SUPPLEMENTARY APPENDIX

The epigenetic regulator RINF (CXXC5) maintains *SMAD7* expression in human immature erythroid cells and sustains red blood cell expansion

Audrey Astori,[1,2,3*] Gabriel Matherat,[1,2,3*] Isabelle Munoz,[1,2,3] Emilie-Fleur Gautier,[1,2,3] Didier Surdez,[3,4,5] Yaël Zermati,[1,2] Frédérique Verdier,[1,2] Sakina Zaidi,[3,4,5] Vincent Feuillet,[1] Amir Kadi,[1] Evelyne Lauret,[1,3] Olivier Delattre,[3,4,5] Carine Lefèvre,[1,2] Michaela Fontenay,[1,2,6] Evelyne Ségal-Bendirdjian,[7] Isabelle Dusanter-Fourt,[1,3] Didier Bouscary,[1,3] Olivier Hermine,[2,3,8] Patrick Mayeux[1,2,3] and Frédéric Pendino[1,2,3]

[1]Université de Paris, Institut Cochin, INSERM, CNRS; [2]Laboratory of Excellence GR-ex; [3]Equipe Labellisée Ligue Nationale Contre le Cancer (LNCC); [4]PSL Research University, Institut Curie Research Center, INSERM U830; [5]SIREDO: Care, Innovation and Research for Children, Adolescents and Young Adults with Cancer, Institut Curie; [6]Service d'Hématologie Biologique, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, Centre-Université de Paris; [7]INSERM UMR-S 1124, Team: Cellular Homeostasis Cancer and Therapies, Université de Paris and [8]Université de Paris, Institut Imagine, INSERM, CNRS, Paris, France.

**AA and GM contributed equally as co-first authors.*

 doi:10.3324/haematol.2020.263558

Received: June 18, 2020.

Accepted: November 16, 2020.

Pre-published: November 26, 2020.

Correspondence: *FRÉDÉRIC PENDINO* - frederic.pendino@inserm.fr

SUPPLEMENTAL METHODS

Flow cytometric cell sorting of MEP cells

Freshly isolated CD34+ were labeled with the following labeled antibodies directed against the following markers: CD34 APC-Cy7 (clone 581), CD123 APC (6H6) from BioLegend; CD38 PE (clone HIT2) and CD45RA-FITC (HI100) from BD Biosciences, CD10 PE-Cy5 (eBioscience; clone CD-CALLA). Cells were sorted using BD FACSAriaIII (BD Biosciences) according to the gating strategy described by Manz *et al*.(1) Purification was obtained to at least 98% and only erythroid colonies were generated in CFC assay (i.e. no G-, M-, or GM-derived colony were noticed).

Chromatin Immunoprecipitation (ChIP) experiments

ChIP was performed with anti-RINF (ProteinTech, Cat.N°16513-1-AP) or anti-H3K4me3 (Diagenode, Cat.N°C15410003), antibodies using iDeal ChIP-seq Kit for Transcription Factors (Diagenode) following manufacturer's instructions and using the following options: K562 cells were crosslinked with 1% formaldehyde for 10 minutes in fresh media at room temperature and quenched with 0.2 M glycine for 5 additional minutes. Chromatin was sonicated using Diagenode Bioruptor (Power High, 20 cycles of 30''on-30''off and by changing ice cold water in the bath every 5 cycles). Sonication was validated by agarose gel electrophoresis (average DNA size ~150-250 bp) using an aliquot of purified sheared chromatin (proteinase K treated followed by and phenol-chloroform extraction). Quantitative-PCR was performed using Power SYBR Green PCR Master Mix (Thermo Fisher Scientific, Cat.N°4367659) using pairs of primers encompassing six regions (R0-R5) of *SMAD7* gene. R0 and R5 regions were used as negative controls. The efficiency of chromatin immunoprecipitation at these loci was determined as the recovery of that locus calculated as the percentage of the input (the relative amount of immunoprecipitated DNA compared to input DNA). The amount used for the input was 1% of the amount used for ChIP, the recovery was calculated as follows: % recovery = $2^{(Ct^{input} - Ct^{sample})}$.

Primer sequences used for quantitative RT-PCR and ChIP-qPCR

Human *PPIA* (Cyclophilin A) endogenous control was obtained from Thermo Fisher Scientific (Cat.N°4333763T). The other primer sequences (5'-3' orientation) used for RT-qPCR and ChIP-qPCR are indicated in the following table.

Q-RT-PCR (Hydrolysis)	**Forward**	**Reverse**	**Probe sequence (5'-3')**
RINF	TCCGCTGCTCTGGAGAAG	CACACGAGCAGTGACATTGC	FAM-aacccaaagctgccctctcc-BBQ
RPLP2	GACCGGCTCAACAAGGTTAT	CCCCACCAGCAGGTACAC	Cy5-agctgaatggaaaaaacattgaagacgtc-BBQ
Q-RT-PCR (SybrGreen)	**Forward**	**Reverse**	
cKIT	TATGCTCTCGCACCTTTCCA	TCTCAATGAAGTGCCCCTGA	
PU1	CACTGGAGGTGTCTGACGG	GCGGATCTTCTTCTTGCTGC	
SMAD7	CAATGACCACGAGTTTATGCA	GTTGAAGATGACCTCTAGCCA	
Murine Rinf	TCCGCTGCTCTGGAGAAG	CACGCGGGCAGTGACATTGC	
murineSmad7	TGCTGTGCAAAGTGTTCAGGTG	CCATCGGGTATCTGGAGTAAGGA	
murineGapdh	TGTGTCCGTCGTCGATCTGA	TTGCTGTTG AAGTCGCAGGAG	
ChIP-qPCR (SybrGreen)	**Forward**	**Reverse**	**Distance to SMAD7 TSS**
SMAD7 R0 (neg.)	ATGGGTGTTTCAAGAGTGGAG	GAAGAGGAAAGCCGTGTAGAG	-4.7kb
SMAD7 R1	CTAGCGCTTCATTCATTGGTTT	TCGGTCCAGTCCGGTATAA	-1962-1840
SMAD7 R2	GCAAACAACAGATCGGGTTTC	CAATCCATTCTGGGAGCTTCT	-1695 -1556
SMAD7 R3	GCCTCGGCTTCCTACATGGA	CTCCCCACCCCCAAATTAAG	-243 -122
SMAD7 R4	CACAGCCTCTTGACTTCCGAG	TCGTCCTACTCGCTCCCCCT	+737 +819
SMAD7-R5 (neg.)	CATCCCATCTATTCTCACGCC	AGACAGCCCCTTAAAGTTAGC	+23.7kb

Target sequences of the shRNAs used to target RINF knockdown

shRNA name	**Target mRNA Sequence**
shControl	CCUGGGCAAAGAAUGGACAAU
shRINF#4	GAAUGGACAAUCAGUUUCCUU
shRINF#3	CCUUUGAUUCUUUCCGACCAU

Detection of SMAD7 protein or 5hmC by immunofluorescence.

CD34+ cells were washed once (5 minutes in ice-cold PBS 1X), and fixed in PFA 2% (8 minutes at room temperature), permeabilized with Triton X100 0.1% (5 minutes at 4°C), incubated 1 hour in blocking buffer (PBS 1X, 5% Fetal Bovine Serum, 2% BSA) before overnight staining with primary mouse monoclonal anti-SMAD7 antibody (1:200, Santa-Cruz, #sc-365846). For 5hmC detection, two additional steps were performed after the permeabilization step, one DNA denaturation step in 3M HCl (5 minutes at

RT) followed by a five-minute neutralization step in TE buffer (10 mM Tris, 1 mM EDTA) at pH8. The primary anti-5hmC polyclonal rabbit antibody (1:2000) was purchased from Active motif (Cat.N°39769). Two secondary antibodies (1:1.000) coupled with AlexaFluor 647 fluorescent dye were purchased from Lifetechnologies, a goat anti-rabbit antibody (Cat # A32733) for 5hmC detection and a goat anti-mouse antibody (Cat # A32723) for SMAD7 detection. Images of 5hmC and SMAD7 staining were acquired on a wide-field Nikon Eclipse TE2000 microscope through a ×20 objective, with a Cascade CDD camera (Photometrics). The quantification of intensity was performed after background subtraction (Fiji). The fluorescent intensity was measured in 100–120 cells for the two conditions.

Cell culture with AML cell lines

Human leukemic cell lines were purchased from DSMZ (MV4-11; Braunschweig, Germany) and from the American Type Culture Collection (K562; Molsheim, France). UT7 5.3 cells were kindly provided by Isabelle Dusanter-Fourt (Cochin Institute, Paris, France). K562 and MV4-11 were cultured in RPMI 1640 medium supplemented with 10% fetal calf serum (FCS), 2 mM L-Glutamine, 50 U/ml penicillin G and 50 µg/ml streptomycin (Life Technologies, Saint-Aubin, France). UT7 5.3 cells were cultured in minimum essential medium (MEM) α medium containing 10% of FCS, 2 mM L-Glutamine, 50 U/ml penicillin G and 50 µg/ml streptomycin (Life Technologies, Saint-Aubin, France) and 2,5 ng/µl of GM-CSF (Miltenyi Biotec, France). Cells were grown in mycoplasma-free conditions, regularly tested.

