Myelodysplastic syndromes (MDS) encompass a heterogeneous group of haematologic disorders collectively defined by aberrant differentiation of myeloid precursors in the bone marrow[1,2]. Because of the ageing of our population, the incidence of the disease is increasing rapidly[3]. MDS is characterized by accumulation of abnormal myeloid precursors in the marrow, which is accompanied by peripheral blood cytopenias. MDS often progresses to acute myeloid leukemia (AML), with a poorer prognosis compared with *de novo* AML[4,5]. Somatic mutations in several crucial genes including *TET2*, *DNMT3A*, *RUNX1*, *ASXL1* and *EZH2* have been implicated as causal genetic alterations in MDS[6,7]. More recently, second-generation sequencing of MDS identified a high frequency of somatic mutations in the genes encoding for the RNA splicing machinery[8]. Recurrent mutations were detected by us and others in *SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2* and other spliceosome genes in independent cohorts of MDS, signifying a novel mechanism regulating the pathogenesis of this disease[9–14]. However, the functional consequence of these somatic mutations in the pathobiology of MDS remains largely unidentified.

RNA splicing is a fundamental process in eukaryotes, which excises the intronic sequences from mRNA precursors to generate functional mRNA species. This function is carried out by the splicing machinery, which comprises RNA–protein complexes called small nuclear ribonucleoprotein particles (snRNP). The major splicing machinery (termed U2 spliceosome) involves 5 snRNPs (U1, U2, U4, U5 and U6), which function in concert with numerous other proteins to effect splicing of introns[15]. In addition, a second class of introns processed by a divergent spliceosome called minor (or U12) spliceosome was later identified[16,17]. The U12 machinery consists of U11, U12, U4atac, U6atac and U5 snRNPs, and recognizes distinct intronic splice sites[18–20]. The U12-type introns coexist with U2-type introns in several genes involved in essential cellular functions such as DNA replication, RNA processing, DNA repair and translation[21].

ZRSR2 (also known as *URP*) is located on X chromosome (Xp22.1) and encodes for a splice factor involved in recognition of 3′-intron splice sites. It interacts with other components of the pre-spliceosome assembly including the U2AF2/U2AF1 heterodimer and SRSF2 (ref. 22). *In vitro* splicing assays suggest that ZRSR2 is required for efficient splicing of both the major and the minor class of introns[23]. In MDS, somatic mutations in *ZRSR2* occur across the entire length of the transcript, which is in contrast to mutational hotspots observed in *SF3B1*, *SRSF2* and *U2AF1*. Moreover, nonsense, splice-site and frame-shift mutations in *ZRSR2* gene frequently occur in males, suggesting a loss of function. Mutations in *ZRSR2* are more prevalent in MDS subtypes without ring sideroblasts and chronic myelomonocytic leukemia, and are associated with elevated percentage of bone marrow blasts and higher rate of progression to AML[8,13]. However, the mechanism linking ZRSR2 deficiency to pathogenesis of MDS has not been explored.

In this study, we have evaluated the cellular and functional consequences of the loss of ZRSR2 in cell lines and patient samples. We show that ZRSR2 plays a pivotal role in splicing of the U12-type introns, while the U2-dependent splicing is largely unaffected. MDS bone marrow harbouring inactivating mutations in *ZRSR2* exhibit overt splicing defects, primarily involving the aberrant retention of U12-type introns. Short hairpin RNA (shRNA)-mediated knockdown of ZRSR2 similarly leads to impaired splicing of U12-type introns. Knockdown of ZRSR2 also inhibits cell growth and alters the *in vitro* differentiation potential of haematopoietic cells. This study uncovers a specific function of ZRSR2 in RNA splicing and also suggests its role in haematopoietic development.

Results

Knockdown of ZRSR2 leads to specific splicing defects. In MDS, somatic mutations in *ZRSR2* are often inactivating alterations (nonsense, frame-shift and splice site mutations), which primarily affect the males, signifying its loss of function in these cases. To replicate the loss of ZRSR2, a lentiviral shRNA approach was used to stably downregulate its expression in human cells. Two shRNA vectors targeting ZRSR2 (ZRSR2 sh1 and sh2) were used to generate stable knockdown cells. These vectors resulted in efficient downregulation of *ZRSR2* transcript and protein levels in 293T cells and leukemia cell lines, TF-1 and K562 (Fig. 1a,b and Supplementary Fig. 1).

First, we examined the effect of ZRSR2 deficiency on splicing, by transfection of minigene constructs in ZRSR2 knockdown and control-transduced 293T cells. Two reporter constructs commonly used to assess splicing—*P120* minigene[24] and *GH1* reporter plasmid[25]—were used in these experiments. *P120* minigene reporter consists of exons 5–8 of human *NOP2* (also known as *NOL1* or *P120*) gene. We observed that the splicing of intron F, a U12-type intron, was reduced on downregulation of ZRSR2 using both ZRSR2 shRNA vectors (Fig. 1c,d). The *GH1* minigene reporter consists of three exons, and on transfection a fully spliced and exon-skipped (missing the second exon) mRNA can be detected. Notably, in a previous study, ectopic expression of mutant U2AF1—a splice factor related to ZRSR2 and also frequently mutated in MDS—results in higher frequency of exon skipping from *GH1* minigene construct[12]. We observed that ZRSR2 knockdown and control 293T cells exhibited comparable rates of exon skipping on transfection of *GH1* reporter construct (Supplementary Fig. 2). Therefore, our results highlight that ZRSR2 functions in a manner distinctive of U2AF1.

Next, we assessed splicing of endogenous introns in the MDS/AML cell line, TF-1 (ref. 26), transduced with either ZRSR2 shRNA or control vector. ZRSR2 has been proposed to be involved in splicing of both major and minor classes of introns[23]; therefore, splicing of both these types of introns was examined. All tested U12-dependent introns were less efficiently spliced in the ZRSR2 knockdown cells (Fig. 1e,f). The average ratio of spliced/unspliced RNA for the ten U12-type introns was 0.30 in sh1-transduced and 0.38 in sh2-transduced cells as compared with 1.0 in the control cells. Notably, the splicing of six U2-type introns was not significantly affected (spliced/unspliced ratio of 1.19 for sh1 and 0.96 for sh2 as compared with 1.0 for cells transduced with control shRNA; Fig. 1e,f). Hence, the inability of ZRSR2 knockdown cells to splice efficiently endogenous and exogenous U12-type introns points towards a specific defect in the minor splicing machinery.

To test whether overexpression of ZRSR2 can rescue the U12 splicing defects, we transiently transfected wild-type (WT) ZRSR2 into stably knock down 293T cells (Supplementary Fig. 3a) and measured the splicing efficiency of U12-type introns. Ectopic expression of ZRSR2 resulted in a significant increase in the splicing efficiency of all U12-type introns tested as compared with cells transfected with empty vector (Fig. 1g and Supplementary Fig. 3b). Therefore, we conclude that aberrant splicing observed in knockdown cells was a consequence of downregulation of ZRSR2. Overall, our experiments to evaluate splicing of endogenous and exogenous introns in ZRSR2 knockdown cells recognize its role in the U12-dependent spliceosome.

Inactivating *ZRSR2* mutations in MDS cause splicing defects. To address the consequences of *ZRSR2* mutations in MDS, a global evaluation of splicing alterations was performed using RNA sequencing (RNA-Seq). RNA was extracted from the bone marrow of eight male MDS patients harbouring either nonsense or frame-

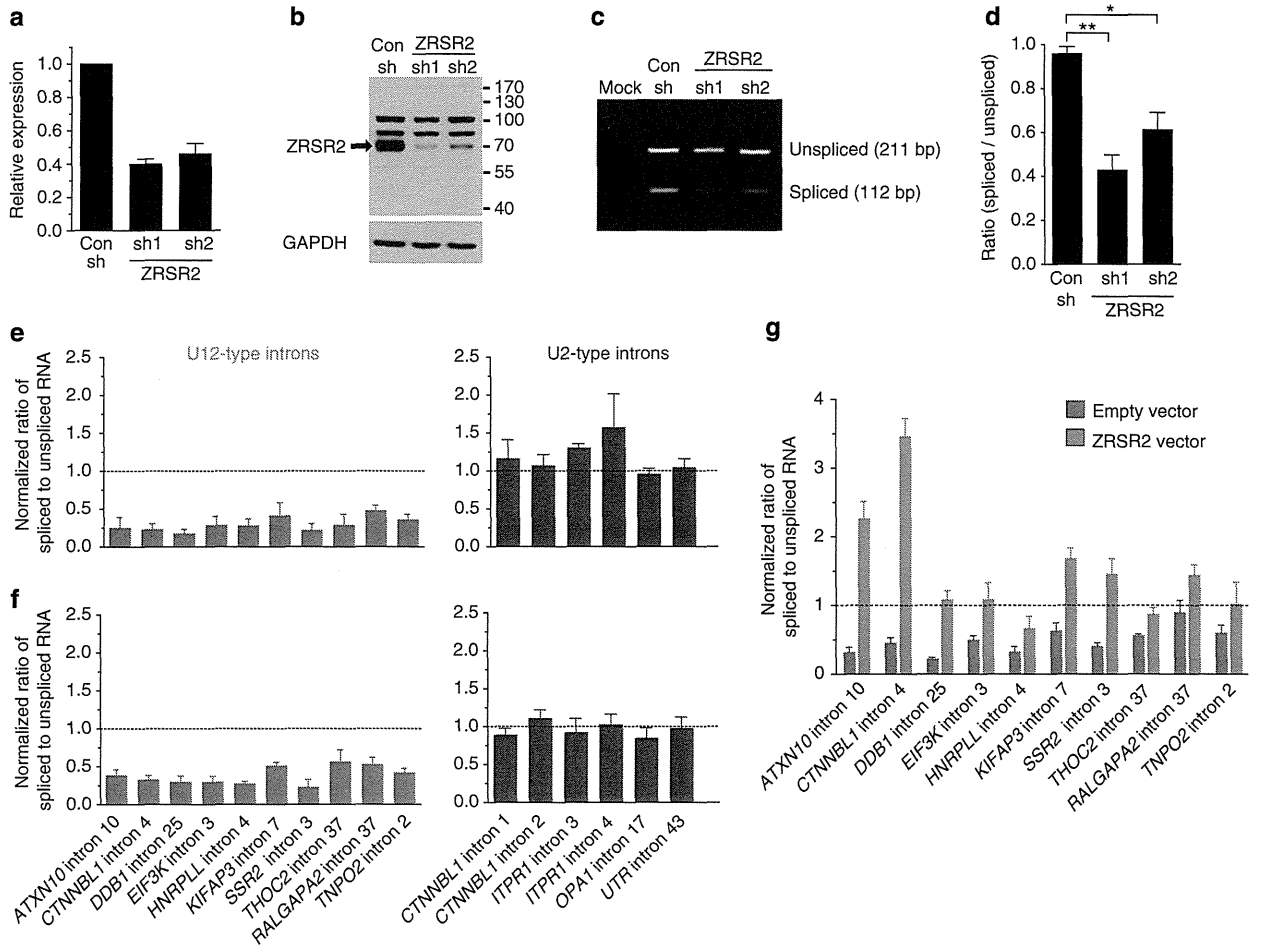

Figure 1 | Knockdown of ZRSR2 induces defects in splicing of U12-type introns. (**a**) Transcript levels of *ZRSR2* in TF-1 cells stably transduced with either lentiviral ZRSR2 shRNA or control (Con) vectors were examined using qRT-PCR. *GAPDH* levels served as endogenous control. (**b**) Western blot analysis to verify the decrease in ZRSR2 protein levels in knockdown TF-1 cells. (**c**) Splicing efficiency of intron F of *P120* minigene construct was measured in ZRSR2 knockdown and control 293T cells. A representative gel picture shows bands corresponding to unspliced and spliced product in RT-PCR analysis performed 48 h after transfection of minigene plasmid. (**d**) The bars depict ratio of intensities of PCR bands corresponding to spliced and unspliced products in *P120* minigene assay. The data represent the mean ± s.e.m. of three independent transfection experiments. $^{*}P<0.05$, $^{**}P<0.01$; unpaired *t*-test. (**e**,**f**) Average ratio of spliced to unspliced pre-mRNA levels of ten U12-type and six U2-type introns in TF-1 cells on knockdown of ZRSR2 using two shRNA vectors, ZRSR2 sh1 (**e**) and ZRSR2 sh2 (**f**). The data are mean ± s.e.m. from at least three independent RNA preparations. Horizontal dotted lines represent the ratio for control transduced cells, which were set as 1.0. *GAPDH* was used as endogenous control. (**g**) Average ratios of spliced to unspliced levels of U12-type introns on transient transfection of ZRSR2 expression plasmid in knockdown 293T cells are depicted. 293T cells stably expressing ZRSR2 sh1 or control vector were transfected with either pCDNA3-hZRSR2 or empty vector, and total RNA was extracted after 72 h. The splicing efficiency was measured using qPCR and spliced/unspliced ratio was set as 1.0 for control cells transfected with empty vector (horizontal dotted line). *GAPDH* was used to normalize for cDNA input. The results are average of five to seven transfection experiments and are represented as mean ± s.e.m.

