAATATAAGGTGTTACAG	CTTCATCATCTTTCCTACAAGTAAACTG	Chr6: 129621915-129621916, C/T
GRIN3A	GACTTGTCCTTTGATACTCCTCCAG	CCATGACACCCTAGCAGGTAGTCTG	Chr9: 104335733-104335734, G/A
SMC5	CTTACAGCTCCTGCAAAATCTTCC	CCAAGAACTCAAGTTCAACCAAGAC	Chr9: 72967244-72967245, T/A
MESDC2	TAGGCAACAAGAGCAAAACTCTGTC	TCTTCCAGAGCACAAGAGACCTTC	Chr15: 81274346-81274347, C/-CT
CCDC40	CAAACAGGACAAAGACGTGACAAC	TCCTAACCTCATGTGATCCACCTG	Chr17: 78011941-78011942, C/T
NRAS	CCTAGATTCTCAATGTCAACAACC	ACTGGGTACTTAATCTGTAGCCTCC	Chr1: 115256528-115256529, T/C
DEFB119	CCTGACTCAATAGCCTCTCCTGC	GCTAGGAAGACAGAAGGGTGAG	Chr20: 29976969-29976970, C/T
IKZF1	GATCAAATTGACCCAGCCAGTG	GTGAGACTTCTGTGTGTGTATGTGC	Chr7: 50450291-50450292, A/G
AK8	CCTGTCATTAAATGCTTTCCTGTG	GATGGAGCACGGGAAGTAGCACC	Chr9: 135730271-135730272, C/T
PPT1	CTCTTCCTATGTCTCCAGCAATG	CCTCAGGTAGTCCACCCACCTC	Chr1: 40557017-40557018, A/G
MDH1B	CCAGATACTCAGAATGTTCCTAGAGG	GAGACCCAAGACCTGGCATCTC	Chr2: 207619827-207619828, G/A Chr2: 207619828-207619829, C/T
GPR98	GTTCTCCACAGGGCTGCCTCC	GAGTCACTATGTTCCAGGTACAGTG	Chr5: 89971181-89971182, C/T
CEL	CCAGCAACCAACGTGACCTAG	CCAGGATAAAGAACGGAAATGTGG	Chr9: 135942014-135942015, G/C
LRP4	GACACAACCTCCTCCACGTTGC	GCCACTCTTCTGGTACTGATGC	Chr11: 46916334-46916335, C/T
CYP2B6	GACTCAGAGCCTTCTTCCAACTTC	CTCCAGTTTCGTCTGTCTCTGTC	Chr19: 41510299-41510300, C/T
BCR	GAGCAGGTGGGAGGGAGCAG	CACAGGGCTGACGCAACGAAC	Chr22: 23610684-23610685, T/G
ASXL1*	ACGTTGGATGTCTGCCACCTCCCTCATCG	ACGTTGGATGATAGAGAGGCGGCCACCAC	chr20:31022441-31022442, -/G
EPB41L3*	ACGTTGGATGGAACTACTGGCACTTCCTTC	ACGTTGGATGCGGAAACCATCAGTTTTGGC	Chr18: 5396285/5396286, G/A
FGFR4*	ACGTTGGATGTACACCTTGCACAGCAGCTC	ACGTTGGATGCATCCTGCAGGCCGGGCTC	Chr5: 176519378-