Microarray and data analysis

Analysis of global gene expression profiles. After validation of RNA quality with Bioanalyzer 2100 (using Agilent RNA6000 nano chip kit), 400 ng of total RNA was reverse transcribed following the Genechip WT plus Reagent kit (Affymetrix). Briefly, the resulting double strand cDNA was used for in vitro transcription with T7 RNA pol. After purification, 15 µg of cRNA was used for reverse transcription with random primers. The cDNA obtained was then purified and fragmented. After control of fragmentation using Bioanalyzer 2100, cDNA was end labeled with biotin using Terminal Transferase (using the WT terminal labeling kit of Affymetrix). cDNA was then hybridized to GeneChip® Human Transcriptome Analysis 2.0 (Affymetrix) at 45°C for

17 hours. After overnight hybridization, chips were washed on the fluidic station FS450 following specific protocols (Affymetrix) and scanned using the GCS3000 7G. The image was then analyzed with Expression Console software (Affymetrix) to obtain raw data (cel files) and metrics for Quality Controls. Minimum information about a microarray experiment (MIAME)–compliant documentation of the microarray experiments have been deposited in Gene Expression Omnibus under the accession number GSE140770.

The Affymetrix HTA2 dataset analysis was performed by GenoSplice technology (www.genosplice.com). Data were normalized using quantile normalization. Background corrections were made with antigenomic probes and probes were selected according to their %GC, cross-hybridization status and potential overlap with repeat region as previously described (2,3). Only probes targeting exons and exon-exon junctions annotated from FAST DB® transcripts (release fastdb_2014_1) were selected (4,5). Only genes expressed in at least one compared condition were analyzed. To be considered to be expressed, the DABG P-value had to be ≤0.05 for at least half of the gene probes. We performed an unpaired Student's t-test to compare gene intensities between *shRINF* and *shControl* cells. Genes were considered significantly regulated when fold-change was ≥1.5, a fold-change noticed for *CXXC5* genes for which we have validated the knock-down by q-RT-PCR and western-blot analysis.

Supplementary Table S1.

ID	Symbol	Entrez Gene Name	Location	Type(s)
AATK	AATK	apoptosis associated tyrosine kinase	Cytoplasm	kinase
EMR3	ADGRE3	adhesion G protein-coupled receptor E3	Plasma Membrane	G-protein cou
ADM5	ADM5	adrenomedullin 5 (putative)	Other	other
ADRA2C	ADRA2C	adrenoceptor alpha 2C	Plasma Membrane	G-protein cou
AGAP4	AGAP6 (includes	ArfGAP with GTPase domain, ankyrin repeat and PH domain 5	Other	other
AKAP7	AKAP7	A-kinase anchoring protein 7	Plasma Membrane	other
ANXA2R	ANXA2R	annexin A2 receptor	Plasma Membrane	other
B3GNT4	B3GNT4	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 4	Plasma Membrane	enzyme
BEND5	BEND5	BEN domain containing 5	Cytoplasm	other
C15orf59	C15orf59	chromosome 15 open reading frame 59	Other	other
C1QB	C1QB	complement C1q B chain	Extracellular Space	peptidase
C2orf66	C2orf66	chromosome 2 open reading frame 66	Other	other
C3orf80	C3orf80	chromosome 3 open reading frame 80	Other	other
C8orf74	C8orf74	chromosome 8 open reading frame 74	Other	other
C9orf131	C9orf131	chromosome 9 open reading frame 131	Other	other
CACNG5	CACNG5	calcium voltage-gated channel auxiliary subunit gamma 5	Plasma Membrane	ion channel
CALML5	CALML5	calmodulin like 5	Cytoplasm	other
CARS	CARS	cysteinyl-tRNA synthetase	Cytoplasm	enzyme
CCDC144NL	CCDC144NL	coiled-coil domain containing 144 family, N-terminal like	Other	other
CDH4	CDH4	cadherin 4	Plasma Membrane	other
CGB8	CGB3 (includes (	chorionic gonadotropin beta subunit 3	Extracellular Space	other
CHADL	CHADL	chondroadherin like	Extracellular Space	other
CNTN4	CNTN4	contactin 4	Plasma Membrane	enzyme
COL18A1	COL18A1	collagen type XVIII alpha 1 chain	Extracellular Space	other
COL4A2	COL4A2	collagen type IV alpha 2	Extracellular Space	other
COL6A1	COL6A1	collagen type VI alpha 1	Extracellular Space	other
COX6B2	COX6B2	cytochrome c oxidase subunit 6B2	Cytoplasm	enzyme
CPOX	CPOX	coproporphyrinogen oxidase	Cytoplasm	enzyme
CRHR1	CRHR1	corticotropin releasing hormone receptor 1	Plasma Membrane	G-protein cou
CRP	CRP	C-reactive protein, pentraxin-related	Extracellular Space	other
CSDC2	CSDC2	cold shock domain containing C2	Cytoplasm	other
CXXC5	**CXXC5**	**CXXC finger protein 5**	**Nucleus**	**other**
CYP4F30P	CYP4F30P	cytochrome P450 family 4 subfamily F member 30, pseudogene	Other	other
DAZ2	DAZ2	deleted in azoospermia 2	Cytoplasm	translation reg
DLGAP3	DLGAP3	DLG associated protein 3	Cytoplasm	other
DNAJC5G	DNAJC5G	DnaJ heat shock protein family (Hsp40) member C5 gamma	Other	other
DND1	DND1	DND microRNA-mediated repression inhibitor 1	Cytoplasm	other
DRGX	DRGX	dorsal root ganglia homeobox	Nucleus	other
DUSP4	DUSP4	dual specificity phosphatase 4	Nucleus	phosphatase
EPN2-IT1	EPN2-IT1	EPN2 intronic transcript 1	Other	other
FAM153C	FAM153C	family with sequence similarity 153, member C	Other	other
FAM229A	FAM229A	family with sequence similarity 229 member A	Other	other
FAM26D	FAM26D	family with sequence similarity 26 member D	Other	other
FGD5P1	FGD5P1	FYVE, RhoGEF and PH domain containing 5 pseudogene 1	Other	other
GATA3	GATA3	GATA binding protein 3	Nucleus	transcription r
GJC1	GJC1	gap junction protein gamma 1	Plasma Membrane	ion channel
GPIHBP1	GPIHBP1	glycosylphosphatidylinositol anchored high density lipoprotein binding protein 1	Plasma Membrane	transporter
GPR139	GPR139	G protein-coupled receptor 139	Plasma Membrane	G-protein cou
GPR55	GPR55	G protein-coupled receptor 55	Plasma Membrane	G-protein cou
GPRIN2	GPRIN2	G protein regulated inducer of neurite outgrowth 2	Other	other
GRASPOS	GRASPOS	GRP1-associated scaffold protein opposite strand	Other	other
GSX2	GSX2	GS homeobox 2	Nucleus	transcription r
HAR1B	HAR1B	highly accelerated region 1B (non-protein coding)	Other	other
HCAR2	HCAR2	hydroxycarboxylic acid receptor 2	Plasma Membrane	G-protein cou
HCG27	HCG27	HLA complex group 27 (non-protein coding)	Other	other
HHIPL1	HHIPL1	HHIP like 1	Other	other
HNRNPKP3	HNRNPKP3	heterogeneous nuclear ribonucleoprotein K pseudogene 3	Other	other
HOXB5	HOXB5	homeobox B5	Nucleus	transcription r
HOXC9	HOXC9	homeobox C9	Nucleus	transcription r
HSPA6	HSPA6	heat shock protein family A (Hsp70) member 6	Nucleus	enzyme
IGFL1	IGFL1	IGF like family member 1	Extracellular Space	other
IL23A	IL23A	interleukin 23 subunit alpha	Extracellular Space	cytokine
IL36A	IL36A	interleukin 36, alpha	Extracellular Space	cytokine
IRF2BP1	IRF2BP1	interferon regulatory factor 2 binding protein 1	Nucleus	transcription r
IRS4	IRS4	insulin receptor substrate 4	Plasma Membrane	other
KCNA1	KCNA1	potassium voltage-gated channel subfamily A member 1	Plasma Membrane	ion channel
KCNC1	KCNC1	potassium voltage-gated channel subfamily C member 1	Plasma Membrane	ion channel
KCNK17	KCNK17	potassium two pore domain channel subfamily K member 17	Plasma Membrane	ion channel
KRTAP2-1	KRTAP2-1	keratin associated protein 2-1	Other	other
KRTAP6-3	KRTAP6-3	keratin associated protein 6-3	Other	other
KRTAP9-8	KRTAP9-8	keratin associated protein 9-8	Other	other
LCE2A	LCE2A	late cornified envelope 2A	Other	other