shift mutations of *ZRSR2* (Fig. 2a) (hereafter referred to as '*ZRSR2* mutant MDS'). We also sequenced RNA from four MDS cases without mutation in either *ZRSR2* or other commonly mutated splice factors (*U2AF1*, *SF3B1* and *SRSF2*; termed '*ZRSR2* WT MDS'). In addition, three non-malignant bone marrows and one remission bone marrow (remission of sample 7; *ZRSR2* mutant MDS; Table 1) were also included as controls (termed 'normal bone marrow (BM)'). RNA-Seq verified the presence of mutations in *ZRSR2* with a high mutant allele frequency (range: 67.9–98.0%) in all the *ZRSR2* mutant MDS samples (Supplementary Fig. 4).

First, the RNA-Seq data were examined for abnormal sequencing reads at all splice junctions to assess the extent of mis-spliced events in different groups. The 'normal' reads comprised those which aligned to the known exon–exon junctions, while 'abnormal' reads either spanned exon–intron junctions or corresponded to splicing involving an ambiguous splice site as illustrated in Fig. 2b. We examined 298,275 unique splice junctions (23,786 RefSeq transcripts in 15,737 genes) for aberrant splicing and each *ZRSR2* mutant MDS sample was compared with a control sample (*ZRSR2* WT MDS or normal BM) as described in the Methods section. Using a false discovery rate (FDR) cutoff of 0.01 and difference in Mis-splicing Index (ΔMSI) >20, significantly higher number of abnormally spliced junctions were detected in *ZRSR2* mutant MDS samples compared with the *ZRSR2* WT MDS samples in a majority of pairwise comparisons (Fig. 2c). Similarly, the number of such abnormal junctions was also elevated in *ZRSR2* mutant group when compared with four normal BM samples (Fig. 2d). These findings suggested a higher incidence of aberrant splicing in samples with *ZRSR2* mutation as compared with controls.

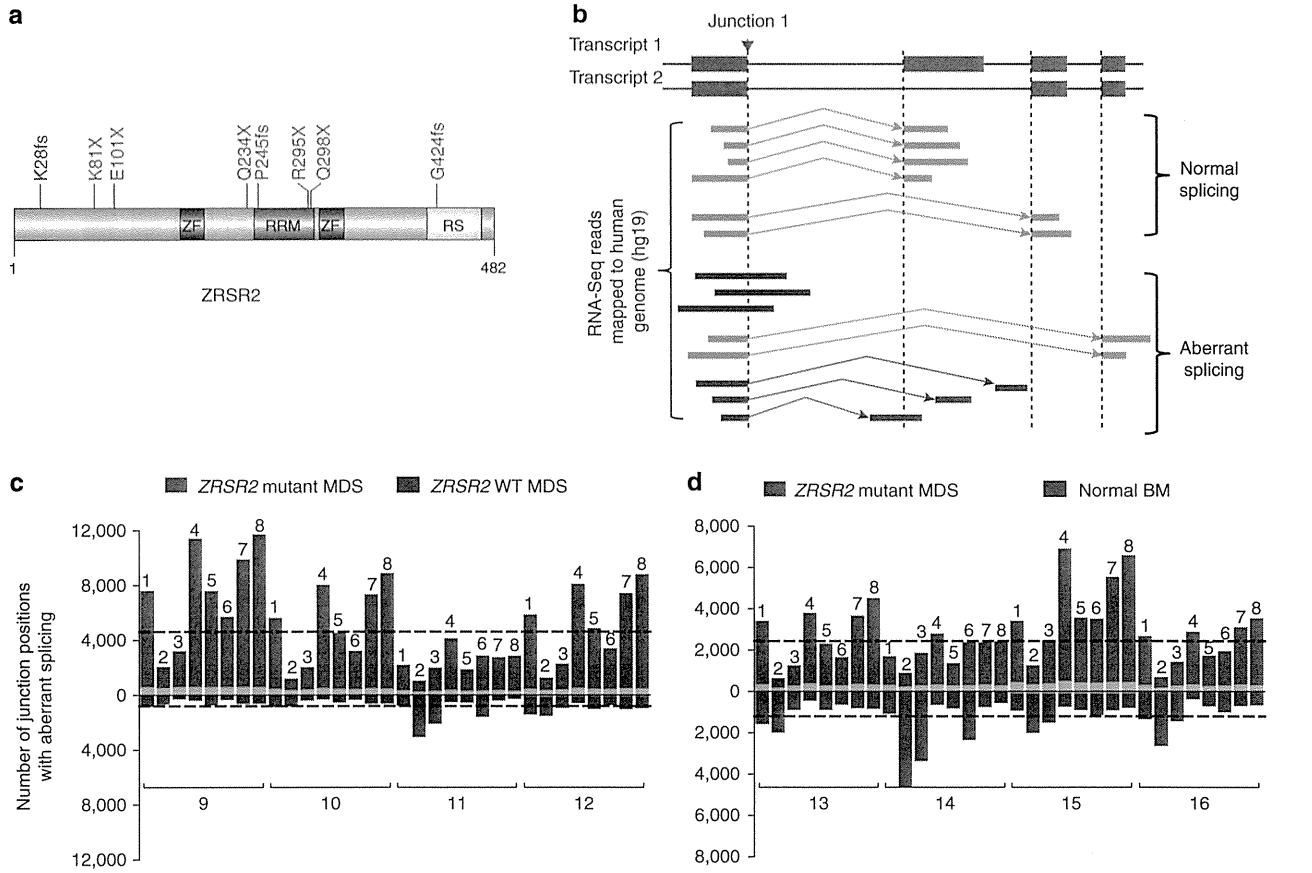

Figure 2 | RNA-Seq of MDS bone marrow harbouring mutations in *ZRSR2* reveals splicing defects. (**a**) Schematic representation of human ZRSR2 protein with position and type of mutations in eight MDS patients used for RNA-Seq. (**b**) Approach used to define the RNA-Seq reads for analysis of splice junctions is shown. All splice junctions corresponding to known RefSeq transcripts were examined. Reads mapped to a representative splice position (Junction 1) are classified as either 'normal' or 'aberrant' in the illustration. (**c**) Number of junction positions with aberrant reads obtained in pairwise comparisons between *ZRSR2* mutant and WT MDS are depicted. Junctions with significant aberrant reads in *ZRSR2* mutant (ΔMSI >20) are shown as red bars, while those in *ZRSR2* WT (ΔMSI < -20) are shown as blue bars for each mutant versus WT pair. The dashed horizontal lines denote the averages of aberrant junction positions in the two genotypes across 32 comparison pairs (8 *ZRSR2* mutant MDS compared individually with 4 *ZRSR2* WT MDS). (**d**) Number of aberrant junctions obtained in pairwise comparisons between *ZRSR2* mutant MDS and normal BM analysed and depicted as described in **c**. The black bars represent number of aberrant junctions detected in normal BM. The green bars in **c** and **d** represent the number of junctions that were identified in all eight mutant/eight control samples in at least one pairwise comparison. Each pair of bars in **c** and **d** is labelled with identifiers for *ZRSR2* mutant and control samples represented in their pair-wise comparison. The sample information including details of *ZRSR2* mutations are described in Table 1.

Notably, 689 mis-spliced junctions were identified in all eight *ZRSR2* mutant MDS cases, signifying a subset of introns that represent *bona-fide* downstream targets of ZRSR2.

Aberrant intron retention in *ZRSR2* mutant MDS. To delineate the splicing defects that occur in *ZRSR2*-mutated cells, we carefully evaluated the RNA-Seq data of mutant and control bone marrows for aberrant intron retention, cryptic splice site usage and exon skipping. For identification of aberrant retention, introns with $\geq$95% coverage and with sequencing reads supporting both 5′ and 3′ exon–intron junctions were considered (described in Methods section). We compared the proportion of aberrant reads (spanning across both exon–intron junctions) in each sample to calculate the MSI (Supplementary Fig. 5). The difference in the MSI values between the mutant and control samples (termed ΔMSI) was used as a measure of aberrant retention for each intron. Using this approach, we tested 110,192 introns and performed pairwise analyses between 8 *ZRSR2* mutant MDS samples and 8 controls (4 *ZRSR2* WT MDS and 4 normal BM samples) to obtain 64 sets of comparisons. We observed that elevated number of retained introns was clearly evident in *ZRSR2* mutant MDS as compared with either *ZRSR2* WT MDS or normal BM samples (Fig. 3a,b). Importantly, significant intron retention was not detected when comparing the *ZRSR2* WT MDS with normal BM samples (Fig. 3c), underlining that the intron retention is specific to mutations in the *ZRSR2* gene.

We further examined the retained introns in *ZRSR2*-mutated cases for the type of intron. The introns were categorized as either U2- or U12-type based on the divergence at the 5′- and 3′-splice sites and the branchpoint sequence[20,21], using the computational method described previously[27]. This analysis revealed a striking overabundance of U12-type introns among the aberrantly retained introns in *ZRSR2* mutant MDS samples (Fig. 3a,b,d,e). This pattern was observed consistently across all the pairwise comparisons between *ZRSR2* mutant and control samples (Fig. 3d,e), while the comparisons between *ZRSR2* WT MDS and normal BM (16 pairwise comparisons) did not show any intron type-specific retention (Fig. 3f). Next, to ascertain the

Table 1 | Details of human bone marrow samples used for RNA-Seq.