			176519379, G/A
BARD1*	ACGTTGGATGTGAGTCGAGTCACACATTTG	ACGTTGGATGTCTGTATAATCGACAGGCCG	Chr2: 215617269-215617270, C/+CGGTT
BSN*	ACGTTGGATGTACTGCTCCTGATAGCCAAC	ACGTTGGATGGTCAAGGACACCTGGTTCTC	Chr3: 49699069-49699070, G/A
EFCAB4B*	ACGTTGGATGAAATGCTCCTAGGAAGGTCG	ACGTTGGATGTCACAGACAAGCAGTCGTTC	Chr12: 3736607/3736608, C/T
KRT7*	ACGTTGGATGTCAAGGATGCTCGTGCCAAG	ACGTTGGATGATGAGTTCCTGGTACTCACG	Chr12: 52639298-52639299, C/T
MUDENG*	ACGTTGGATGTCACCCTTGTGTAACTTCTC	ACGTTGGATGTGGAAATTTGTAAGGCCCAC	Chr14: 57747057-57747058, G/A
XPO1*	ACGTTGGATGGTTTTTTGAGAGCTCACTGG	ACGTTGGATGAGAAAGAGATTTACCATGC	Chr2: 61719472-61719473, C/T
HMCN1	CTTTAGACACTGGGCAATA	AATAGTGCTGCTTTCAGTCA	Chr1: 185970471-185970472, C/T
UBE2A_CDS1	CTCTCTCTGCTCTCAGGTTGGTTC	ATTCCACTCAAGCCTTTAGCAG	-
UBE2A_CDS2	CTTTCCTCTCTACCCTGTATCTTTG	TCTAGGACAAGACAGCCACAGAC	-
UBE2A_CDS3	CTGATTTCCTGGATAATAGGGCAGC	AAGGAAGATGGAAACAGCACAACAG	-

Supplemental Figures

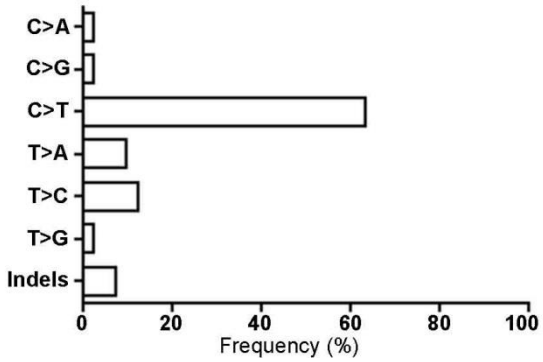


Figure S1. Graphic representation of somatic single nucleotide variants as frequency on the total of non-synonymous variants identified in the 10 BC samples.

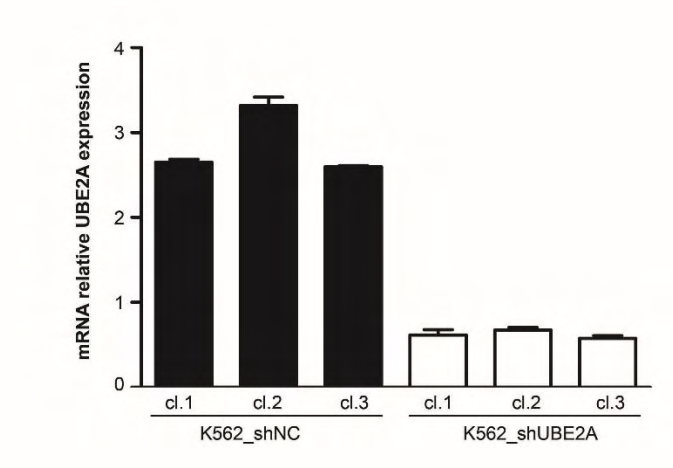


Figure S2. RT-qPCR for UBE2A expression levels in K562_shUBE2A and controls (shNC) populations selected for RNA-seq analysis. Beta-glucoronidase (GUSB) gene has been used as internal reference.