LCE2C	LCE2C/LCE2D	late cornified envelope 2D	Other	other
LEMD1	LEMD1	LEM domain containing 1	Other	other
LIN28A	LIN28A	lin-28 homolog A	Cytoplasm	other
LINC00641	LINC00641	long intergenic non-protein coding RNA 641	Other	other
LINC00927	LINC00927	long intergenic non-protein coding RNA 927	Other	other
LINC00968	LINC00968	long intergenic non-protein coding RNA 968	Other	other
LINC00987	LINC00987	long intergenic non-protein coding RNA 987	Other	other
LINC00996	LINC00996	long intergenic non-protein coding RNA 996	Other	other
LINC01010	LINC01010	long intergenic non-protein coding RNA 1010	Other	other
LMOD3	LMOD3	leiomodin 3	Other	other
LRRC32	LRRC32	leucine rich repeat containing 32	Plasma Membrane	other
LRRC4B	LRRC4B	leucine rich repeat containing 4B	Plasma Membrane	other
LRRC52	LRRC52	leucine rich repeat containing 52	Other	other
MACROD2	MACROD2	MACRO domain containing 2	Nucleus	enzyme
MAFG-AS1	MAFG-AS1	MAFG antisense RNA 1 (head to head)	Other	other
MAP4K1	MAP4K1	mitogen-activated protein kinase kinase kinase kinase 1	Cytoplasm	kinase
MIR103A1	mir-103	microRNA 107	Cytoplasm	microRNA
MIR1257	mir-1257	microRNA 1257	Cytoplasm	microRNA
MIR1288	mir-1288	microRNA 1288	Cytoplasm	microRNA
MIR129-2	mir-129	microRNA 129-2	Cytoplasm	microRNA
MIR1301	mir-1301	microRNA 1301	Cytoplasm	microRNA
MIR152	mir-148	microRNA 148a	Cytoplasm	microRNA
MIR149	mir-149	microRNA 149	Cytoplasm	microRNA
MIR377	mir-154	microRNA 494	Other	microRNA
MIR1972-1	mir-1972	microRNA 1972-1	Cytoplasm	microRNA
MIR211	mir-204	microRNA 204	Cytoplasm	microRNA
MIR29C	mir-29	microRNA 29a	Cytoplasm	microRNA
MIR371A	mir-290	microRNA 372	Other	microRNA
MIR3065	mir-3065	microRNA 3065	Cytoplasm	microRNA
MIR320B2	mir-320	microRNA 320a	Cytoplasm	microRNA
MIR3689C	mir-3689	microRNA 3689b	Cytoplasm	microRNA
MIR4654	mir-4654	microRNA 4654	Cytoplasm	microRNA
MIR4803	mir-4803	microRNA 4803	Cytoplasm	microRNA
MIR489	mir-489	microRNA 489	Cytoplasm	microRNA
MIR550A1	mir-550	microRNA 550a-1	Cytoplasm	microRNA
MIR569	mir-569	microRNA 569	Cytoplasm	microRNA
MIR581	mir-581	microRNA 581	Cytoplasm	microRNA
MIR615	mir-615	microRNA 615	Cytoplasm	microRNA
MIR647	mir-647	microRNA 647	Cytoplasm	microRNA
MIR200A	mir-8	microRNA 200c	Cytoplasm	microRNA
MIR4291	MIR4291	microRNA 4291	Cytoplasm	microRNA
MIR4440	MIR4440	microRNA 4440	Cytoplasm	microRNA
MIR4692	MIR4692	microRNA 4692	Cytoplasm	microRNA
MRGPRG	MRGPRG	MAS related GPR family member G	Plasma Membrane	G-protein cou
MUC6	MUC6	mucin 6, oligomeric mucus/gel-forming	Extracellular Space	other
MYADML	MYADML	myeloid associated differentiation marker like (pseudogene)	Other	other
NEURL2	NEURL2	neuralized E3 ubiquitin protein ligase 2	Cytoplasm	other
NGB	NGB	neuroglobin	Cytoplasm	transporter
NHSL2	NHSL2	NHS like 2	Other	other
NKX6-3	NKX6-3	NK6 homeobox 3	Other	other
NOXA1	NOXA1	NADPH oxidase activator 1	Other	other
NPIPB7	NPIPB7	nuclear pore complex interacting protein family member B7	Other	other
NRGN	NRGN	neurogranin	Other	other
NRTN	NRTN	neurturin	Extracellular Space	growth factor
NUDT16L1	NUDT16L1	nudix hydrolase 16 like 1	Cytoplasm	other
OGFR	OGFR	opioid growth factor receptor	Plasma Membrane	other
OLFML3	OLFML3	olfactomedin like 3	Extracellular Space	other
OMD	OMD	osteomodulin	Extracellular Space	other
OR2A25	OR2A25	olfactory receptor family 2 subfamily A member 25	Plasma Membrane	G-protein cou
OR2Y1	OR2Y1	olfactory receptor family 2 subfamily Y member 1	Plasma Membrane	G-protein cou
OR5C1	OR5C1	olfactory receptor family 5 subfamily C member 1	Plasma Membrane	G-protein cou
OR5R1	OR5R1	olfactory receptor family 5 subfamily R member 1 (gene/pseudogene)	Plasma Membrane	G-protein cou
OXCT2	OXCT2	3-oxoacid CoA-transferase 2	Cytoplasm	other
P2RY13	P2RY13	purinergic receptor P2Y13	Plasma Membrane	G-protein cou
PARVG	PARVG	parvin gamma	Cytoplasm	other
PAX7	PAX7	paired box 7	Nucleus	transcription r
PCDHA1	PCDHA1	protocadherin alpha 1	Plasma Membrane	other
PF4V1	PF4V1	platelet factor 4 variant 1	Extracellular Space	cytokine
PHF13	PHF13	PHD finger protein 13	Nucleus	other
POTEKP	POTEKP	POTE ankyrin domain family member K, pseudogene	Cytoplasm	other
POU5F1P3	POU5F1P3	POU class 5 homeobox 1 pseudogene 3	Other	other
PRR21	PRR21	proline rich 21	Other	other
PRRT3-AS1	PRRT3-AS1	PRRT3 antisense RNA 1	Other	other
PRSS41	PRSS41	protease, serine 41	Plasma Membrane	peptidase
PTPRF	PTPRF	protein tyrosine phosphatase, receptor type F	Plasma Membrane	phosphatase
PWRN1	PWRN1	Prader-Willi region non-protein coding RNA 1	Other	other
RAMP2-AS1	RAMP2-AS1	RAMP2 antisense RNA 1	Other	other
RARRES2	RARRES2	retinoic acid receptor responder 2	Plasma Membrane	transmembrar

RBMY1E	RBMY1A1 (inclu	RNA binding motif protein, Y-linked, family 1, member A1	Nucleus	other
C15orf60	REC114	REC114 meiotic recombination protein	Other	other
RFPL3	RFPL1/RFPL3	ret finger protein like 3	Other	other
RHBDL2	RHBDL2	rhomboid like 2	Plasma Membrane	peptidase
RPH3AL	RPH3AL	rabphilin 3A-like (without C2 domains)	Plasma Membrane	other
S1PR1	S1PR1	sphingosine-1-phosphate receptor 1	Plasma Membrane	G-protein cou
SCARA3	SCARA3	scavenger receptor class A member 3	Plasma Membrane	transmembrar
SCGB3A1	SCGB3A1	secretoglobin family 3A member 1	Extracellular Space	cytokine
SDC1	SDC1	syndecan 1	Plasma Membrane	enzyme
SIAH3	SIAH3	siah E3 ubiquitin protein ligase family member 3	Cytoplasm	other
SIGLEC16	SIGLEC16	sialic acid binding Ig like lectin 16 (gene/pseudogene)	Plasma Membrane	other
SIX3	SIX3	SIX homeobox 3	Nucleus	transcription r
SIX6	SIX6	SIX homeobox 6	Nucleus	transcription r
SLC17A6	SLC17A6	solute carrier family 17 member 6	Plasma Membrane	transporter
SLC52A1	SLC52A1	solute carrier family 52 member 1	Plasma Membrane	other
SMAD7	**SMAD7**	**SMAD family member 7**	**Nucleus**	**transcription**
SMIM18	SMIM18	small integral membrane protein 18	Other	other
SMIM22	SMIM22	small integral membrane protein 22	Other	other
SMPD3	SMPD3	sphingomyelin phosphodiesterase 3	Cytoplasm	enzyme
SMTNL1	SMTNL1	smoothelin like 1	Cytoplasm	other
SNX21	SNX21	sorting nexin family member 21	Cytoplasm	transporter
SOWAHD	SOWAHD	sosondowah ankyrin repeat domain family member D	Other	other
SOX2	SOX2	SRY-box 2	Nucleus	transcription r
SPATA21	SPATA21	spermatogenesis associated 21	Other	other
SPRR2D	SPRR2D	small proline rich protein 2D	Cytoplasm	other
SSTR1	SSTR1	somatostatin receptor 1	Plasma Membrane	G-protein cou
SYNGR4	SYNGR4	synaptogyrin 4	Plasma Membrane	other
TAAR5	TAAR5	trace amine associated receptor 5	Plasma Membrane	G-protein cou
TEPP	TEPP	testis, prostate and placenta expressed	Other	other
TFAP2E	TFAP2E	transcription factor AP-2 epsilon	Nucleus	transcription r
TIE1	TIE1	tyrosine kinase with immunoglobulin like and EGF like domains 1	Plasma Membrane	kinase
TMED6	TMED6	transmembrane p24 trafficking protein 6	Cytoplasm	other
TMPRSS4-AS	TMPRSS4-AS1	TMPRSS4 antisense RNA 1	Other	other
TOLLIP-AS1	TOLLIP-AS1	TOLLIP antisense RNA 1 (head to head)	Other	other
TPSD1	TPSD1	tryptase delta 1	Extracellular Space	peptidase
TTC30A	TTC30A	tetratricopeptide repeat domain 30A	Other	other
TUSC1	TUSC1	tumor suppressor candidate 1	Other	other
USP3-AS1	USP3-AS1	USP3 antisense RNA 1	Other	other
WFDC2	WFDC2	WAP four-disulfide core domain 2	Extracellular Space	other
ZDHHC15	ZDHHC15	zinc finger DHHC-type containing 15	Cytoplasm	enzyme
ZMIZ1	ZMIZ1	zinc finger MIZ-type containing 1	Nucleus	other
ZNF423	ZNF423	zinc finger protein 423	Nucleus	transcription r
ZNF836	ZNF836	zinc finger protein 836	Other	other

