De novo UBE2A mutations are recurrently acquired during chronic 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, South Australia, Australia; [4]GalSeq s.r.l., Monza, Italy; [5]Department of Bioscience and Biotechnology, University of Milano Bicocca, Milano, Italy; [6]Bloodwise Molecular Haematology Unit, John Radcliffe Hospital, University of Oxford, 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, South Korea; [10]University of South Australia, Adelaide, South Australia, Australia and [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*

 doi:10.3324/haematol.2017.179937

Received: January 19, 2018.

Accepted: February 26, 2019.

Pre-published: February 28, 2019.

Correspondence: *VERA MAGISTRONI* - vera.magistroni@unimib.it

ROCCO PIAZZA - rocco.piazza@unimib.it

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

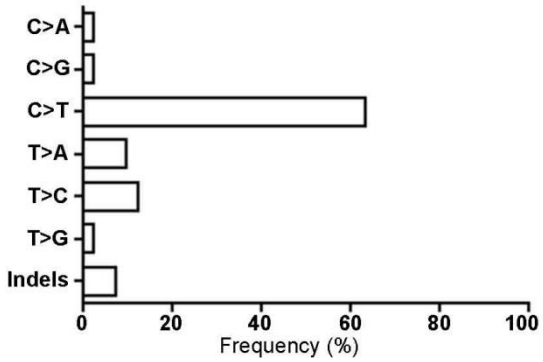

Figure S1. Graphic representation of somatic single nucleotide variants as frequency on the total of non-synonymous variants identified in the 10 BC samples.

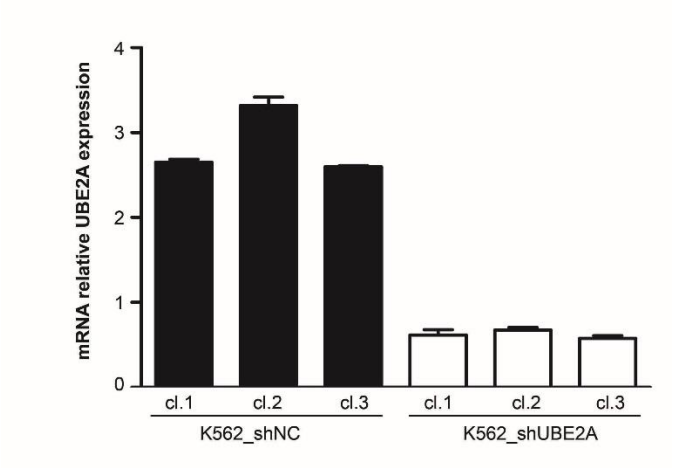

Figure S2. RT-qPCR for UBE2A expression levels in K562_shUBE2A and controls (shNC) populations selected for RNA-seq analysis. Beta-glucoronidase (GUSB) gene has been used as internal reference.

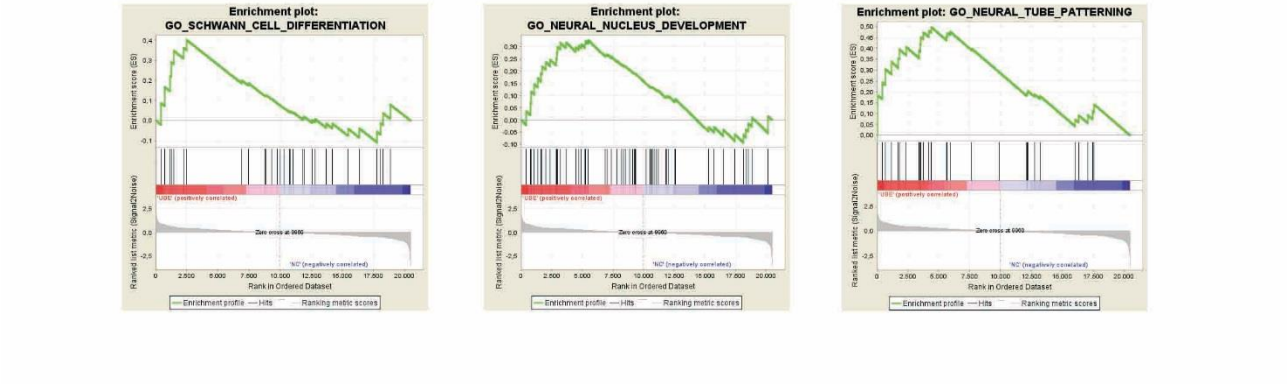

Figure S3. Gene set enrichment analysis (GSEA) from RNA-seq data of differentially expressed genes comparing K562_shUBE2A with controls (shNC) displaying three of the most enriched categories associated with neuronal development.