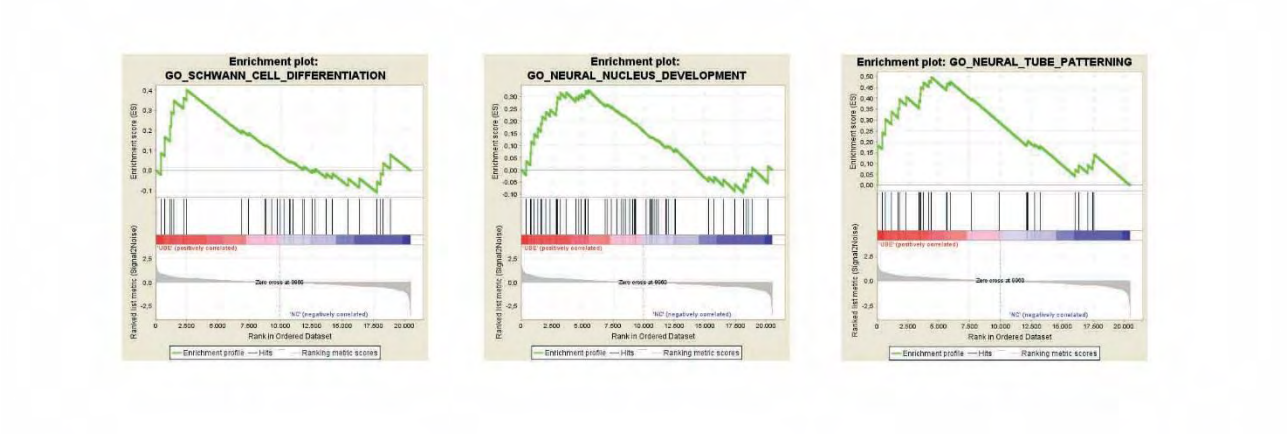


Figure S3. Gene set enrichment analysis (GSEA) from RNA-seq data of differentially expressed genes comparing K562_shUBE2A with controls (shNC) displaying three of the most enriched categories associated with neuronal development.

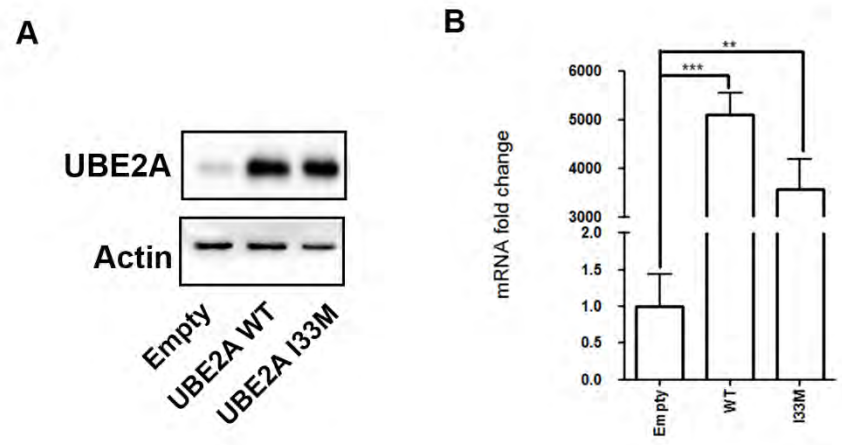


Figure S4. Expression of UBE2A WT and I33M in the BCR-ABL1-positive 32Dcl3. **A**) Western Blot analysis of total cell lysates from BCR-ABL1-positive 32Dcl3 cell lines stably transfected with pMIGR-UBE2A vectors encoding for the UBE2A (isoform 1) wild-type (WT) or I33M form. Empty vector has been used as negative control. **B**) RT-qPCR of total RNA extracted from BCR-ABL1-positive 32Dcl3 pMIGR/UBE2A cell lines. The values are normalized on the EMPTY cells (Empty vs WT *** $p<0.001$ and Empty vs I33M **$p<0.01$).