Supplementary Table S1. List of genes commonly down-regulated in K562 and UT7$_{5.3}$ knockdown cells (n=193). For the microarray experiments, we used the GeneChip™ Human Transcriptome Array 2.0 (HTA2).

LEGENDS TO SUPPLEMENTAL FIGURES

Figure S1. The *CXXC5/RINF* gene is located at chromosome 5q31.2 within the CDR associated with higher risk in MDS and AML. Genes, as well as sequence tag site markers, are listed in the order of their physical positions (megabase, Mb) along chromosome 5q from centromere (upper) to telomere (bottom), according to the most recent human genome assembly at http://ensembl.org (Human, GRCh38.p12). *CXXC5* gene is located at less than 20kb from *UBE2D2*, the more distal gene of the high-risk CDR (grey) investigated by Liu *et al*. in 2007 (6). According to a SNP-array study published in 2012 by Jerez *et al*. (7), and according to the GRCH37/hg19 genome assembly version (released in 2009) that was probably used for this study, the *CXXC5* gene is located inside the refined-CDR associated with high-risk.

Figure S2. The pattern of *RINF* mRNA expression profile during erythroid maturation of human hematopoietic progenitors cultured *in vitro* was confirmed in three independent datasets. (A) *CXXC5* mRNA expression profile was obtained by querying the Human Erythroblast Maturation Database described by Merryweather-Clarke *et al*.(8) In that study, PBMC-derived erythroblasts from 38 adult donors were grown *in vitro* and cells at the CFU-E, Pro-E, Int-E and Late-E stages were sorted by FACS according to CD36, CD71 and glycophorin A expression. **(B)** *CXXC5* mRNA expression data were extracted from an RNA-seq dataset performed by An *et al.*(9) CD34+ cells from cord blood were cultured in a 3-phases culture system to produce primary human erythroid cells at different stages of terminal differentiation. Erythroblastic populations were isolated through sorting by FACS according to expression of cell surface markers glycophorin A, band 3, and α4-integrin. **(C)** *CXXC5* mRNA expression was extracted from a public microarray dataset performed by Keller *et al.*(10) Briefly, CD34+ cells isolated from adult PBMC were cultures in a serum-free medium supplemented with IL3, SCF, and EPO for the indicated number of days.

Figure S3. In human CD34+ cells, *RINF* silencing did not affect the total number of colonies. Human CD34+ cells were isolated from bone marrow of healthy donors and transduced with a lentiviral vector that drives shRNA expression (pTRIPDU3). Cells were then seeded (500 cells per well) into methylcellulose plates containing

cytokines (IL3, SCF, G-CSF, GM-CSF, and EPO) enabling the formation of erythroid colonies (both large and small BFU-Es), myeloid colonies (CFU-G, CFU-M, or CFU-GM), or mixed colonies (CFU-GEMM). Colony Forming Cells were counted at 14 days of culture. **(A)** Schematic representation of the pTRIPDU3 (*lower panel*) lentiviral vector used to express a short hairpin RNA. The specificity and the efficacy of the sequences were previously described (11). The pTRIPDU3 vector drives the expression of the shRNA under a H1 promoter and the GFP expression gene under an EF1α promoter. After two days of expansion in (IL3, SCF, IL6), human CD34+ cells were transduced with the pTRIPDU3 vector. Of note, pTRIPDU3-transduced cells were FACS-sorted for GFP expression at two days after transduction. **(B)** This histogram represents the level of *RINF* mRNA obtained with the two sequences targeting RINF (*shRINF#3* or *shRINF#4*) and detected by quantitative RT-PCR. Values are expressed in percentage of the *shControl* condition. (student T-test, * p< 0.05) **(C)** CD34+ cells were isolated from bone marrow of healthy donors (n=4). The histogram represents the total number of colonies enumerated (*left panel*) and the relative proportion of each type of colony (*right panel*). For each of the four donors, the proportions are also visualized by pairs of red dots. (student T-test, * p< 0.05) **(D)** CD34+ were isolated from bone marrow of healthy donors (n=4, *left panel*) or cord-blood (n=6, *right panel*). Histograms give the mean number of small and large BFU-E-derived colonies. The error bars indicate donor-to-donor variability for each vector (± SEM). The trend was not significant (student t-test). **(E)** Histogram represents the number of BFU-E-derived colonies enumerated from 1000 MEP (Megakaryocyte–Erythroid Progenitor)-enriched seeded cells (*right panel*). Data shown are means ± SD (n=3). The increased number of BFU-Es enumerated in the shRNA-RINF condition was statistically significant (student T-test, * p< 0.05). The gating strategy used for MEP-enrichment is shown (*right panel*). The protocol is adapted from Manz *et al*.(1) For this, CD34+ cells were isolated from primary human bone marrow and stained with CD34, CD38, CD10, CD123 and CD45RA (see Supplemental Methods). We have verified that these MEP cells only generate erythroid colonies (BFU-Es) in our experimental conditions.

Figure S4. *RINF* silencing alters the TGFβ signaling pathway in K562 and UT7$_{5.3}$ cells and accelerates cell hemoglobinization triggered by hemin and EPO, respectively. The pTRIPDU3 lentiviral system was used to constitutively express a

shRNA targeting *RINF* expression in K562 (*upper panels*) and $UT7_{5.3}$ (*lower panels*) cell line. Two days after transduction, cells positive for GFP expression were sorted. **(A and B)** Cells were then respectively treated for the indicated time with hemin (40µM) for K562, or EPO (5UI/ml) for $UT7_{5.3}$. **(A)** The relative expression of RINF was measured by Western-Blot analysis (see Methods). ACTIN (ACTB) was used as a loading control. **(B)** For hemoglobin production analysis, the benzidine test was used. The percentage of benzidine positive cells was established by counting at least 300 cells per sample. **(C)** Transcriptomic analysis was performed with HTA 2.0 Human Transcriptome Array 2.0 GeneChip (Affymetrix), and previously described (12). Venn diagram representing the number of genes downregulated after *RINF* knockdown. 193 genes were downregulated in K562 and $UT7_{5,3}$ cell lines with at least a 1.25-fold-change (log 2), corresponding to a non-stringent condition. In absolute value, this cut-off corresponds to a 2.38 expression fold-change. We applied QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) tool to this gene list. IPA revealed an enrichment in genes of the TGFβ signaling pathway and *SMAD7* as a potential RINF target. **(D)** Histograms represent the *RINF* or *SMAD7* mRNA levels detected by quantitative RT-PCR for cells. Values are expressed in percentage of the *shControl* condition. A student T test (unpaired) was performed to assess if the observed relative difference was statistically significant. **(E)** Among the list of 193 genes, 39 were also commonly downregulated in MV4-11 cell line. This gene list is presented as well as the heatmap (*right panel*) of the 39 genes representing the relative mRNA expression fold-change (log2) between *shRNA-RINF* and *shRNA-Control* expressing cell lines. *CXXC5* and *SMAD7* are indicated in red.