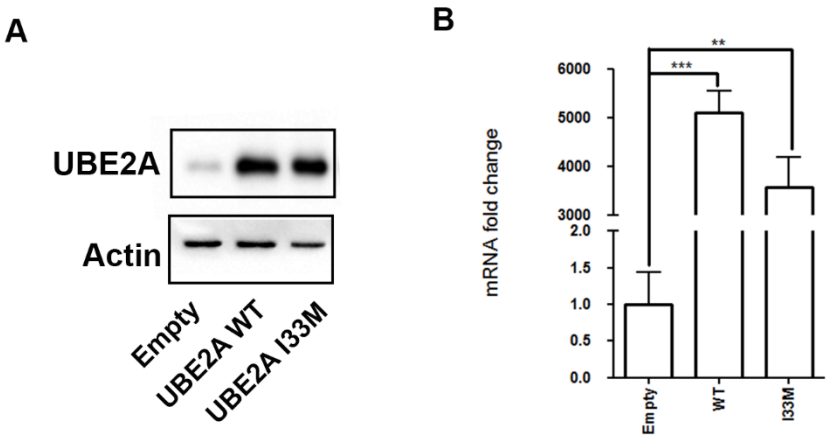

Figure S4. Expression of UBE2A WT and I33M in the BCR-ABL1-positive 32Dcl3. **A**) Western Blot analysis of total cell lysates from BCR-ABL1-positive 32Dcl3 cell lines stably transfected with pMIGR-UBE2A vectors encoding for the UBE2A (isoform 1) wild-type (WT) or I33M form. Empty vector has been used 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 **$p<0.01$).

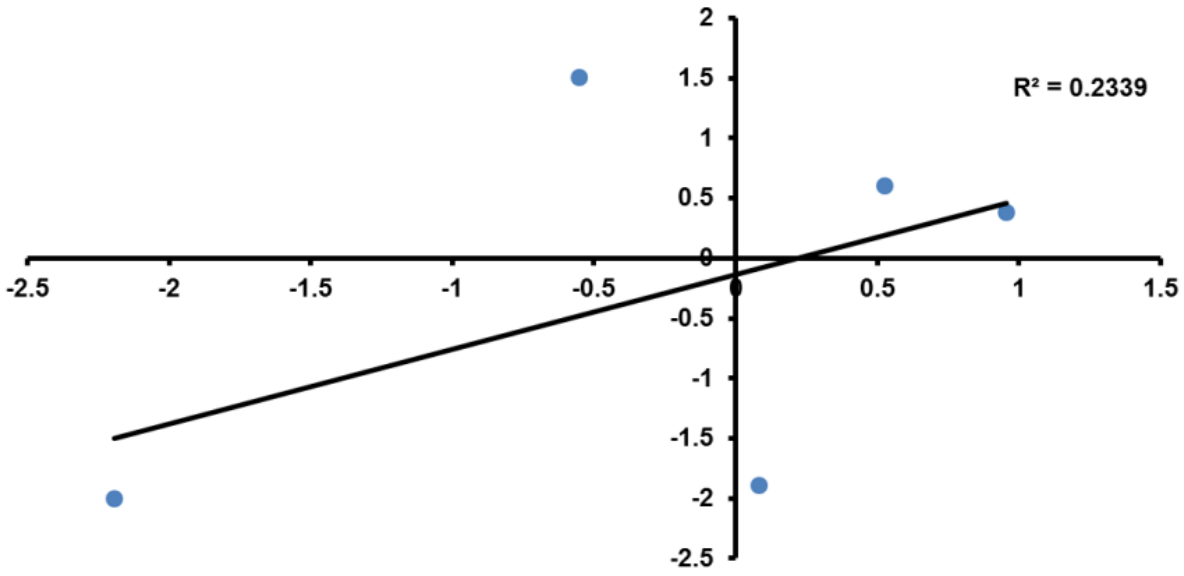

Figure S5: Linear correlation of Log2FC of 32Dcl3-BCR/ABL1 UBE2A WTvs 32Dcl3-BCR/ABL1 UBE2A I33M (y-axys) and known BC data (GEO_GSE47927) vs HSC data (x-axys) target genes; $R^2 = 0.234$.