Sample	MDS/normal	Age (years)	Gender	*ZRSR2* mutation*		WHO subtype	IPSS†	Cytogenetics
				Nucleotide	Protein			
1	MDS	59	Male	C892T	Gln298X	RAEB-1	1	46,XY
2	MDS	81	Male	G301T	Glu101X	RA	1	45,X, − Y
3	MDS	70	Male	A241T	Lys81X	RCMD	0	46,XY
4	MDS	74	Male	81delA	Lys28fsX11	RCMD	0	46,XY
5	MDS	72	Male	C883T	Arg295X	RCMD	0	46,XY
6	MDS	74	Male	C700T	Gln234X	RAEB-2	2	46,XY
7	MDS	62	Male	1271delG	Gly424fsX93	RAEB-2	2	46,XY
8	MDS	72	Male	735delC	Pro245fsX9	RCMD	0	46,XY
9	MDS	50	Male	—	—	RA	1	46,XY,t(2;11)(p21;q23)
10	MDS	76	Male	—	—	RAEB-2	3	45,XY, − 7
11	MDS	65	Female	—	—	RAEB-1	1	46,XX
12	MDS	71	Male	—	—	RA	0	45,X, − Y
13	Remission state of 7	62	Male	—	—	—	—	NA‡
14	Normal BM	51	Male	—	—	—	—	46,XY
15	Normal BM	38	Male	—	—	—	—	46,XY
16	Normal BM	75	Male	—	—	—	—	46,XY

BM, bone marrow; IPSS, International Prognostic Scoring System; MDS, myelodysplastic syndrome; RNA-Seq, RNA sequencing; WHO, World Health Organization.
*None of the MDS samples harbour mutation in other frequently mutated spliceosome genes: *SF3B1*, *SRSF2* and *U2AF1*.
†IPSS (0 = low; 1 = int-1; 2 = int-2; 3 = high).
‡Not available.

subset of introns, which were consistently retained in *ZRSR2* mutant samples (ΔMSI >20), we focused on introns recognized in a large number of pairwise comparisons. Expectedly, the number of aberrantly retained introns identified in successively increasing numbers of pairwise comparisons gradually decreased (Supplementary Fig. 6); however, the proportion of U12-type introns among the retained introns steadily climbed (Fig. 3g). In fact, 43 out of 45 introns retained in the *ZRSR2* mutant MDS in all 64 comparisons were U12 dependent. On the other hand, specific intron retention in the *ZRSR2* WT and normal BM groups (ΔMSI < -20) was not apparent and no intron was retained in more than 41 pairwise comparisons (Fig. 3g). Consequently, among the high-ranking set of introns consistently retained (present in >41 pairwise comparisons) in the *ZRSR2* mutant MDS, a disproportionately higher prevalence (85%) of U12-type introns occurred (Fig. 3h and Supplementary Data 1), thereby underscoring the involvement of ZRSR2 in the minor spliceosome machinery. Moreover, among the retained U2-type introns, 72% were contained in a transcript also harbouring a U12-type intron, with the retained U2-type intron typically located immediately downstream of the U12-type intron (Fig. 3h and Supplementary Fig. 7). This indicates that the inefficient splicing of U12-type introns can also cause mis-splicing of neighbouring U2-type introns within the transcript. Only 11 U2-type introns, which were independent of U12 transcripts, were identified as aberrantly retained in our computational approach (Supplementary Data 1). These introns were indistinguishable from other unaffected U2-type introns (data not shown) and invariably displayed a weaker retention phenotype compared with the U12-type introns.

Although our preceding analysis identified that the U12-type introns were significantly retained in *ZRSR2* mutant MDS, we further inquired whether the splicing of all U12-type was affected. To address this, we examined the aberrant retention of each U12-type intron using the average ΔMSI values for mutant versus control groups. Of all the genes harbouring U12-type introns in the human genome[27,28], genes containing 558 U12 introns were expressed at sufficient levels ($FPKM_{max} > 1$; $FPKM_{average} > 0.5$). First, ΔMSI calculations showed that practically all U12-type introns were mis-spliced, albeit to varying extent, in the *ZRSR2* mutant MDS group (Supplementary Fig. 8). This is also evident in individual pairwise comparisons between *ZRSR2* mutant MDS and control samples (Fig. 3d,e). Next, we classified the U12-type introns based on the ΔMSI values and found that the splicing of only 29 introns (ΔMSI ≤ 0) (5% of expressed U12-type introns) was unaffected (Supplementary Data 2). Although, we did not detect any difference in the distribution of GT-AG versus AT-AC intron type with the retention phenotype, the U12-type introns, which are not retained in *ZRSR2* mutant cells, tend to be longer (median length 1,958 nucleotides compared with 1,039 for significantly retained introns; Supplementary Data 2). Interestingly, the unaffected GT-AG introns had relatively weaker splice sites as indicated by lower 5′-splice site score compared with the retained introns (Supplementary Data 2).

We also sequenced mRNA from TF-1 cells in which ZRSR2 was downregulated, using either sh1 or sh2 lentiviral vectors. Stable knockdown cells from two independent transduction experiments were used to examine for intron retention as described above. Expression analysis confirmed 60–70% reduction in *ZRSR2* transcript levels in the knockdown cells (Fig. 3i). The suppression of ZRSR2 in these knockdown cells resulted in marked retention of U12-type introns as compared with control shRNA-transduced cells (Fig. 3j,k). Overall, the number of retained introns obtained in knockdown TF-1 cells was lower than those identified in *ZRSR2* mutant MDS cases. This is conceivable because of low levels of ZRSR2 present in the knockdown cells as compared with complete absence in males with either nonsense or frame-shift mutations. Importantly, among the introns retained in both sh1 and sh2 knockdown cells, a large proportion were U12-type introns (Fig. 3l). Notably, we observed a sizeable overlap of U12-type introns retained in both *ZRSR2* mutant MDS and knockdown TF1 cells (Supplementary Fig. 9).

Next, to validate the aberrant intron retention, quantitative reverse transcriptase–PCR (qRT–PCR) was used to measure the normalized intronic expression in *ZRSR2* mutant MDS and control samples. Using this approach, we tested eight representative U12-type introns detected as aberrantly retained in our computational analysis. We observed markedly higher expression of all tested introns in each of the mutant samples as compared

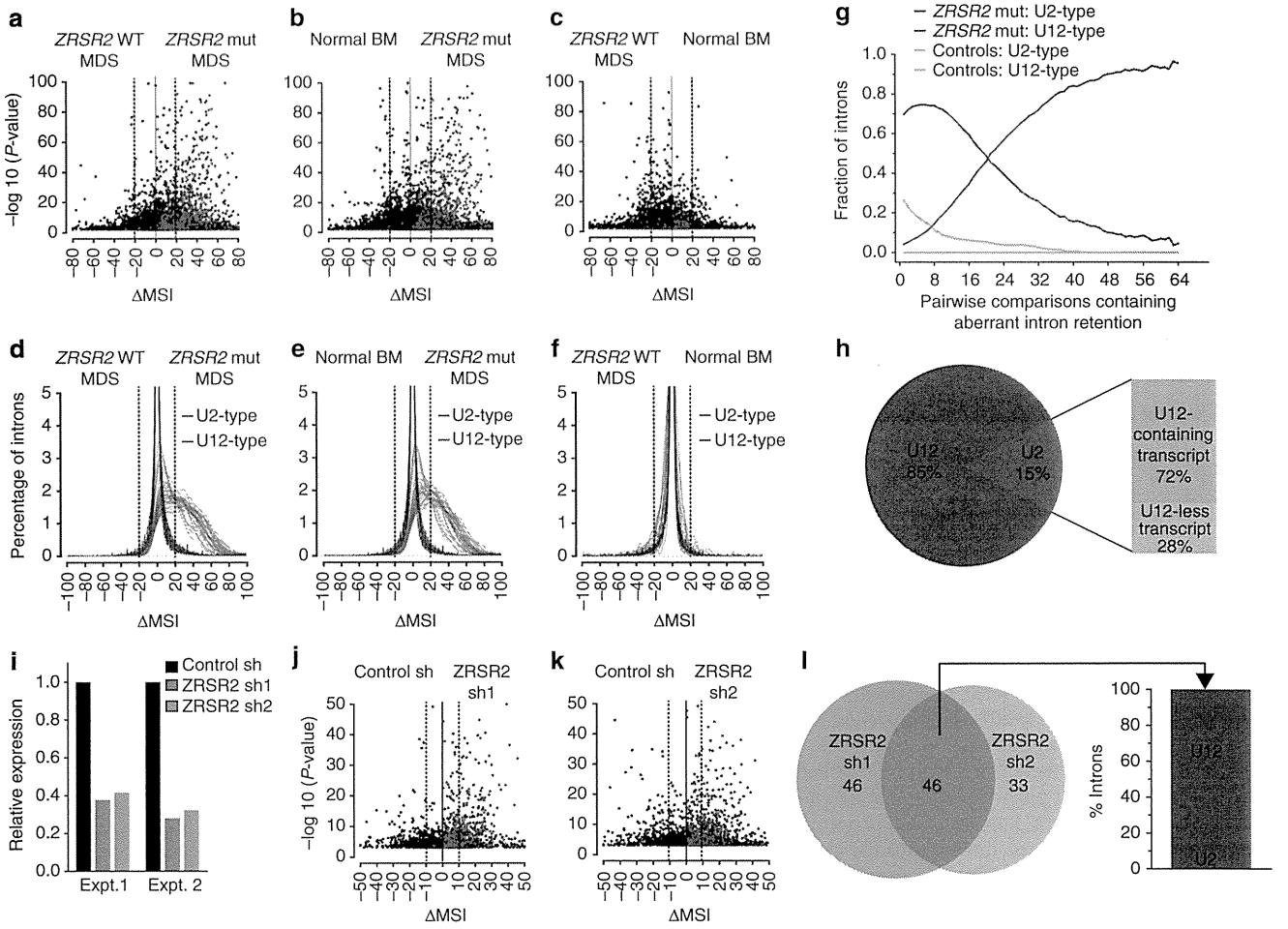

Figure 3 | *ZRSR2*-mutated MDS bone marrow and ZRSR2 knockdown cells are characterized by aberrant retention of U12-type introns. (**a**–**c**) Dot plots display aberrantly retained introns in a representative pairwise analysis of *ZRSR2* mutant MDS versus *ZRSR2* WT MDS (**a**), *ZRSR2* mutant MDS versus normal BM (**b**), and normal BM versus *ZRSR2* WT MDS (**c**). Each dot denotes an intron and U12-type introns are shown in red. *P*-value was calculated using Fisher's exact test and data points with $P<0.01$ are shown. (**d**–**f**) Histograms depict frequencies of U2-type and U12-type introns plotted against ΔMSI values in pairwise comparisons of *ZRSR2* mutant and control samples. Each curve represents a pairwise comparison for U2-type (blue) and U12-type (red) intron. Thirty-two comparisons between *ZRSR2* mutant MDS and *ZRSR2* WT MDS (**d**), 32 comparisons between *ZRSR2* mutant MDS and normal BM (**e**), and 16 comparisons between *ZRSR2* WT MDS and normal BM (**f**) were performed. (**g**) The proportion of U2-type and U12-type introns among aberrantly retained introns in either *ZRSR2* mutant or controls (*ZRSR2* WT MDS + normal BM) are shown for 64 pairwise comparisons. The number of introns is plotted against the number of pairwise comparisons in which they were identified. (**h**) Distribution of intron type for significantly retained introns (ΔMSI >20; FDR ≤0.01) in *ZRSR2* mutant MDS is shown. The bar graph shows the distribution of retained U2-type introns into those present in either transcript containing a U12-type intron or without U12-type intron. (**i**) Relative expression of *ZRSR2* was computed as fragments per kilobase of transcript per million fragments mapped (FPKM) values from RNA-Seq data in two independently transduced control and knockdown TF-1 cells. (**j**,**k**) Dot plots depict aberrantly retained introns in control versus ZRSR2 sh1 (**j**) and control versus ZRSR2 sh2 (**k**). Only data points corresponding to $P<0.01$ are displayed. (**l**) Venn diagram shows an overlap of retained introns between ZRSR2 sh1 and sh2 knockdown TF-1 cells. The introns that are significantly retained ($P<0.01$; ΔMSI >10) for sh1- or sh2-transduced cells in both experiments are included. The bar graph on the right depicts the proportion of U2-type and U12-type introns among the introns retained in both sh1- and sh2-transduced cells.

with the control samples (Fig. 4a–h). In a parallel analysis, we also detected higher expression of these introns in ZRSR2 knockdown TF1 cells (Supplementary Fig. 10), signifying consistent findings in our two experimental models.