Figure S5. Ectopic overexpression of *RINF* delays hemoglobinization of human hematopoietic cell lines K562 and $UT7_{5.3}$ triggered by hemin and EPO, respectively. The MigR1 retroviral system (schematized in *upper panel* A) was used to constitutively express RINF in two cell lines. A few days after transduction, cells were sorted for GFP expression. Cells were then respectively treated with hemin (40µM) for K562 or EPO (5UI/ml) for $UT7_{5.3.}$ **(A)** The relative expression of RINF was measured by Western-Blot analysis using a polyclonal rabbit antibody (see Methods) that detects a specific band at 33 kDa in total extracts. ACTIN (ACTB) was used as a loading control. **(B)** For hemoglobin production analysis, the benzidine test was used.

Percentage of benzidine positive cells was established by counting about 300 cells per sample (Student T-test). Pictures represent one of three independent experiments.

Figure S6. *SMAD7* gene expression correlates with *CXXC5* in CD34+ cells isolated from healthy donors or MDS patients with del(5q) and is a favorable prognostic indicator in MDS. For these analyses, two microarray datasets (Affymetrix GeneChip Human Genome U133 Plus 2.0 arrays) respectively performed by Pellagatti *et al.* (183 patients and 17 healthy control donors) or Gerstund *et al*. (159 patients and 17 healthy control donors) were downloaded from the Gene Expression Omnibus website (http://www.ncbi.nlm.nih.gov/geo/) by using the followings accession numbers: GSE19429 or GSE58831. **(A)** Heatmap of Pearson correlation coefficient matrix represents, in red, the gene probesets that positively correlate with *SMAD7* (and in blue, the one that negatively correlate with *SMAD7*). To create this heatmap, gene expression data of the high-risk CDR genes (n=43) were extracted from GSE19429 performed by Pellagatti *et al*. (13) and correlated with *SMAD7* mRNA expression (204790_at). **(B)** Genes of the high risk CDR, as well as sequence tag site markers, are listed in the order of their physical positions (megabase, Mb) along chromosome 5q from centromere (upper) to telomere (bottom), according to the most recent human genome assembly at http://ensembl.org (Human, GRCh38.p12). **(C)** Histogram represents the *SMAD7* mRNA levels (probeset 204790_at) detected in the microarray dataset performed by Gerstund *et al*.(14) Values are expressed in arbitrary units. A student T test (unpaired) was performed to assess if the observed relative difference was statistically significant between the 3 groups. **(D)** Overall survival data were available for 123 patients, and a tertile stratification was performed. For this, the patients have been classified in 3 equivalent groups according to a high (n=41), an intermediate (n=41), or a low (n=41) *SMAD7* mRNA expression level. The Kaplan-Meier curves (for survival analysis) and the log-rank test were performed by using GraphPad Prism 7.0. P value (log-rank test) of the comparison of the various groups of patients is indicated. In green, the survival curve of the third of patients with the *lowest SMAD7* expression (n=41). The median survival was significantly shorter for this group than for the group with *intermediary* or *high SMAD7* expression (n=82). Note that the median *SMAD7* mRNA level was still significantly lower in the MDS group with "*high" SMAD7* than in the healthy control donors (Student t-Test, P=0.005).

Figure S7. Schematic representation of the RINF/SMAD7 signaling pathway that occurs during human erythropoiesis. The black box recapitulates the mains erythropoietic features obtained from immature erythroid cells (i.e. progenitors and proerythroblasts) with a high *RINF* mRNA expression, and the red box, the ones with attenuated/knocked-down *RINF* mRNA expression. RINF directly binds to *SMAD7* promoter to maintain its expression. *RINF* knockdown cells have a moderate loss of SMAD7, that is enough to sensitize erythroid cells to TGFβ-signaling and mitigate RBC expansion. This signaling could contribute to the inefficient erythropoiesis observed in malignant hemopathies. The *RINF* gene is rarely mutated in hematopoietic malignancies but several cytogenetic abnormalities (i.e. del(5q)(11,15-17), t(15;17)(11) MLL rearrangements, and t(8;21)(15) and gene mutations (such as GATA2 (15), WT1(15, 18) have been shown to affect *RINF* expression in MDS or AML.

Supplemental References

1. Manz MG, Miyamoto T, Akashi K, Weissman IL. Prospective isolation of human clonogenic common myeloid progenitors. Proceedings of the National Academy of Sciences of the United States of America 2002;99(18):11872-11877

2. Gandoura S, Weiss E, Rautou PE, *et al.* Gene- and exon-expression profiling reveals an extensive LPS-induced response in immune cells in patients with cirrhosis. J Hepatol 2013;58(5):936-948

3. Wang E, Aslanzadeh V, Papa F, Zhu H, de la Grange P, Cambi F. Global profiling of alternative splicing events and gene expression regulated by hnRNPH/F. PLoS One 2012;7(12):e51266

4. de la Grange P, Dutertre M, Correa M, Auboeuf D. A new advance in alternative splicing databases: from catalogue to detailed analysis of regulation of expression and function of human alternative splicing variants. BMC Bioinformatics 2007;8:180

5. de la Grange P, Dutertre M, Martin N, Auboeuf D. FAST DB: a website resource for the study of the expression regulation of human gene products. Nucleic acids research 2005;33(13):4276-4284

6. Liu TX, Becker MW, Jelinek J, *et al.* Chromosome 5q deletion and epigenetic suppression of the gene encoding alpha-catenin (CTNNA1) in myeloid cell transformation. Nature medicine 2007;13(1):78-83

7. Jerez A, Gondek LP, Jankowska AM, et al. Topography, clinical, and genomic correlates of 5q myeloid malignancies revisited. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012;30(12):1343-1349.

8. Merryweather-Clarke AT, Atzberger A, Soneji S, et al. Global gene expression analysis of human erythroid progenitors. Blood. 2011;117(13):e96-108.

9. An X, Schulz VP, Li J, et al. Global transcriptome analyses of human and murine terminal erythroid differentiation. Blood. 2014;123(22):3466-3477.

10. Keller MA, Addya S, Vadigepalli R, et al. Transcriptional regulatory network analysis of developing human erythroid progenitors reveals patterns of coregulation and potential transcriptional regulators. Physiol Genomics. 2006;28(1):114-128.

11. Pendino F, Nguyen E, Jonassen I, et al. Functional involvement of RINF, retinoid-inducible nuclear factor (CXXC5), in normal and tumoral human myelopoiesis. Blood. 2009;113(14):3172-3181.

12. Bruserud O, Reikvam H, Fredly H, et al. Expression of the potential therapeutic target CXXC5 in primary acute myeloid leukemia cells - high expression is associated with adverse prognosis as well as altered intracellular signaling and transcriptional regulation. Oncotarget. 2015;6(5):2794-2811.

13. Pellagatti A, Cazzola M, Giagounidis AA, et al. Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. Blood. 2006;108(1):337-345.

14. Gerstung M, Pellagatti A, Malcovati L, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. Nat Commun. 2015;6:5901.

15. Kuhnl A, Valk PJ, Sanders MA, et al. Downregulation of the Wnt inhibitor CXXC5 predicts a better prognosis in acute myeloid leukemia. Blood. 2015;125(19):2985-2994.

16. Stoddart A, Qian Z, Fernald AA, et al. Retroviral insertional mutagenesis identifies the del(5q) genes, CXXC5, TIFAB and ETF1, as well as the Wnt pathway, as potential targets in del(5q) myeloid neoplasms. Haematologica. 2016;101(6):e232-236

17. Treppendahl MB, Mollgard L, Hellstrom-Lindberg E, Cloos P, Gronbaek K. Downregulation but lack of promoter hypermethylation or somatic mutations of the potential tumor suppressor CXXC5 in MDS and AML with deletion 5q. Eur J Haematol. 2013;90(3):259-260.

18. Kim MS, Yoon SK, Bollig F, et al. A novel Wilms tumor 1 (WT1) target gene negatively regulates the WNT signaling pathway. J Biol Chem. 2010;285(19):14585-14593.