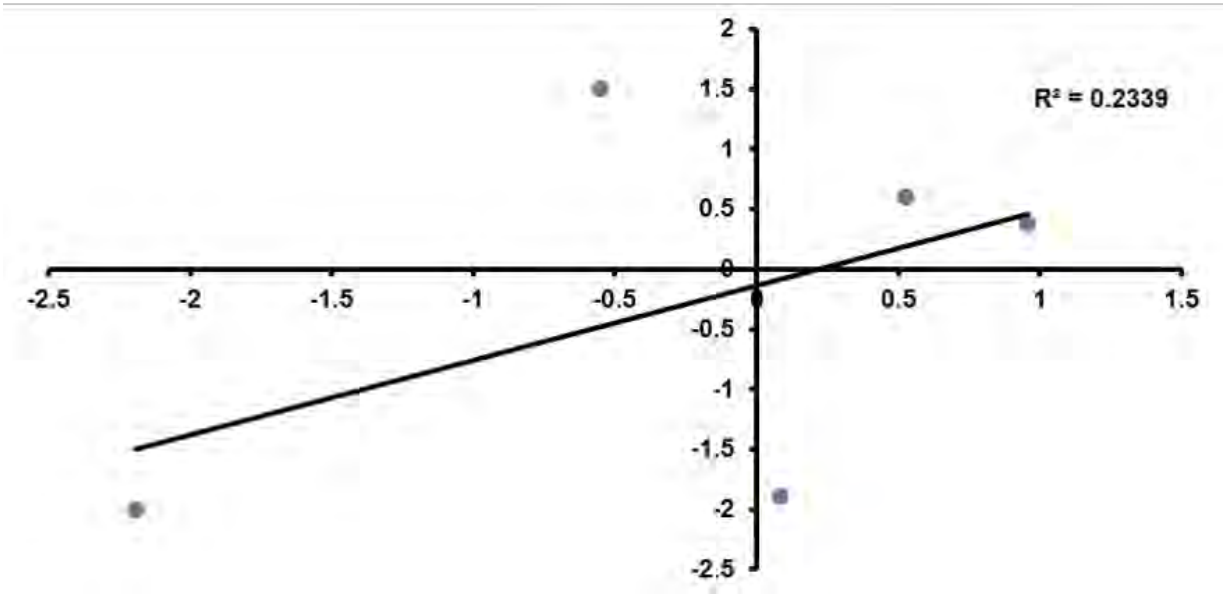


Figure S5: Linear correlation of Log2FC of 32Dcl3-BCR/ABL1 UBE2A WTvs 32Dcl3-BCR/ABL1 UBE2A I33M (y-axys) and known BC data (GEO_GSE47927) vs HSC data (x-axys) target genes; $R^2 = 0.234$.