The experimental evidence of U12-type intron retention specifically in ZRSR2-deficient MDS samples further substantiates ZRSR2 as a crucial component of the U12-dependent spliceosome.

Additional mis-splicing events in *ZRSR2* mutant MDS. Further, using an approach similar to the one to detect intron retention, we searched for mis-splicing events involving abnormal recognition of 5′- and 3′-splice sites (Supplementary Fig. 5). We identified several loci where cryptic splice sites were observed in *ZRSR2* mutant MDS. We focused on loci, which displayed aberrant splicing in all mutant cases, and found that the majority of them were associated with transcripts containing U12-type introns. These events usually occurred either within the U12-type introns or in their vicinity. The incorrect recognition of splice sites resulted in a varied pattern of mis-splicing involving ambiguous splice donor and acceptor sites, which invariably generated cryptic U2 splice junctions (Fig. 5). As representative examples of abnormal splice site recognition in *ZRSR2* mutant

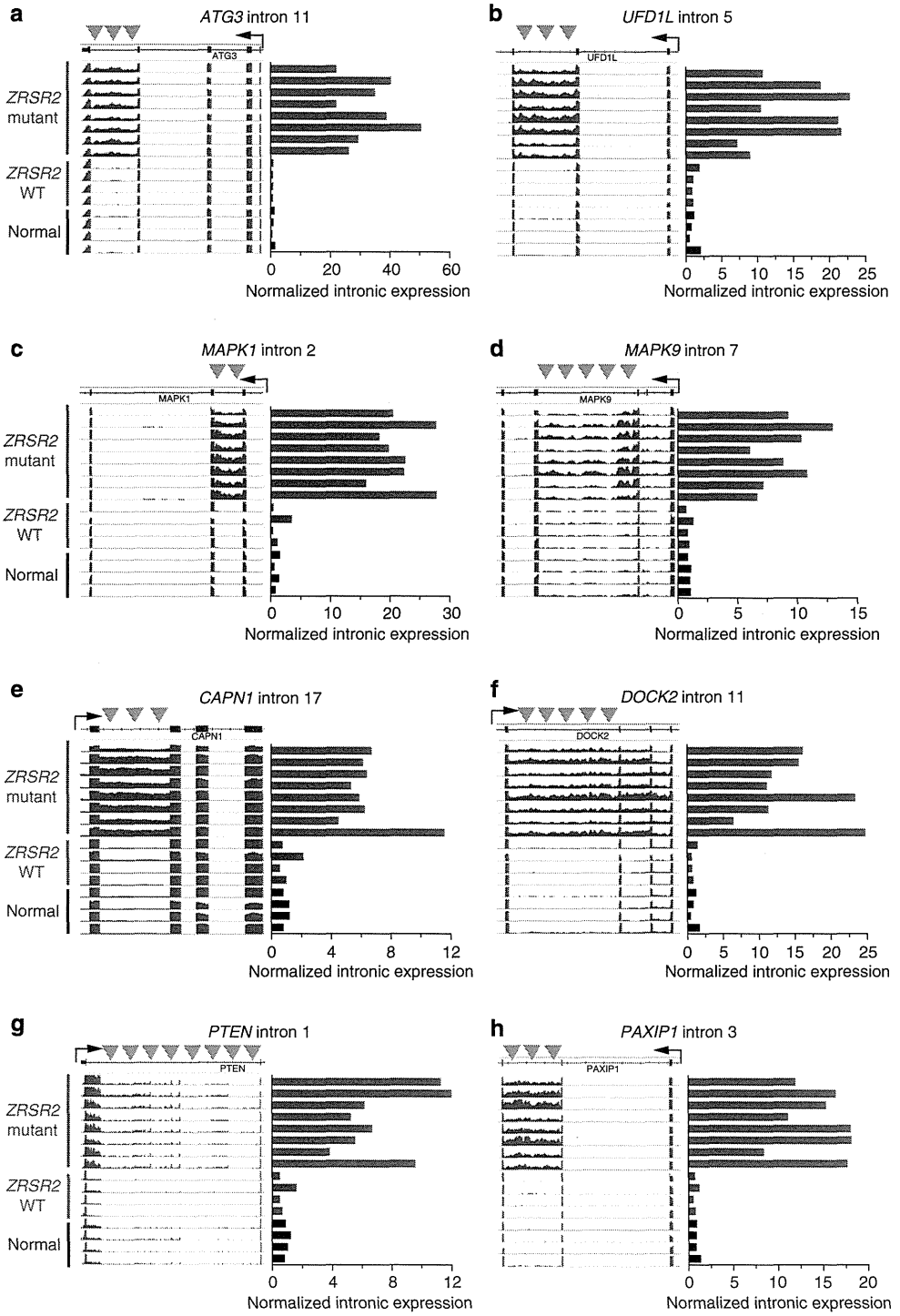

Figure 4 | Detection of U12-type intron retention in *ZRSR2* mutant MDS. (**a–h**) Normalized intron expression for eight representative U12-type introns was determined using qPCR. The RNA-Seq reads normalized to total number of mappable reads for all 16 cases are depicted using IGV 2.3 in the left panels. Read counts are shown using an identical scale in all samples, and the U12-type introns are indicated by orange arrow heads. For each gene, only the genomic locus containing the retained intron is shown. Right panels: the expression of U12-type introns was measured relative to the expression of flanking exons and is shown by horizontal bars (red bars, *ZRSR2* mutant MDS; blue bars, *ZRSR2* WT MDS; black bars, normal BM).

MDS, mis-spliced U12-type introns in *WDR41*, *FRA10AC1* and *SRPK2* genes were experimentally validated.

The intron 4 of human *WDR41* gene is a U12-type intron. RNA-Seq revealed a distinctive pattern of mis-splicing in all *ZRSR2* mutant samples, which includes retention of the 5′-portion of the intron followed by multiple mis-splicing events employing cryptic U2-type splice sites within intron 4 (Fig. 5a,b and Supplementary Data 3 and 4). We verified the presence of such aberrant splice junctions across the exons 4 and 5 involving two cryptic exons ('4A' and '4B') using qRT–PCR. The levels of mis-spliced products

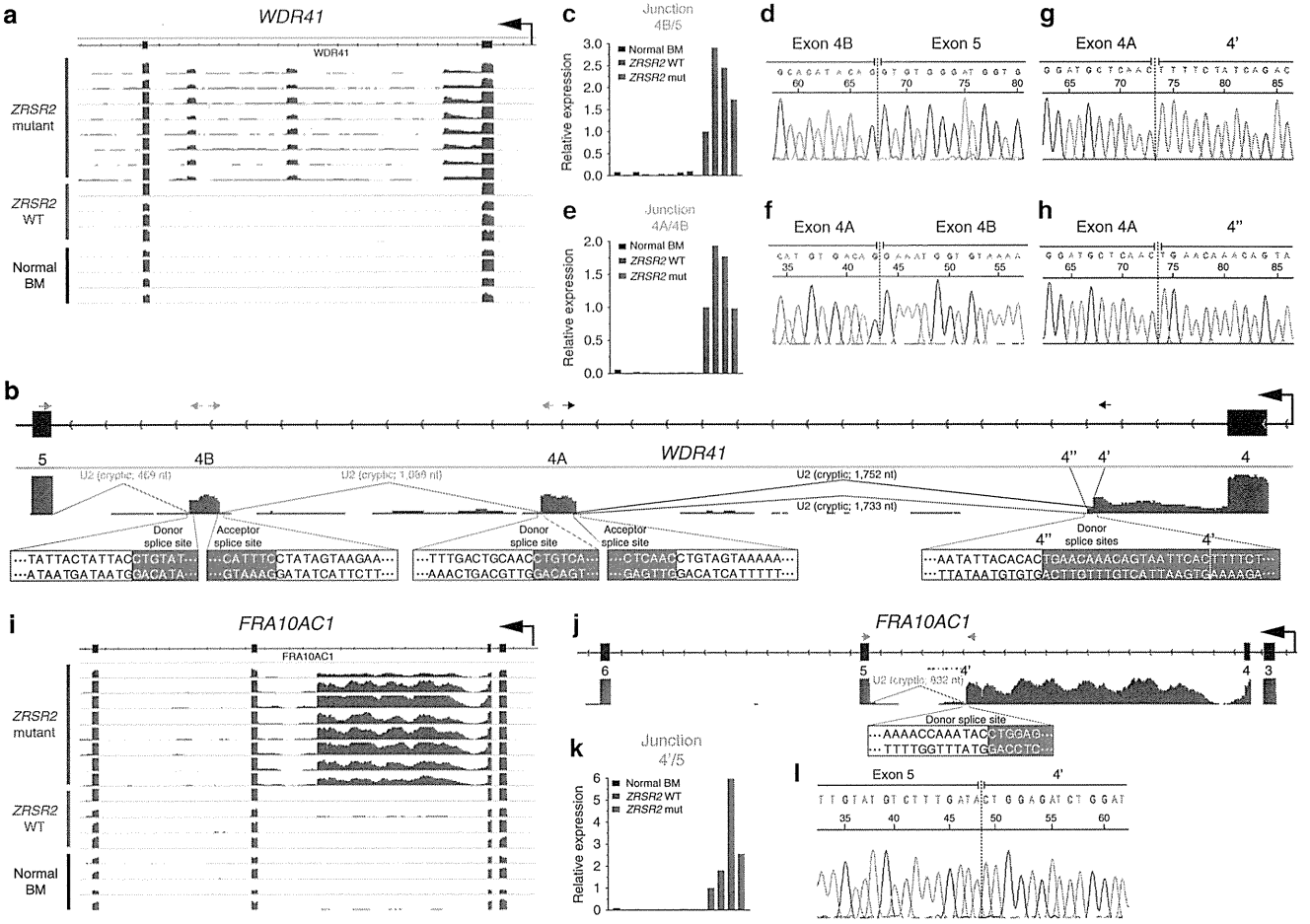

Figure 5 | Cryptic splice junctions in U12-type introns of *WDR41* and *FRA10AC1*. (**a**) Normalized RNA-Seq reads mapped to the genomic region encompassing exons 4 and 5 of *WDR41* gene are displayed using IGV 2.3 for all 16 samples. Reads in all samples are shown on identical scale. (**b**) Aberrant splice junctions in intron 4 (U12-type intron) of *WDR41* gene are depicted. The mis-spliced regions (designated 4A and 4B) and two alternative splice donor sites in intron 4 (upstream of 4A; marked as 4′ and 4″) in *ZRSR2* mutant samples are illustrated. The intron type and length for the cryptic junctions are indicated. The cryptic splice acceptor and donor sequences, which are activated in *ZRSR2* mutant MDS, are shown below. The PCR primers used in **c** and **e** are indicated by arrows. (**c**,**d**) Experimental verification of splicing junction between 4B and exon 5 of *WDR41*. (**c**) qRT-PCR using primers located in 4B and exon 5 to determine relative levels of mis-spliced transcript in *ZRSR2* mutant and control samples. (**d**) The PCR product amplified from cDNA of *ZRSR2* mutant samples in **c** was Sanger sequenced. The junction is shown by dashed vertical line. (**e**,**f**) Splicing between 4A and 4B was analysed as described above in **c** and **d**, respectively. (**g**,**h**) Sanger sequencing of the PCR amplicon obtained from cDNA of *ZRSR2* mutant samples using primers located in 4A and retained intron (upstream of 4′). The PCR product was cloned into TOPO TA vector and individual clones were sequenced to verify two alternative splice donor sites (4′ and 4″). (**i**) RNA-Seq reads mapped to the genomic region encompassing exons 3–6 of *FRA10AC1* gene. (**j**) Aberrant splice donor site (denoted 4′) in intron 4 of *FRA10AC1* is shown for a representative *ZRSR2* mutant sample. Sequence and intron-type for the cryptic splice junction are indicated. (**k**) qRT-PCR to measure the relative levels of mis-spliced RNA in *ZRSR2* mutant and control samples using primers located in intron 4 and exon 5. *GAPDH* was used to normalize the levels of transcripts. (**l**) Sanger sequencing of the PCR product obtained from *ZRSR2* mutant samples in **j**. The junction between intron 4 and exon 5 is indicated by the dashed vertical line.