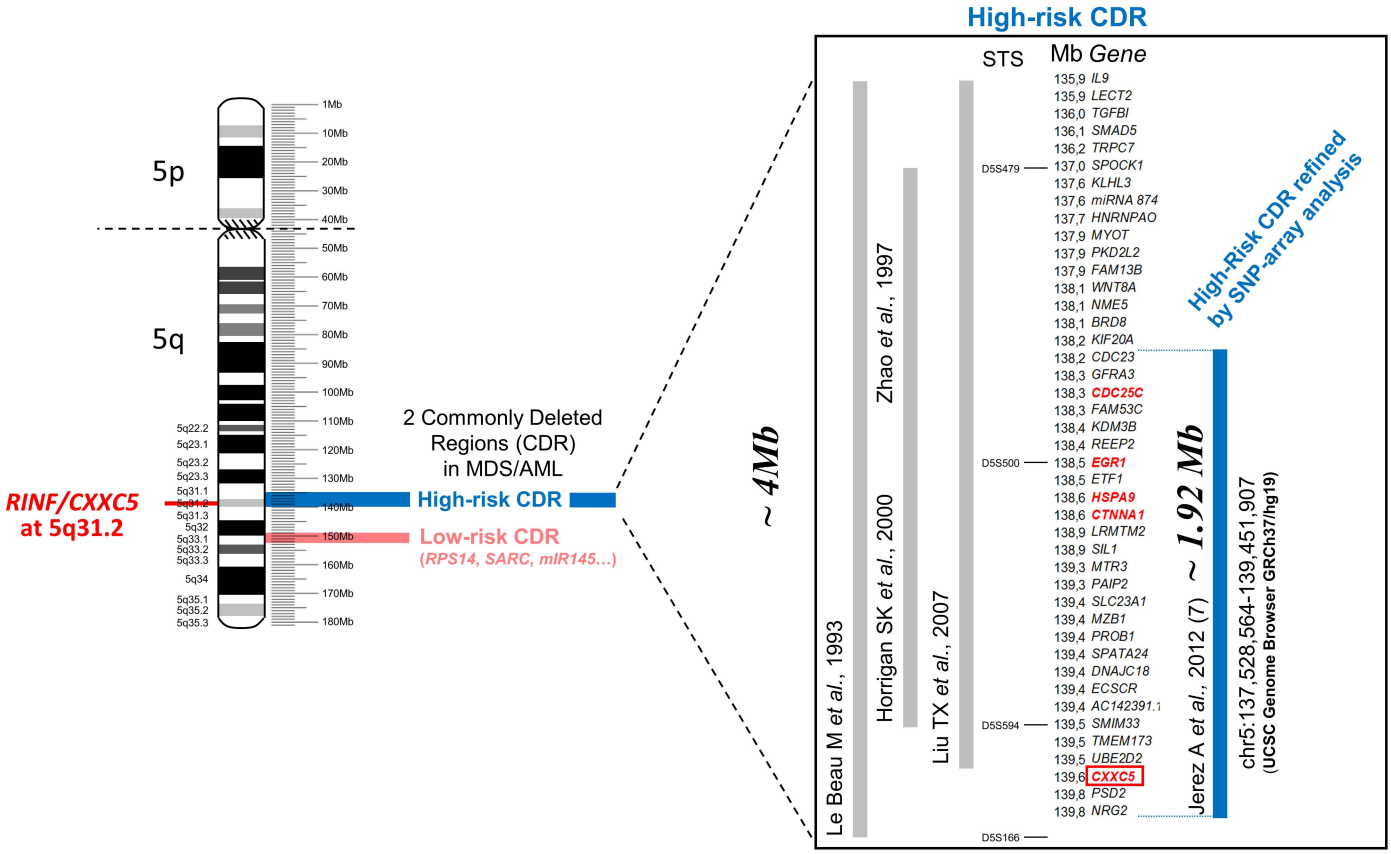


Figure S1
Astori and Matherat *et al.*

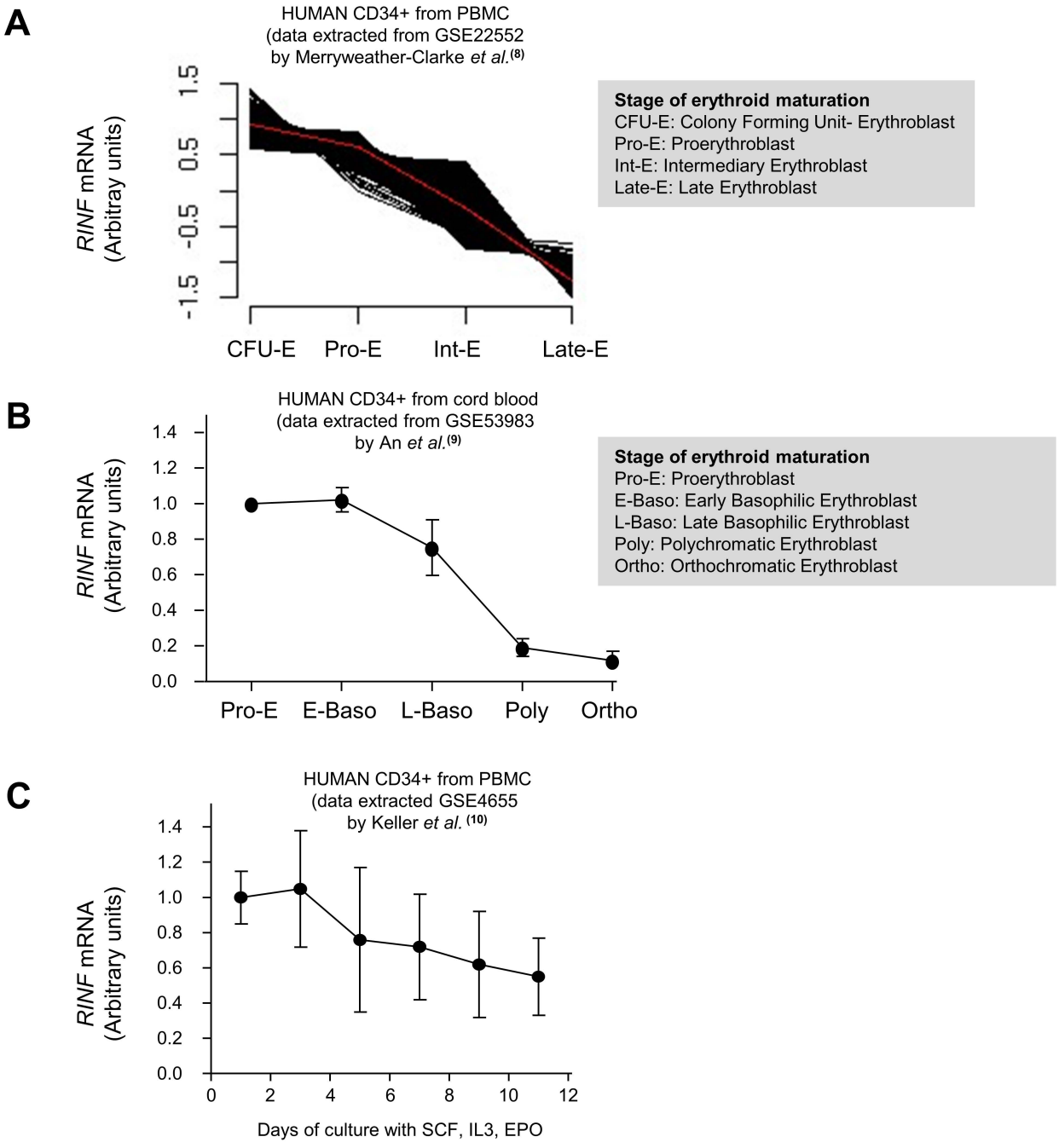


Figure S2
Astori and Matherat *et al.*

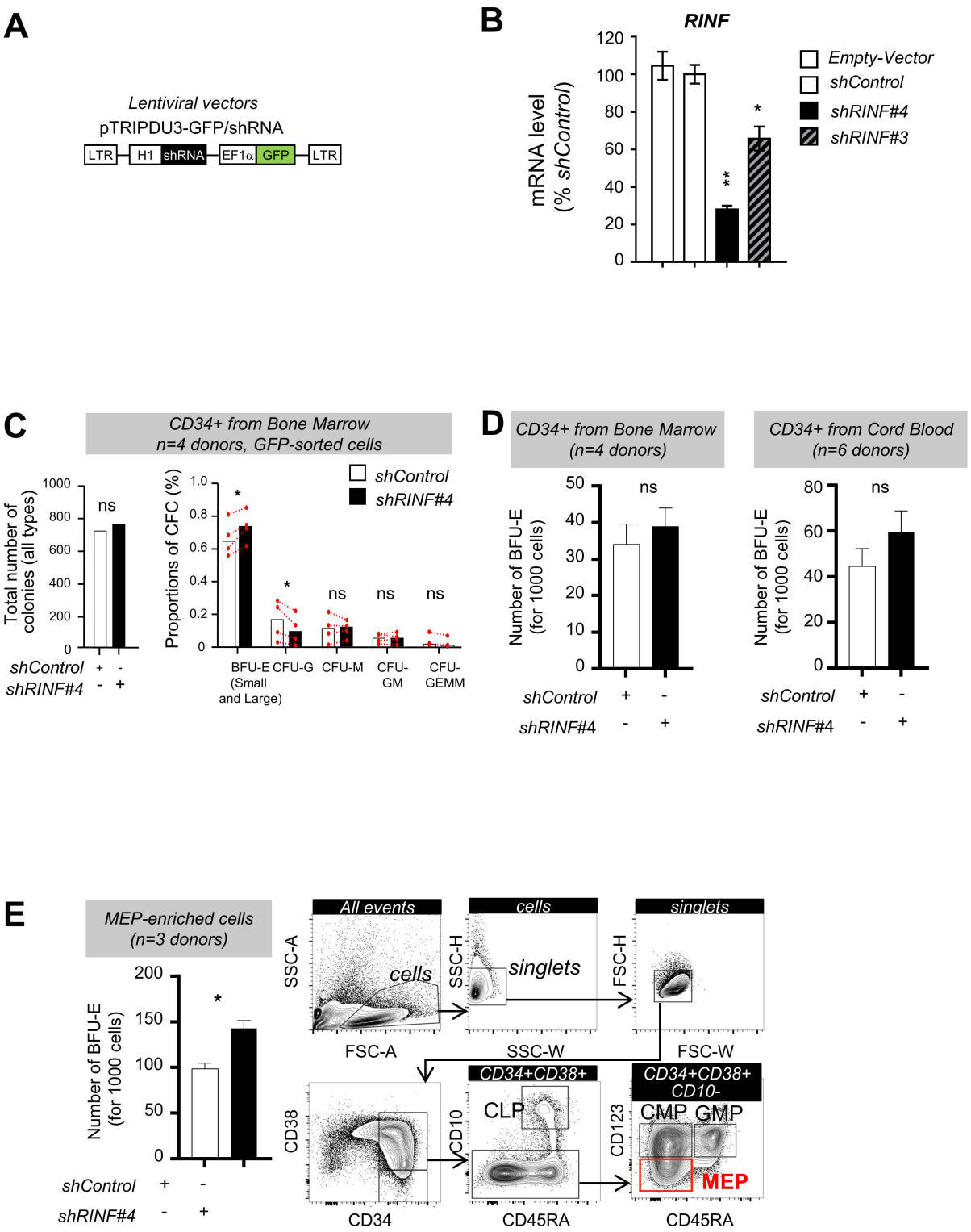