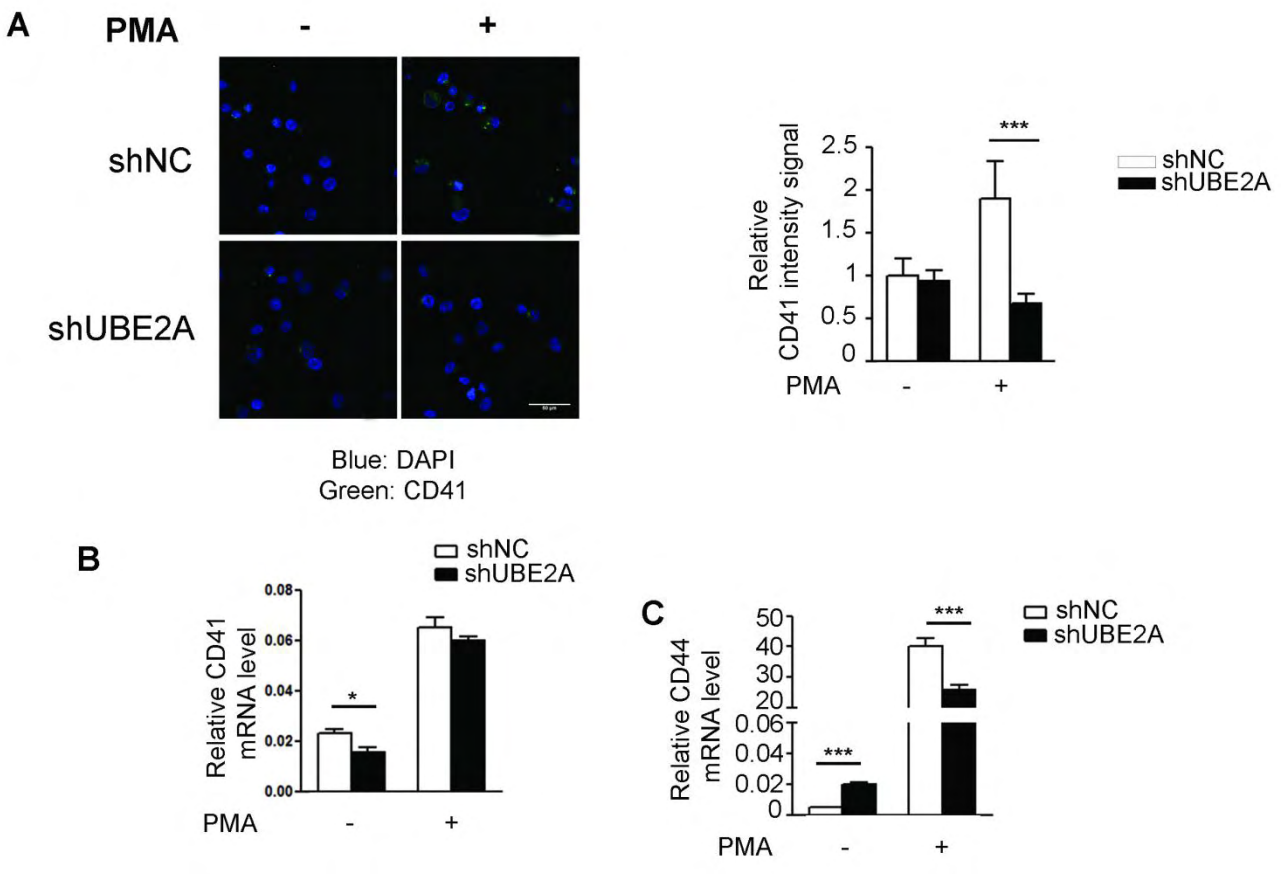


Figure S6. A) Immunofluorescence staining for CD41 in UBE2A-silenced-K562 (shUBE2A) and controls (shNC) after 72 hours of Phorbol 12-myristate 13-acetate (PMA) treatment at 10nM. Green: CD41-FITC; Blue: DAPI. The graph represents the average intensity of CD41 signal obtained from acquiring 10 fields from two independent experiments for each sample (approximately 100 cells each). **B-C**) CD41 and CD44 mRNA expression level assessed through RT-qPCR in PMA treated (+) and untreated K562 cells. Beta-glucoronidase (GUSB) gene has been used as internal reference.

References

1. Branford S, Rudzki Z, Walsh S, Grigg A, Arthur C, Taylor K, et al. High frequency of point mutations clustered within the adenosine triphosphate-binding region of BCR/ABL in patients with chronic myeloid leukemia or Ph-positive acute lymphoblastic leukemia who develop imatinib (STI571) resistance. Blood. 2002 May 1;99(9):3472-5.
2. Parker WT, Lawrence RM, Ho M, Irwin DL, Scott HS, Hughes TP, et al. Sensitive detection of BCR-ABL1 mutations in patients with chronic myeloid leukemia after imatinib resistance is predictive of outcome during subsequent therapy. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011 Nov 10;29(32):4250-9.
3. Pear WS, Miller JP, Xu L, Pui JC, Soffer B, Quackenbush RC, et al. Efficient and rapid induction of a chronic myelogenous leukemia-like myeloproliferative disease in mice receiving P210 bcr/abl-transduced bone marrow. Blood. 1998 Nov 15;92(10):3780-92.
4. Magistroni V, Mologni L, Sanselicio S, Reid JF, Redaelli S, Piazza R, et al. ERG deregulation induces PIM1 over-expression and aneuploidy in prostate epithelial cells. PloS one. 2011;6(11):e28162.
5. Gambacorti-Passerini CB, Donadoni C, Parmiani A, Pirola A, Redaelli S, Signore G, et al. Recurrent ETNK1 mutations in atypical chronic myeloid leukemia. Blood. 2015 Jan 15;125(3):499-503.
6. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010 Mar 1;26(5):589-95.
7. Piazza R, Magistroni V, Pirola A, Redaelli S, Spinelli R, Galbiati M, et al. CEQer: a graphical tool for copy number and allelic imbalance detection from whole-exome sequencing data. PloS one. 2013;8(10):e74825.
8. Piazza R, Valletta S, Winkelmann N, Redaelli S, Spinelli R, Pirola A, et al. Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. Nature genetics. 2013 Jan;45(1):18-24.
9. Spinelli R, Pirola A, Redaelli S, Sharma N, Raman H, Valletta S, et al. Identification of novel point mutations in splicing sites integrating whole-exome and RNA-seq data in myeloproliferative diseases. Molecular genetics & genomic medicine. 2013 Nov;1(4):246-59.
10. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010 Sep;38(16):e164.
11. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15-21.