were substantially higher in the *ZRSR2* mutant MDS as compared with either the *ZRSR2* WT MDS or normal BM samples (Fig. 5c,e). Moreover, Sanger sequencing of the products amplified from *ZRSR2* mutant samples verified the predicted splice junctions (Fig. 5d,f). We detected two alternative splice donor sites (4′ and 4″ located 19 bp apart) both of which resulted in splicing to cryptic exon 4A (Fig. 5g,h). Similarly, in *FRA10AC1* and *SRPK2*, anomalous splice junctions were created from cryptic U2-type splice sites, which resulted in partial retention of U12-type introns in *ZRSR2* mutant samples (Fig. 5i,j, Supplementary Fig. 11a,b and Supplementary Data 3 and 4). The presence of these alternative splicing junctions was also validated by PCR and Sanger sequencing (Fig. 5k,l and Supplementary Fig. 11c,d). Likewise, in conventional RT–PCR analysis of *WDR41*, *FRA10AC1* and *SRPK2*, aberrantly spliced transcripts were detectable only in *ZRSR2* mutant MDS samples (Supplementary Fig. 12).

We also detected a few instances of increased exon skipping in the *ZRSR2* mutant group. The exon skipping phenotype was observed in the transcripts containing the U12-type introns and often involved exons flanking the U12-type intron (Supplementary Data 5).

Downregulation of ZRSR2 alters growth and differentiation. We investigated the consequences of ZRSR2 suppression on cell growth and haematopoietic differentiation, using shRNA-mediated knockdown. We observed that the ZRSR2-deficient leukemia cells divided moderately slower than the cells transduced with control shRNA vector (data not shown). In soft agar colony assay,

downregulation of ZRSR2 resulted in pronounced reduction in the number of colonies obtained in TF-1 and K562 cells (Fig. 6a). In propidium iodide staining of steady-state TF-1 cells, fewer cells were detected in S-phase of the cell cycle (Supplementary Fig. 13a). In addition, lower proportion of ZRSR2 knockdown cells incorporated 5-bromodeoxyuridine in *in vitro* labelling assay (Supplementary Fig. 13b), indicating that the ZRSR2-deficient cells divide slower than the control cells. We further tested the *in vivo* tumorigenic potential of ZRSR2 knockdown K562 cells in mice. Cells were injected subcutaneously in NOD-scid-gamma mice and the tumour growth was assessed. ZRSR2-deficient K562 cells produced smaller tumours as compared with the cells transduced with control shRNA (Fig. 6b). These results illustrate that downregulation of ZRSR2 suppresses cellular growth both *in vitro* and *in vivo*.

Next, the implications of ZRSR2 deficiency on myeloid differentiation were investigated using an *in vitro* model of differentiation of human $CD34^+$ haematopoietic stem cells (HSCs). $CD34^+$ cells enriched from cord blood were transduced with either ZRSR2 shRNA or control shRNA lentivirus, and analysed for *in vitro* clonogenic growth in methylcellulose media. We observed a marked decrease in the number of Burst-forming unit-erythroid (BFU-E) obtained after 9 days. On the other hand, a notable increase in the number of colony-forming unit-macrophage (CFU-M) occurred, while the number of colony-forming unit-granulocyte (CFU-G) was unaffected (Fig. 6c). We confirmed an effective knockdown of *ZRSR2* in transduced cells using qRT–PCR (Supplementary Fig. 14). Next, we evaluated the differentiation profile of $CD34^+$ cells using flow cytometry. Following transduction with lentivirus, cells were cultured in the presence of cytokines for 2 weeks. A significant increase in the proportion of $CD11b^+$ myeloid cells was observed on knockdown of ZRSR2 (Fig. 6d,e). Downregulation of ZRSR2 also resulted in a reduced proportion of erythroid precursors co-expressing Glycophorin A and CD71 surface antigens (Fig. 6f,g). These results indicate that suppression of ZRSR2 alters erythroid and myeloid differentiation of HSCs, presumably as a result of dysregulated splicing of genes implicated in haematopoiesis.

Pathways regulated by mis-spliced genes in *ZRSR2* mutant MDS. To reveal the enrichment of functionally related genes among the significantly mis-spliced transcripts in *ZRSR2* mutant MDS (Supplementary Data 1,4 and 5), Gene Ontology (GO) analyses was performed using the standard enrichment computation method. A significant enrichment was obtained for several pathways ($P<0.05$) including mitogen-activated protein kinase (MAPK) signalling, ErbB signalling and for genes associated with chronic myeloid leukemia and AML (Fig. 7a). We sought to identify downstream targets of aberrant splicing of ZRSR2, which may contribute to the disease phenotypes. Several genes, which participate in either haematopoietic differentiation or are

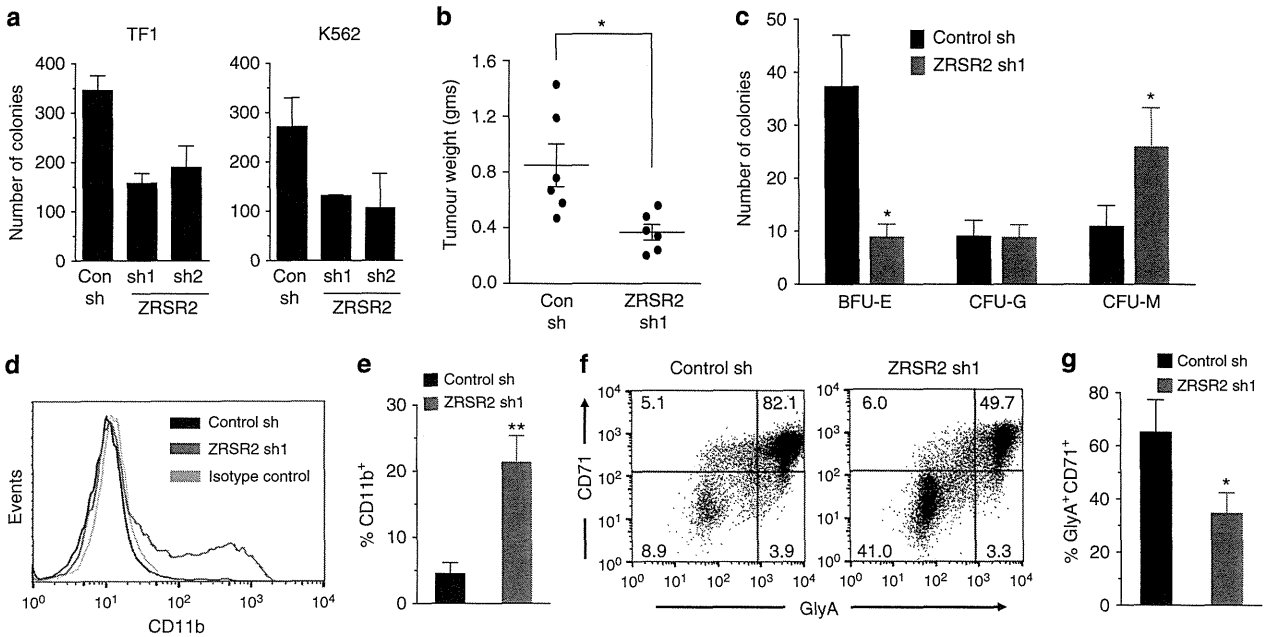

Figure 6 | Stable knockdown of ZRSR2 alters cell growth and differentiation. (**a**) Colony-forming ability of TF-1 and K562 cells transduced with ZRSR2 shRNA and control vectors was evaluated using soft-agar colony assay. One thousand and five hundred cells were seeded in each well of a 24-well plate and colonies were enumerated after 2 weeks. Cells were plated in triplicate and the data are the mean ± s.e.m. from multiple experiments (TF-1: $n=8$ for sh1 and $n=3$ for sh2; K562: $n=2$). (**b**) ZRSR2 knockdown and control K562 cells were transplanted into the flank of NOD-scid-gamma (NSG) mice; tumours were dissected after 2 weeks and weighed. Data represent mean ± s.e.m. (**c**) Cord blood-derived $CD34^+$ cells were transduced with ZRSR2 shRNA lentivirus and plated in methylcellulose media containing stem cell factor (SCF), interleukin-3 (IL-3), granulocyte colony-stimulating factor (G-CSF), granulocyte macrophage colony-stimulating factor (GM-CSF) and erythropoietin (EPO), as well as 1 μg ml^{-1} puromycin. Burst-forming unit-erythroid (BFU-E), colony-forming unit-granulocyte (CFU-G) and colony-forming unit-macrophage (CFU-M) colonies were counted after 9 days. Data represent the mean ± s.e.m. of three experiments. (**d–g**) $CD34^+$ cells were transduced as in **c**, cultured in liquid media containing SCF, IL-3, GM-CSF and EPO for 2 weeks and analysed using flow cytometry. (**d**) A representative overlay of histograms showing staining with αCD11b antibody used as a marker of differentiation towards myeloid lineage. (**e**) Percentage of $CD11b^+$ cells obtained after *in vitro* differentiation of $CD34^+$ cells in four experiments are depicted (mean ± s.e.m.). (**f**) Erythroid differentiation was measured using surface expression of Glycophorin A and CD71 antigens in cells cultured for 2 weeks in the presence of cytokines. Representative dot plots for flow cytometric analysis of ZRSR2 knockdown and control cells are shown. Percentage of cells in each quadrant are indicated. (**g**) Bar graphs represent percentages of $GlyA^+CD71^+$ cells from three experiments (mean ± s.e.m.). *P <0.05, **P <0.01. P-values were calculated using Student's t-test.