Figure S3
Astori and Matherat *et al.*

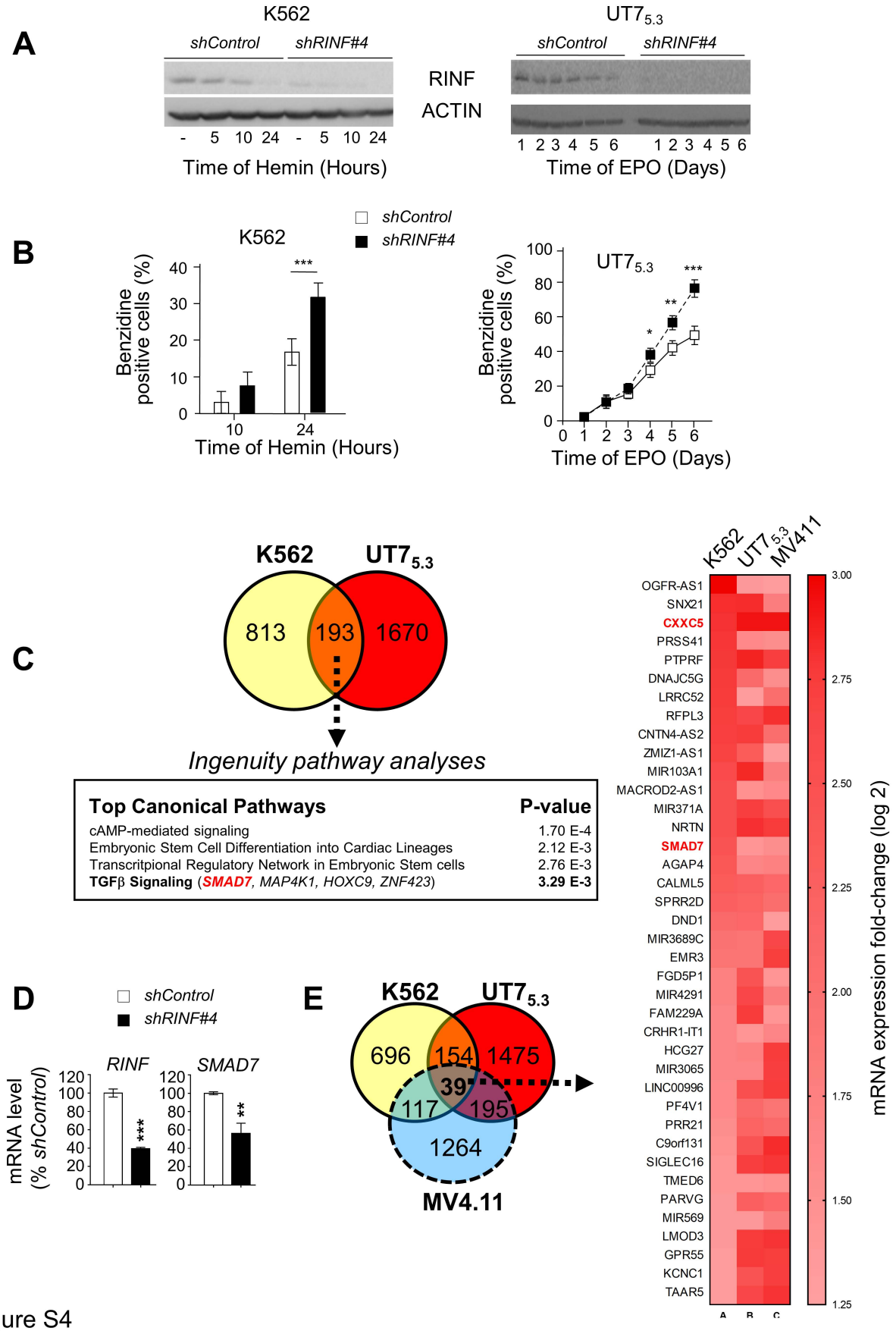


Figure S4
Astori and Matherat *et al*.

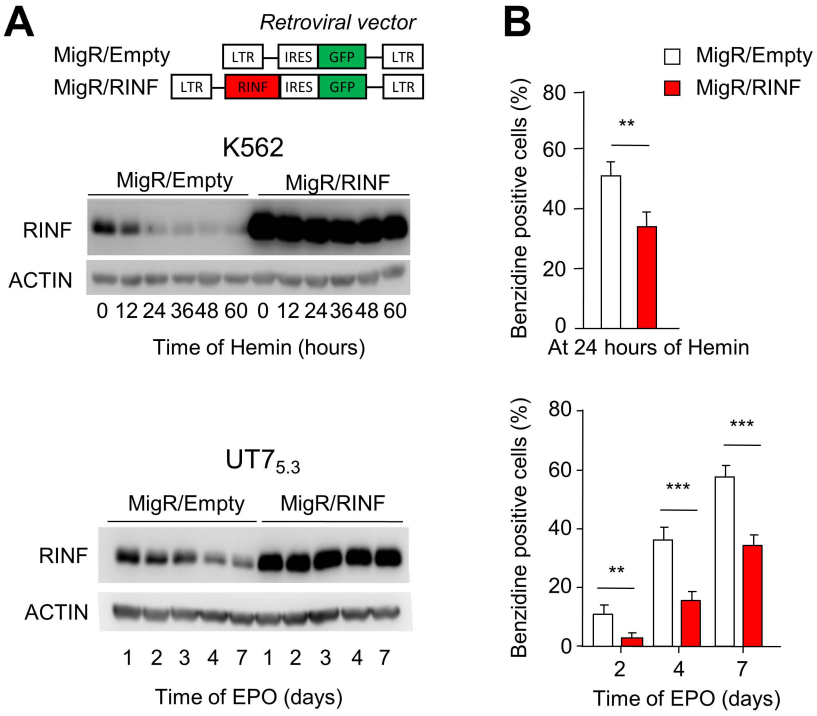
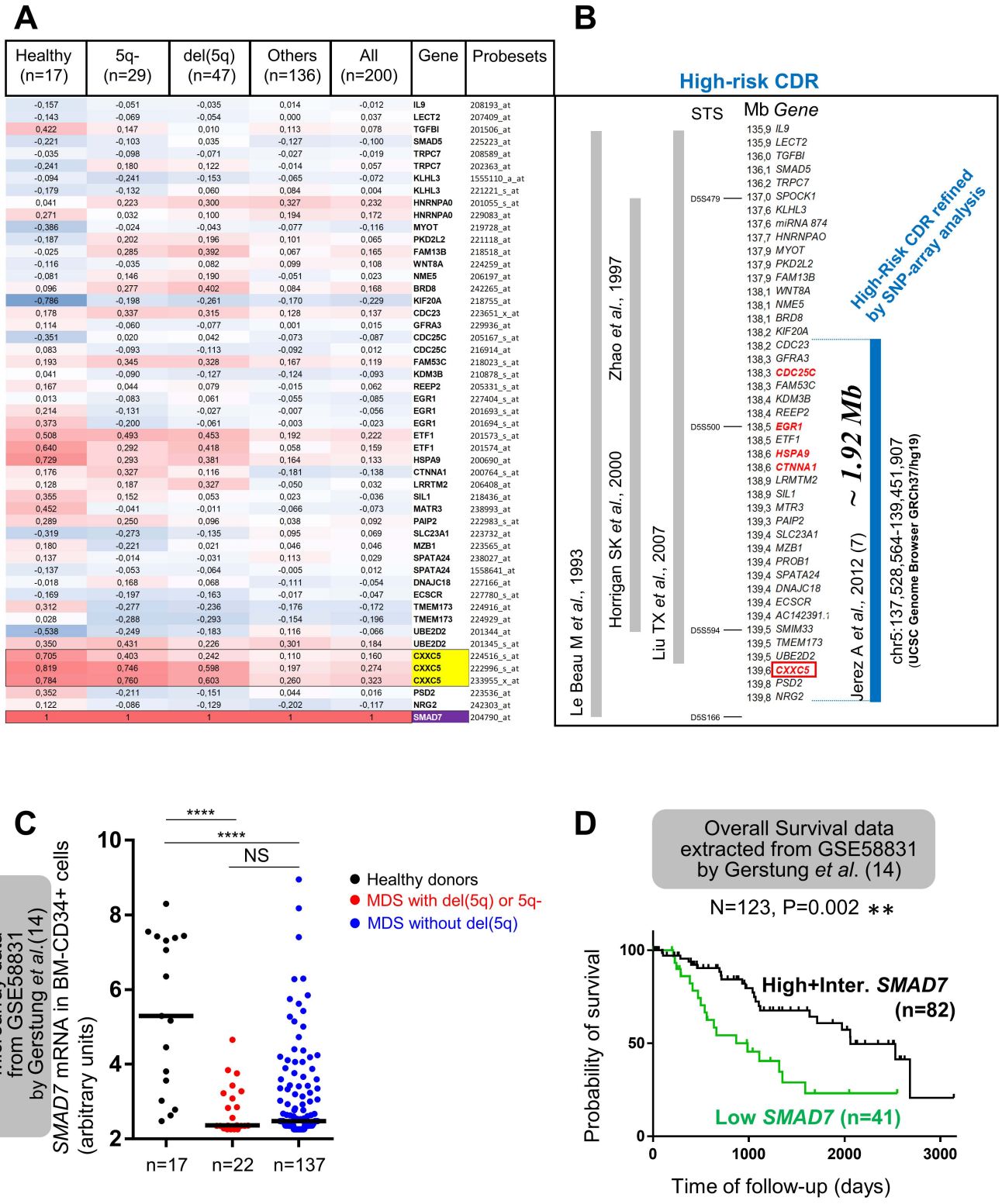