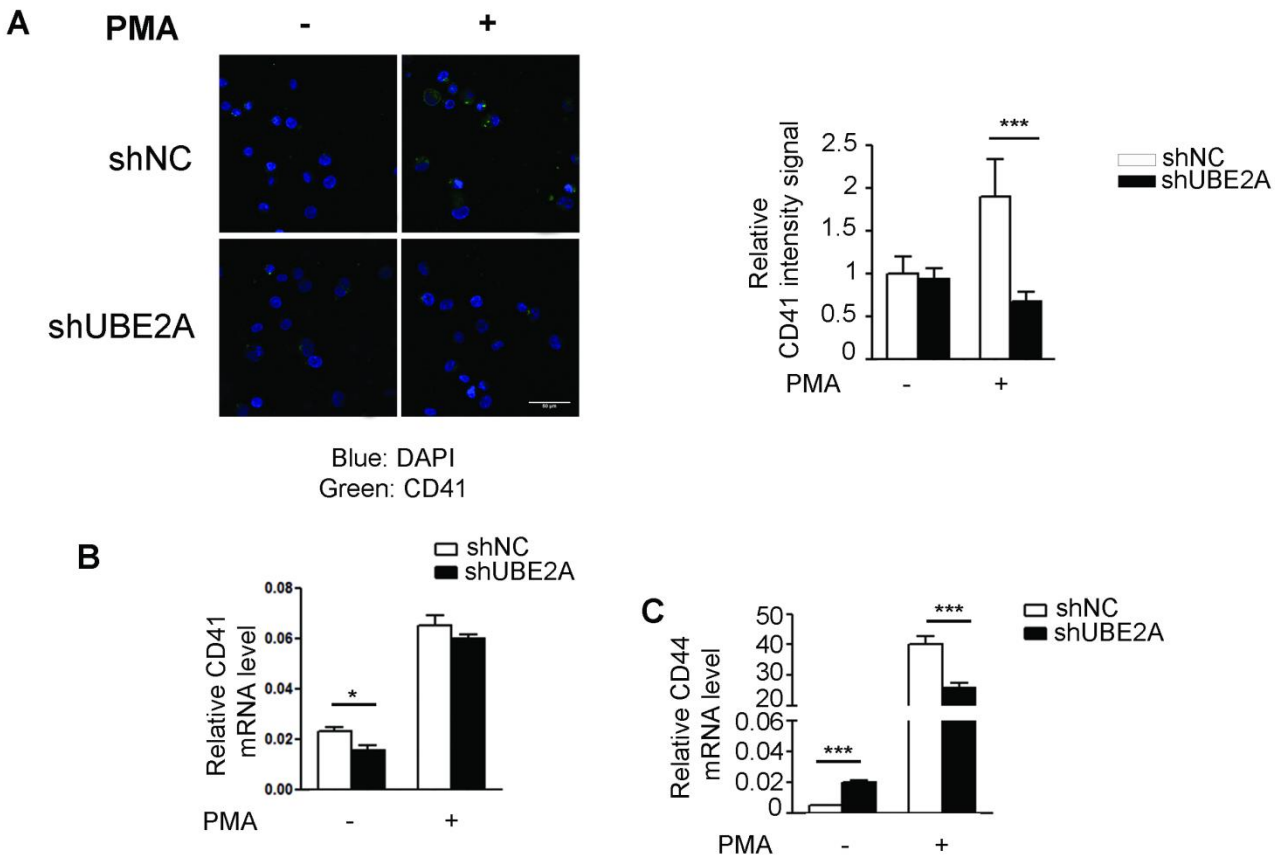

Figure S6. A) Immunofluorescence staining for CD41 in UBE2A-silenced-K562 (shUBE2A) and controls (shNC) after 72 hours of Phorbol 12-myristate 13-acetate (PMA) treatment at 10nM. Green: CD41-FITC; Blue: DAPI. The graph represents the average intensity of CD41 signal obtained from acquiring 10 fields 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.