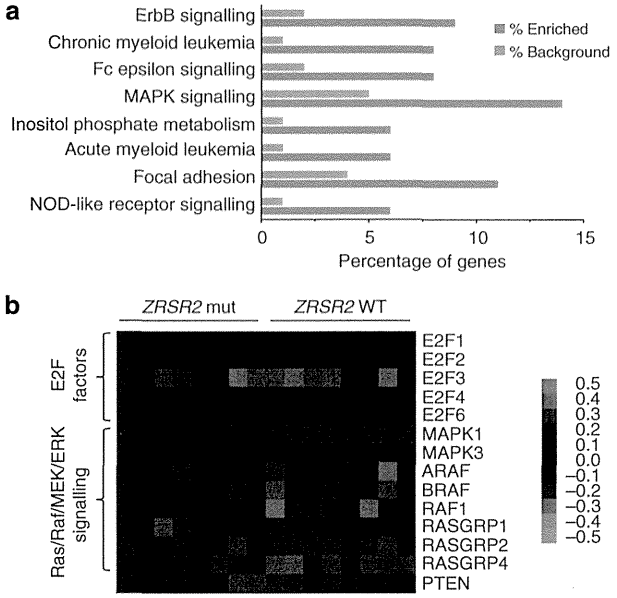

Figure 7 | GO analyses of mis-spliced genes in *ZRSR2* mutant MDS. (**a**) GO analysis of 251 significantly mis-spliced genes in *ZRSR2* mutant MDS shows their enrichment in a number of essential biological pathways ($P < 0.05$). (**b**) Heat map portrays the relative scale of aberrant retention of U12-type introns in genes involved in haematopoietic development. MSI values for U12-type introns in *ZRSR2* mutant and control samples were utilized for depiction of mis-splicing.

implicated in myeloid malignancies, were consistently mis-spliced in all *ZRSR2* mutant MDS samples (Fig. 7b). For instance, members of E2F transcription factors—E2F1, E2F2, E2F3, E2F4 and E2F6—which function during myeloid differentiation[29–35], exhibit splicing defects in *ZRSR2* mutant cells. Similarly, various regulators of MAPK signalling, including MAPK1, MAPK3, RAS guanyl releasing proteins and RAF serine/threonine protein kinases contain U12-type introns and were mis-spliced in *ZRSR2* mutant cells. These proteins mediate vital signalling cascades and their role in maintaining physiological haematopoiesis has been identified[36–44]. Another interesting candidate is the tumour suppressor gene, *PTEN*, loss of which impairs HSC activity, alters their lineage commitment and leads to myeloid disorders in mice[45,46]. Dysregulation of these genes through aberrant splicing can potentially have direct implications on haematopoietic differentiation and may contribute to the pathogenesis of MDS. Further investigations are necessary to identify effector(s) of the MDS phenotype in *ZRSR2* mutant cells.

Overall, our results demonstrate that *ZRSR2* mutations lead to aberrant splicing, primarily involving U12-type introns. These splicing defects are also corroborated in knockdown cells, which display alterations in growth and differentiation, substantiating an essential and non-redundant role of ZRSR2 in the U12-dependent spliceosome.

Discussion

Discovery of mutations in several spliceosome genes in MDS strongly suggest existence of dysfunctional splicing machinery in this disease. Interestingly, majority of the mutated genes including *SF3B1*, *U2AF1*, *ZRSR2*, *SRSF2*, *SF1* and *SF3A1* (ref. 8), encode for components of E/A splicing complex, which is involved in the recognition of splice sites. Recurrent somatic mutations in these genes, which occur mutually exclusively, indicate that the disruption of initial splicing steps is a common feature in MDS. Therefore, these mutations can be predicted to cause widespread alterations in splicing and gene expression. However, contrary to this hypothesis, mutations in *U2AF1* and *SF3B1* exert splicing changes in a specific subset of introns and exons[11,47,48]. Hence, this raises the question whether mutations of each splice factor has a distinctive effect on the splicing machinery, which possibly results in diverse phenotypes. In fact, evidence shows an association between splice factor mutation and the clinical phenotype. *SF3B1* mutations are found at high frequency in MDS subtypes characterized by the presence of ring sideroblasts, *SRSF2* mutations are highly associated with chronic myelomonocytic leukemia, while mutations in *ZRSR2* are often observed in RAEB-1 and RAEB-2 subtypes[8,13,47,49–52]. In this study, we demonstrate that the depletion of ZRSR2 leads to a specific splicing defect by disrupting the splicing of the entire subset of U12-type introns. Therefore, our study and previous reports indicate that the mutations in individual splice factors are likely to alter the function of splicing machinery in an exclusive manner.

Our results define an essential role of ZRSR2 in splicing of U12-type introns and propose it as one of the key components of the minor spliceosome. Our lentiviral shRNA-mediated knockdown approach demonstrates a specific defect in splicing of U12-type introns, which can be reversed by transient overexpression of ZRSR2. Most notably, RNA-Seq revealed that the transcriptome of *ZRSR2* mutant MDS exhibited several splicing defects, invariably resulting from inefficient splicing of the U12-type introns. Surprisingly, the U2-type introns were spliced with similar efficiency as the control cells. Although ZRSR2 was shown to be required for *in vitro* splicing of U2-type introns[22,23], our results provide the first evidence that *in vivo* splicing of U2-type introns is essentially unaffected in the absence of ZRSR2. We detect significant retention of U2-type introns only in transcripts that contain U12-type introns. Rare U2-type introns independent of the U12 transcripts appear in our analysis of mis-spliced introns. However, these introns exhibit weak mis-splicing phenotype and do not indicate any specificity with respect to either splice site strength or intron length, and therefore, are probable outliers in our analysis. These results signify ZRSR2 is crucial to the U12 spliceosome machinery, while it may have either a limited or redundant role in the U2 machinery. In addition, ZRSR2 has been identified as a component of U11/U12 snRNP[53]. The importance of ZRSR2 in U12-dependent spliceosome was also noted by the inability of cell extract lacking ZRSR2 to form a U12 splicing complex assembly in *in vitro* assays[23]. Further, a perfect correlation exists in the phylogenetic distribution of organisms that have both ZRSR2- and U12-dependent splicing, which also strongly supports the involvement of this splice factor in the U12 machinery[23]. Our computational analysis also demonstrates that *ZRSR2* mutations cause mis-splicing of a majority of U12-type introns, except for a minor subset with weaker U12-type splice sites.

ZRSR2 contacts the 3′-splice site of the U12-type intron of *P120* transcript by binding the A residue of the AC dinucleotide at the 3′-splice site[23]. This suggests that ZRSR2 is needed for recognition of 3′-splice sites in U12-type introns. Interestingly, the U12-spliceosome complexes (Complex A and Complex B/C) failed to assemble at *P120* pre-mRNA in its absence[23], indicating that recognition of the whole intron is impaired. Analysis of our RNA-Seq data also suggests that the ZRSR2-mediated 3′-splice site recognition is required for the 5′-splice site and branch site recognition. This is evident by activation of cryptic 5′-splice site sequences in certain U12-type intron containing transcripts such as DRAM2, TAPT1 and VPRBP in *ZRSR2* mutant MDS (Supplementary Fig. 15). Similar instances of cryptic splice sites were previously reported for deficiency of RNPC3 and U11–48K,

proteins also involved in the U12 splicing machinery[54,55]. The cryptic splice junctions observed in *ZRSR2* mutant cells were invariably U2 type.

Unlike the U2-dependent spliceosome, the U12-dependent machinery has been less well understood and proteins involved in splice-site recognition are not clearly defined. U12 spliceosome was initially uncovered as a machinery instrumental in excising introns lacking the consensus GT-AG splice site termini[56]. However, it was soon recognized that this splicing complex was responsible for a unique subset of evolutionary conserved introns, which had a highly conserved splice-site recognition sequences distinct from the U2-type introns[16,17,21,24] and which used different snRNPs than those used in U2 machinery[18,19]. Although the U12-type introns comprise ~0.5% of all human introns, they exist in several crucial genes involved in vital cellular processes[57]. Germline mutations in the *RNU4ATAC* gene, which encodes for an essential component of the U12 spliceosome, have been shown to cause MOPD1/TALS (Microcephalic Osteodysplastic Primordial Dwarfism type 1/Taybi–Linder Syndrome), a rare developmental disorder in humans[58,59]. Biallelic mutations in the *RNPC3* gene, which encodes for a 65-kDa constituent of U11/U12 di-snRNP, also lead to isolated familial growth hormone deficiency[55]. Moreover, an intact U12 spliceosome is essential for the development of *Drosophila* (despite the presence of <20 U12-type introns in its genome) and zebrafish[60–62]. Thus, the questions raised in the context of myeloid neoplasms are: what are the consequences of the impairment of U12 splicing specifically in HSCs and how the *ZRSR2* mutations (and other spliceosome gene mutations, in general) contribute to leukemogenesis. We show that knockdown of ZRSR2 in human HSCs alters their *in vitro* differentiation potential. Reduced differentiation occurs towards erythroid lineage accompanied by a higher proportion of $CD11b^+$ myeloid cells on knockdown of ZRSR2. These results suggest that a competent U12 spliceosome is required for normal myeloid differentiation. Further investigations will focus on identifying downstream target genes, which on mis-splicing lead to pathogenic consequences in MDS. In this study, the GO analysis of U12-type genes identifies few such molecular pathways with potential involvement in the MDS phenotype. We find that genes encoding E2F transcription factors and several components of the Ras/Raf/MEK/ERK signalling exhibit aberrant splicing of U12-type introns consistently in all *ZRSR2* mutant MDS. These proteins have been implicated in normal and malignant haematopoiesis and represent possible candidates for future investigations. Studies using murine models will help identify downstream targets and better understand the role of ZRSR2-mediated splicing in haematopoietic differentiation. Pathway analysis of mis-spliced genes also revealed a strong enrichment of key cellular functions such as RNA transport, cell cycle, cellular response to stress, response to DNA damage stimulus, protein transport, protein serine/threonine kinase activity and ribonucleotide binding among the mis-spliced genes (Supplementary Fig. 16). These functional classes have been previously attributed to the genes containing the U12-type introns[57].

Downregulation of ZRSR2 impaired *in vitro* clonogenic ability and suppressed tumour formation in mice. Overall, the ZRSR2 knockdown leukemia cells showed a general tendency to grow slower than the control cells. This observation is similar to the effect on cell growth reported for other spliceosome mutations. Expression of mutant *U2AF1* suppressed the growth of cell lines and resulted in lower reconstitution potential in mice[8]. Cell cycle arrest and inhibition of growth in leukemia cell lines is also attributed to downregulation of SF3B1 (ref. 63). Therefore, the spliceosome mutations do not seem to contribute towards a proliferative advantage of the haematopoietic precursors in MDS. How these cells harbouring spliceosome mutations achieve clonal expansion remains unclear. One explanation can be that other accompanying genetic alterations are necessary to confer a proliferative advantage. In fact, mutations in several components of epigenetic machinery co-occur with spliceosome mutations. A high incidence of co-occurrence of mutations in the *TET2* and *ZRSR2* genes has been noted in MDS[8,13]. Loss of TET2 has been shown to lead to myeloproliferation and transformation of haematopoietic precursors in mice[64,65]. Therefore, mutations in *TET2* are likely to promote clonal dominance. Mutations in spliceosome genes, which occur early during leukemogenesis[66], might play a prominent role in modifying the differentiation potential of myeloid precursors, thus contributing to abnormal precursor phenotype, which is a hallmark of MDS. Further studies into co-operativity between mutations in spliceosome and epigenetic modifier genes in pathogenesis of myeloid disorders are therefore warranted.

Methods

Generation of stable ZRSR2 knockdown cell lines. Lentiviral shRNA vectors (pLKO.1) were either purchased from Sigma or assembled by cloning the shRNA hairpin loop sequence into the AgeI/EcoRI sites of the empty vector. The target sequence for the ZRSR2 shRNAs sh1 and sh2 are 5′-CAACAGTTCCTA-GACTTCTAT-3′ and 5′-AGCAGCCCTTTCTCTGTTTAA-3′, respectively. MISSION pLKO.1-puro Non-Mammalian shRNA Control (SHC002; Sigma) was used as control vector for transduction. To generate lentiviral particles, 293T cells (kindly provided by Dr Bing Lim, Genome Institute of Singapore, Singapore) were co-transfected with shRNA plasmid and the packaging plasmids, pMISSIONgagpol and pMISSIONvsvg (Sigma) using Lipofectamine 2000 (Invitrogen). Virus containing supernatant was collected after 48 and 72 h, filtered through 0.45 μm filter and stored in aliquots at −80 °C. TF-1 and K562 cells (American Type Culture Collection) were infected with lentivirus for two rounds, 24 h apart, in the presence of 5 μg ml^{-1} protamine sulfate. Transduced cells were selected in puromycin to generate stable knockdown cell lines. The knockdown was verified using qPCR and western blotting.