Figure S5
Astori and Matherat *et al*.



Healthy (n=17)	5q- (n=29)	del(5q) (n=47)	Others (n=136)	All (n=200)	Gene	Probesets
-0,157	-0,051	-0,035	0,014	-0,012	IL9	208193_at
-0,143	-0,069	-0,054	0,000	0,037	LECT2	207409_at
0,422	0,147	0,010	0,113	0,078	TGFBI	201506_at
-0,221	-0,103	0,035	-0,127	-0,100	SMAD5	225223_at
-0,035	-0,098	-0,071	-0,027	-0,019	TRPC7	208589_at
-0,241	0,180	0,122	-0,014	0,057	TRPC7	202363_at
-0,094	-0,241	-0,153	-0,065	-0,072	KLHL3	1555110_a_at
-0,179	-0,132	0,060	0,084	0,004	KLHL3	221221_s_at
0,041	0,223	0,300	0,327	0,232	HNRNPA0	201055_s_at
0,271	0,032	0,100	0,194	0,172	HNRNPA0	229083_at
-0,386	-0,024	-0,043	-0,077	-0,116	MYOT	219728_at
-0,187	0,202	0,196	0,101	0,065	PKD2L2	221118_at
-0,025	0,285	0,392	0,067	0,165	FAM13B	218518_at
-0,116	-0,035	0,082	0,099	0,108	WNT8A	224259_at
-0,081	0,146	0,190	-0,051	0,023	NME5	206197_at
0,096	0,277	0,402	0,084	0,168	BRD8	242265_at
-0,786	-0,198	-0,261	-0,170	-0,229	KIF20A	218755_at
0,178	0,337	0,315	0,128	0,137	CDC23	223651_x_at
0,114	-0,060	-0,077	0,001	0,015	GFRA3	229936_at
-0,351	0,020	0,042	-0,073	-0,087	CDC25C	205167_s_at
0,083	-0,093	-0,113	-0,092	0,012	CDC25C	216914_at
0,193	0,345	0,328	0,167	0,119	FAM53C	218023_s_at
0,041	-0,090	-0,127	-0,124	-0,093	KDM3B	210878_s_at
0,167	0,044	0,079	-0,015	0,062	REEP2	205331_s_at
0,013	-0,083	0,061	-0,055	-0,085	EGR1	227404_s_at
0,214	-0,131	-0,027	-0,007	-0,056	EGR1	201693_s_at
0,373	-0,200	-0,061	-0,003	-0,023	EGR1	201694_s_at
0,508	0,493	0,453	0,192	0,222	ETF1	201573_s_at
0,640	0,292	0,418	0,058	0,159	ETF1	201574_at
0,729	0,293	0,381	0,164	0,133	HSPA9	200690_at
0,176	0,327	0,116	-0,181	-0,138	CTNNA1	200764_s_at
0,128	0,187	0,327	-0,050	0,032	LRRTM2	206408_at
0,355	0,152	0,053	0,023	-0,036	SIL1	218436_at
0,452	-0,041	-0,011	-0,066	-0,073	MATR3	238993_at
0,289	0,250	0,096	0,038	0,092	PAIP2	222983_s_at
-0,319	-0,273	-0,135	0,095	0,069	SLC23A1	223732_at
0,180	-0,221	0,021	0,046	0,046	MZB1	223565_at
0,137	-0,014	-0,031	0,113	0,029	SPATA24	238027_at
-0,137	-0,053	-0,064	-0,005	0,012	SPATA24	1558641_at
-0,018	0,168	0,068	-0,111	-0,054	DNAJC18	227166_at
-0,169	-0,197	-0,163	-0,017	-0,047	ECSCR	227780_s_at
0,312	-0,277	-0,236	-0,176	-0,172	TMEM173	224916_at
0,028	-0,288	-0,293	-0,154	-0,196	TMEM173	224929_at
-0,538	-0,249	-0,183	0,116	-0,066	UBE2D2	201344_at
0,350	0,431	0,226	0,301	0,184	UBE2D2	201345_s_at
0,705	0,403	0,242	0,110	0,160	CXXC5	224516_s_at
0,819	0,746	0,598	0,197	0,274	CXXC5	222996_s_at
0,784	0,760	0,603	0,260	0,323	CXXC5	233955_x_at
0,352	-0,211	-0,151	0,044	0,016	PSD2	223536_at
0,122	-0,086	-0,129	-0,202	-0,117	NRG2	242303_at
1	1	1	1	1	SMAD7	204790_at

Figure S6
Astori and Matherat *et al.*

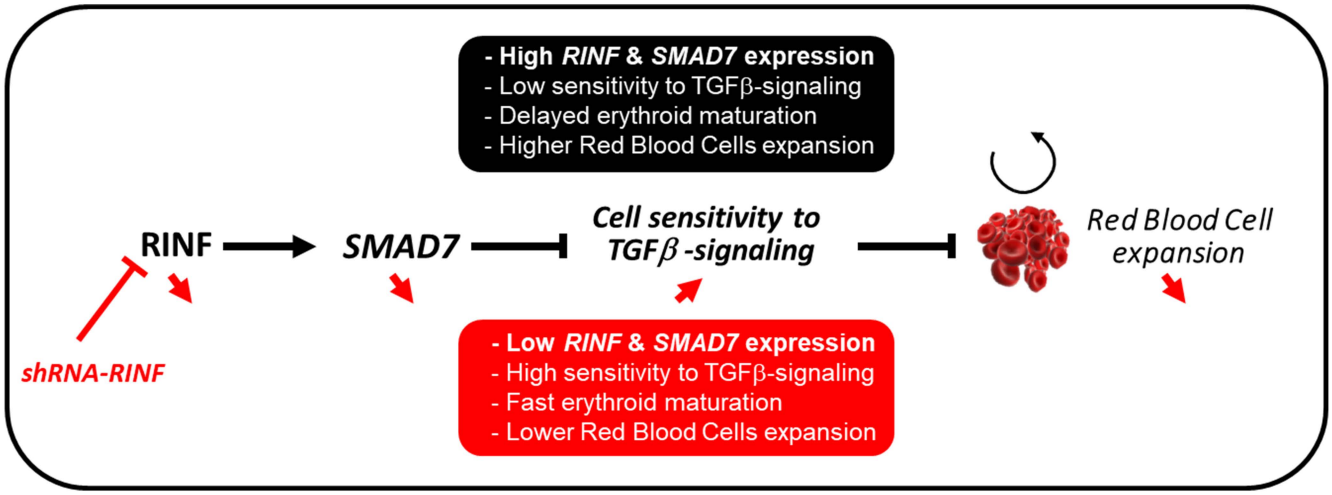


Figure S7
Astori and Matherat *et al*.