Western blot analysis. Total protein lysates were prepared using M-PER Mammalian Protein Extraction Reagent (Thermo Scientific) containing protease inhibitor cocktail (Roche). Twenty-five micrograms of protein was resolved on 10% SDS–PAGE gel and transferred to Immobilon-P polyvinylidene difluoride membrane (Millipore). Anti-ZRSR2 antibody (1:2,500 dilution; kindly provided by Dr Michael Green, University of Massachusetts Medical School, MA) was used to determine ZRSR2 protein expression. The membrane was stripped before probing with anti-GAPDH antibody (1:5,000 dilution; Cell Signaling Technology).

Minigene splicing assays. *P120* reporter construct was generously provided by Dr Richard A. Padgett, Cleveland Clinic, Cleveland, OH, and the *GH1* minigene plasmid was provided by Dr Kinji Ohno, Center for Neurological Diseases and Cancer, Nagoya, Japan. 293T cells were transfected with *P120* or *GH1* plasmids using jetPRIME transfection reagent (Polyplus Transfection SA). Briefly, cells in 60 mm dish were transfected with 200 ng of reporter plasmid. Cells were harvested 48 h after transfection and total RNA was extracted using AxyPrep Multisource Total RNA Miniprep Kit (Axygen). RNA was treated with DNase I followed by complementary DNA synthesis using RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific). For *P120* minigene assay, plasmid-specific primer was used for reverse transcription. The cDNA was used as a template to amplify the region of *P120* gene containing intron F at 94 °C for 2 min followed by 25 cycles of 94 °C for 30 s, 60 °C for 40 s, 72 °C for 40 s and a final extension at 72 °C for 7 min. Amplified PCR product was resolved on agarose gel, stained with ethidium bromide and photographed. Bands corresponding to spliced (112 bp) and unspliced (211 bp) products were quantified using Image Lab software (Biorad).

For *GH1* minigene, reverse transcription was performed using random primers. The PCR to amplify the *GH1* transcript was performed at 94 °C for 2 min followed by 28 cycles of 94 °C for 30 s, 60 °C for 40 s, 72 °C for 40 s and a final extension at 72 °C for 7 min. Amplicons corresponding to fully spliced and exon skipped transcripts measured 447 and 327 bp, respectively. The sequences of primers used for RT–PCR are provided in the Supplementary Table 1.

RT-PCR to determine splicing efficiency of U2-type and U12-type introns. Splicing of U2-type and U12-type introns was measured using qRT–PCR[47,59]. Briefly, RNA was treated with DNase I (Thermo Scientific) followed by reverse transcription using RevertAid M-MuLV Reverse Transcriptase (RevertAid First Strand cDNA Synthesis Kit, Thermo Scientific) in the presence of random primers. Spliced and unspliced levels of introns were measured using two separate qPCRs and normalized to *GAPDH* transcript levels. PCR conditions included an initial

denaturation at 95 °C followed by 40–50 cycles of denaturation at 95 °C for 15 s and annealing/extension at 60 °C for 30 s. Primer sequences used for U2-type and U12-type introns are provided in Supplementary Table 1. Splicing efficiency was calculated as a ratio of relative quantities of spliced and unspliced pre-mRNA levels and was set as 1 for the control transduced cells.

Overexpression of ZRSR2. Full-length coding sequence of human *ZRSR2* gene (1449 bp) was amplified using PCR and cloned into BamHI/NotI sites of pCDNA3.1 expression vector (Invitrogen). Two micrograms of plasmid was used to transfect 293T cells in a 60-mm petri dish using jetPRIME transfection reagent (Polyplus Transfection), according to the manufacturer's protocol. RNA was extracted 72 h after transfection and overexpression of *ZRSR2* was verified using qPCR. RT–PCR to assess the splicing efficiency of U12-type introns was performed as described above.

Bone marrow samples. Bone marrow aspirates of diagnostic samples, including MDS and non-malignant cases, were obtained at the MLL Munich Leukemia Laboratory. Informed consent was obtained in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the MLL. Unfractionated bone marrow mononuclear cells were lysed and total RNA was extracted using RNeasy Kit (Qiagen).

Soft-agar colony assay. Colony-forming ability of TF-1 and K562 cells was determined by plating the cells in semi-solid media containing agar. The bottom layer consisted of 0.5% agar supplemented with 20% fetal bovine serum, RPMI-1640 medium and antibiotics/antimycotic (Gibco). One thousand and five hundred cells were mixed with the top agar layer, which contained 0.35% agar and RPMI-1640 medium, fetal bovine serum, L-glutamine, β-mercaptoethanol and antibiotics/antimycotic, which was plated in each well of 24-well dish. Two weeks after plating, colonies were enumerated under the microscope from triplicate wells.

Xenograft tumour model. All mice experiments were approved by the Institutional Animal Care and Use Committee, National University of Singapore, Singapore. Ten million K562 cells were resuspended in 50% matrigel and injected subcutaneously into flank of 8-week-old female NOD-scid-gamma mice. Each mouse was injected with control cells in one flank and the ZRSR2 knockdown cells in the other. Once tumours were palpable, mice were killed, tumours were harvested and weighed.

***In vitro* differentiation of human CD34$^+$ cells.** CD34$^+$ cells were enriched from fresh cord blood and transduced twice, 24 h apart, with ZRSR2 shRNA lentivirus. For colony assay, 2,500 transduced cells were plated per well in a six-well dish in methylcellulose media (Stemcell Technologies) supplemented with stem cell factor, granulocyte colony-stimulating factor, granulocyte macrophage colony-stimulating factor, interleukin-3 and erythropoietin, and containing puromycin. Colonies were enumerated after 9 days. In some experiments, cells were harvested from colonies and used to extract RNA. Downregulation of *ZRSR2* was examined using qRT–PCR. For liquid culture, cells were maintained in the presence of single cell factor, granulocyte macrophage colony-stimulating factor, interleukin-3, erythropoietin and puromycin for 2 weeks. Expression of Glycophorin A and CD71 (erythroid lineage) and CD11b (myeloid lineage) was determined using flow cytometry. Data were analysed using Flowjo software.

RNA sequencing. *Library preparation and sequencing.* cDNA libraries were prepared using TruSeq RNA Sample Preparation Kit (Illumina), according to the manufacturer's protocol. One microgram of total RNA was used for library preparation and paired-end adapters were ligated to DNA fragments before amplification and sequencing on a HiSeq 2000 instrument (Illumina) with 100 bp paired-end reads, according to manufacturer's protocol. We first mapped the sequenced reads to the reference transcript in the database obtained from RefSeq, Ensemble and UCSC known genes using bowtie, and unmapped or poorly mapped reads were realigned to human reference genome (hg19) with the Blat software. This two-step mapping procedure is included in genomon-fusion (http://genomon.hgc.jp/rna/) pipeline.

The mapped reads were organized into five different libraries as follows: (a) exon read library, (b) intron read library, (c) exon–intron junction read library, (d) exon–exon junction read library and (e) intergenic read library. The gene classification was done using all known RefSeq transcripts. To be considered as a real junction, two criteria were applied. First, we required the splice junction to be supported by at least five reads and aligned reads spanned a minimal of 4 bp on each side of the junction.

Differential splicing analysis. To identify splicing events, we introduced a parameter called MSI, which is equivalent to Percent Spliced-in values used for alternative splicing analysis[67,68] and modified to cater to different splicing events such as intron retention, exon skip and incorrect splice site usage (see Supplementary Fig. 5 for mathematical expression of MSI). For identification of differential splicing between two samples, the difference in MSI (ΔMSI) was applied: $\Delta MSI = MSI_{ZRSR2\ mutant} - MSI_{control}$. Furthermore, we used the Fisher's exact test to evaluate the significance of such difference and adjusted *P*-value by FDR analysis to minimize the false positives. For identification of differentially spliced events among two genotypes, we performed all possible pairwise analyses between *ZRSR2* mutant MDS and control (*ZRSR2* WT MDS and normal BM) samples. These include 32 comparisons each between '*ZRSR2* mutant MDS' and '*ZRSR2* WT MDS' or '*ZRSR2* mutant MDS', and 'normal BM' (total 64 pairwise comparisons). As control, *ZRSR2* WT MDS samples were compared with normal BM samples ($4 \times 4 = 16$ comparisons). A similar approach was used to compare the mis-splicing in ZRSR2 knockdown versus control-transduced TF-1 cells. Statistical difference of $P \leq 0.01$ and difference in $\Delta MSI > 20$ were considered as significant differential splicing in each comparison. To identify an overall significant mis-splicing between the two genotypes (*ZRSR2* mutant versus controls), we considered the frequency of occurrence of mis-splicing events in pairwise comparisons and applied an FDR of ≤ 0.01 for intron retention and ≤ 0.02 for abnormal splice site recognition and exon skip analysis, using the control-versus-control comparisons as background.

For intron retention analysis, we selected events where at least one read overlaps an intron and flanking exons at each of the two junctions, and the total number of such junction read counts was a minimum of 4, in at least one sample. We also applied the criteria that $\geq 95\%$ of the intron was covered by at least one read (Supplementary Fig. 17).

Classification of introns. First, position weight matrices were generated based on the 5′- and 3′-splice sites and the branch site sequence. These matrices were used to scan the introns of interest and the introns were then categorized as U2-type or U12-type based on the mapping score[27].

Gene expression analysis. The relative abundance of transcripts was quantified using normalized fragments per kilobase of transcript per million fragments mapped, which were calculated using bedtools with a transcriptome reference.

Gene function analysis for mis-spliced genes. Genes significantly mis-spliced in *ZRSR2* mutant MDS cases were analysed using GO tools, to identify enriched GO terms for biological pathways, biological processes and molecular functions. Significant enrichment was computed based on the Fishers' exact test, using the numbers of mis-spliced genes compared with the numbers in the genome for each GO term. The obtained *P*-value was further corrected for FDR as described[69]. GO pathway analysis tool used was the MetaCore from GeneGO (https://portal.genego.com/) and the GO analysis for biological processes and molecular functions was done with DAVID (http://david.abcc.ncifcrf.gov). The corrected *P*-value cutoff of 0.05 was used for significant enriched GO terms. Heat maps were created for enriched biological processes and molecular functions, using Cluster 3.0 software.

Validation of mis-spliced genes. Intron retention in *ZRSR2* mutant MDS was verified using qRT-PCR[70]. Expression of each intron was normalized against the expression of flanking exons. For validation of other mis-spliced introns, qPCR or conventional PCR were used. *GAPDH* was used as the normalization control for cDNA input. Primer sequences are available in the Supplementary Table 1.

References

1. Hofmann, W. K. & Koeffler, H. P. Myelodysplastic syndrome. *Annu. Rev. Med.* **56,** 1–16 (2005).
2. Greenberg, P. L. The multifaceted nature of myelodysplastic syndromes: clinical, molecular, and biological prognostic features. *J. Natl Compr, Canc. Netw.* **11,** 877–884 quiz 885 (2013).
3. Ma, X. Epidemiology of myelodysplastic syndromes. *Am. J. Med.* **125,** S2–S5 (2012).
4. Tefferi, A. & Vardiman, J. W. Myelodysplastic syndromes. *New Engl. J. Med.* **361,** 1872–1885 (2009).
5. Sekeres, M. A. & Bejanyan, N. The revolution of myelodysplastic syndromes. *Ther. Adv. Hematol.* **2,** 33–43 (2011).
6. Shih, A. H., Abdel-Wahab, O., Patel, J. P. & Levine, R. L. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat. Rev. Cancer* **12,** 599–612 (2012).
7. Schlegelberger, B., Gohring, G., Thol, F. & Heuser, M. Update on cytogenetic and molecular changes in myelodysplastic syndromes. *Leuk. Lymphoma* **53,** 525–536 (2012).
8. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478,** 64–69 (2011).
9. Maciejewski, J. P. & Padgett, R. A. Defects in spliceosomal machinery: a new pathway of leukaemogenesis. *Br. J. Haematol.* **158,** 165–173 (2012).
10. Thol, F. *et al.* Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood* **119,** 3578–3584 (2012).
11. Makishima, H. *et al.* Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood* **119,** 3203–3210 (2012).
12. Graubert, T. A. *et al.* Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat. Genet.* **44,** 53–57 (2012).
13. Damm, F. *et al.* Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood* **119,** 3211–3218 (2012).

14. Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122,** 3616–3627 (2013).
15. Lamond, A. I. The spliceosome. *Bioessays* **15,** 595–603 (1993).
16. Hall, S. L. & Padgett, R. A. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* **239,** 357–365 (1994).
17. Tarn, W. Y., Yario, T. A. & Steitz, J. A. U12 snRNA in vertebrates: evolutionary conservation of 5' sequences implicated in splicing of pre-mRNAs containing a minor class of introns. *RNA* **1,** 644–656 (1995).
18. Tarn, W. Y. & Steitz, J. A. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* **84,** 801–811 (1996).
19. Tarn, W. Y. & Steitz, J. A. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* **273,** 1824–1832 (1996).
20. Will, C. L. & Luhrmann, R. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol. Chem.* **386,** 713–724 (2005).
21. Patel, A. A. & Steitz, J. A. Splicing double: insights from the second spliceosome. *Nat. Rev. Mol. Cell Biol.* **4,** 960–970 (2003).
22. Tronchere, H., Wang, J. & Fu, X. D. A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. *Nature* **388,** 397–400 (1997).
23. Shen, H., Zheng, X., Luecke, S. & Green, M. R. The U2AF35-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev.* **24,** 2389–2394 (2010).
24. Hall, S. L. & Padgett, R. A. Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* **271,** 1716–1718 (1996).
25. Fu, Y., Masuda, A., Ito, M., Shinmi, J. & Ohno, K. AG-dependent 3'-splice sites are predisposed to aberrant splicing due to a mutation at the first nucleotide of an exon. *Nucleic Acids Res.* **39,** 4396–4404 (2011).
26. Kitamura, T. *et al.* Establishment and characterization of a unique human cell line that proliferates dependently on GM-CSF, IL-3, or erythropoietin. *J. Cell Physiol.* **140,** 323–334 (1989).
27. Sheth, N. *et al.* Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* **34,** 3955–3967 (2006).
28. Alioto, T. S. U12DB: a database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res.* **35,** D110–D115 (2007).
29. Chen, H. Z., Tsai, S. Y. & Leone, G. Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat. Rev. Cancer* **9,** 785–797 (2009).
30. Gery, S., Gombart, A. F., Fung, Y. K. & Koeffler, H. P. C/EBPepsilon interacts with retinoblastoma and E2F1 during granulopoiesis. *Blood* **103,** 828–835 (2004).
31. Kikuchi, J. *et al.* E2F-6 suppresses growth-associated apoptosis of human hematopoietic progenitor cells by counteracting proapoptotic activity of E2F-1. *Stem Cells* **25,** 2439–2447 (2007).
32. Trikha, P. *et al.* E2f1-3 are critical for myeloid development. *J. Biol. Chem.* **286,** 4783–4795 (2011).
33. Kadri, Z. *et al.* Direct binding of pRb/E2F-2 to GATA-1 regulates maturation and terminal cell division during erythropoiesis. *PLoS Biol.* **7,** e1000123 (2009).
34. Zhang, J. *et al.* pRB and E2F4 play distinct cell-intrinsic roles in fetal erythropoiesis. *Cell Cycle* **9,** 371–376 (2010).
35. Kinross, K. M., Clark, A. J., Iazzolino, R. M. & Humbert, P. O. E2f4 regulates fetal erythropoiesis through the promotion of cellular proliferation. *Blood* **108,** 886–895 (2006).
36. Chan, G., Gu, S. & Neel, B. G. Erk1 and Erk2 are required for maintenance of hematopoietic stem cells and adult hematopoiesis. *Blood* **121,** 3594–3598 (2013).
37. Chung, E., Hsu, C. L. & Kondo, M. Constitutive MAP kinase activation in hematopoietic stem cells induces a myeloproliferative disorder. *PLoS ONE* **6,** e28350 (2011).
38. Chung, E. & Kondo, M. Role of Ras/Raf/MEK/ERK signaling in physiological hematopoiesis and leukemia development. *Immunol. Res.* **49,** 248–268 (2011).
39. Hsu, C. L., Kikuchi, K. & Kondo, M. Activation of mitogen-activated protein kinase kinase (MEK)/extracellular signal regulated kinase (ERK) signaling pathway is involved in myeloid lineage commitment. *Blood* **110,** 1420–1428 (2007).
40. Geest, C. R. & Coffer, P. J. MAPK signaling pathways in the regulation of hematopoiesis. *J. Leukoc. Biol.* **86,** 237–250 (2009).
41. Stone, J. C. Regulation and function of the RasGRP family of Ras activators in blood cells. *Genes Cancer* **2,** 320–334 (2011).
42. Kamata, T. *et al.* A critical function for B-Raf at multiple stages of myelopoiesis. *Blood* **106,** 833–840 (2005).
43. Machnicki, M. M. & Stoklosa, T. BRAF--a new player in hematological neoplasms. *Blood Cells Mol. Dis.* **53,** 77–83 (2014).
44. Nelson, D. S. *et al.* Somatic activating ARAF mutations in Langerhans cell histiocytosis. *Blood* **123,** 3152–3155 (2014).
45. Zhang, J. *et al.* PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. *Nature* **441,** 518–522 (2006).
46. Yilmaz, O. H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441,** 475–482 (2006).
47. Visconte, V. *et al.* SF3B1 haploinsufficiency leads to formation of ring sideroblasts in myelodysplastic syndromes. *Blood* **120,** 3173–3186 (2012).
48. Przychodzen, B. *et al.* Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood* **122,** 999–1006 (2013).
49. Damm, F. *et al.* SF3B1 mutations in myelodysplastic syndromes: clinical associations and prognostic implications. *Leukemia* **26,** 1137–1140 (2012).
50. Malcovati, L. *et al.* Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* **118,** 6239–6246 (2011).
51. Wu, S. J. *et al.* The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. *Blood* **120,** 3106–3111 (2012).
52. Meggendorfer, M. *et al.* SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). *Blood* **120,** 3080–3088 (2012).
53. Will, C. L. *et al.* The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *RNA* **10,** 929–941 (2004).
54. Turunen, J. J., Will, C. L., Grote, M., Luhrmann, R. & Frilander, M. J. The U11-48K protein contacts the 5' splice site of U12-type introns and the U11-59K protein. *Mol. Cell. Biol.* **28,** 3548–3560 (2008).
55. Argente, J. *et al.* Defective minor spliceosome mRNA processing results in isolated familial growth hormone deficiency. *EMBO Mol. Med.* **6,** 299–306 (2014).
56. Jackson, I. J. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* **19,** 3795–3798 (1991).
57. Turunen, J. J., Niemela, E. H., Verma, B. & Frilander, M. J. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* **4,** 61–76 (2013).
58. Edery, P. *et al.* Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science* **332,** 240–243 (2011).
59. He, H. *et al.* Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science* **332,** 238–240 (2011).
60. Otake, L. R., Scamborova, P., Hashimoto, C. & Steitz, J. A. The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in *Drosophila*. *Mol. Cell* **9,** 439–446 (2002).
61. Konig, H., Matter, N., Bader, R., Thiele, W. & Muller, F. Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. *Cell* **131,** 718–729 (2007).
62. Markmiller, S. *et al.* Minor class splicing shapes the zebrafish transcriptome during development. *Proc. Natl Acad. Sci. USA* **111,** 3062–3067 (2014).
63. Boultwood, J., Dolatshad, H., Varanasi, S. S., Yip, B. H. & Pellagatti, A. The role of splicing factor mutations in the pathogenesis of the myelodysplastic syndromes. *Adv. Biol. Regul.* **54,** 153–161 (2013).
64. Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20,** 11–24 (2011).
65. Li, Z. *et al.* Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* **118,** 4509–4518 (2011).
66. Haferlach, T. *et al.* Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* **28,** 241–247 (2014).
67. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–476 (2008).
68. Shapiro, I. M. *et al.* An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet.* **7,** e1002218 (2011).
69. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).
70. Wong, J. J. *et al.* Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* **154,** 583–595 (2013).

Acknowledgements

We thank Dr Michael Green (University of Massachusetts Medical School) for sharing the αZRSR2 antibody, Professor Richard Padgett (Cleveland Clinic) for *P120* minigene plasmid and Dr Kinji Ohno (Nagoya University Graduate School of Medicine) for *GH1* reporter plasmid. We are indebted to Dr Ravi Sachidanandam (Mount Sinai Hospital) for valuable advice on classification of introns and Drs Ramon Tiu, Jarnail Singh and Valeria Visconte (Cleveland Clinic) for suggestions regarding experimental validation of intron retention. We are grateful to Drs Yasunobu Nagata (Kyoto), Tamara Alpermann, Andreas Roller (MLL Munich) and Abhijit Patel (Yale University School of Medicine) for useful discussions. We are also thankful to Sally Chi (Cedars-Sinai Medical Center) for help in arranging cord blood for experiments. This work was funded by the Singapore Ministry of Health's National Medical Research Council (NMRC) under its Singapore Translational Research (STaR) Investigator Award to H. Phillip Koeffler and the NMRC Centre Grant (NCIS Centre Grant Seed Funding) awarded to the National University Cancer Institute of Singapore, the NIH grant R01CA026038-33, NCIS Centre Grant Seed

Funding, the National Research Foundation Singapore and the Singapore Ministry of Education under its Research Centres of Excellence initiative.

Author contributions

V.M. designed the study, performed majority of experiments, interpreted the data and wrote the manuscript. D.K. and R.O. designed and performed the experiments. J.L. and A.S.-O. performed bioinformatics and statistical analyses. A.K. and V.G. prepared and provided primary patient specimen. M.S. performed RNA sequencing and provided experimental and analytical advice. J.S. performed experiments. Y.S. and S.M. designed the bioinformatics tools. F.T. and A.G. provided patient specimen. H.Y. designed and performed the bioinformatics and statistical analysis, and wrote the manuscript. T.H. provided primary patient samples and critical advice. S.O. provided support for RNA sequencing and data analysis, and provided critical advice. H.P.K. conceived and guided the study, interpreted the data and wrote the manuscript. All authors reviewed and approved the final manuscript.

Additional information

Accession codes: The RNA-Seq data have been deposited in the NCBI Gene Expression Omnibus under accession code GSE63816.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: T. H. declares part ownership of MLL Munich Leukemia Laboratory. A.K. and V.G. are employees of MLL Munich Leukemia Laboratory. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Madan, V. *et al.* Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat. Commun.* 6:6042 doi: 10.1038/ncomms7042 (2015).

ELSEVIER

CrossMark

Experimental Hematology

Experimental Hematology 2015;43:32–43

Variegated clonality and rapid emergence of new molecular lesions in xenografts of acute lymphoblastic leukemia are associated with drug resistance

Daniel Nowak[a,b], Natalia L.M. Liem[c], Maximilian Mossner[b], Marion Klaumünzer[b], Rachael A. Papa[c], Verena Nowak[a,b], Johann C. Jann[b], Tadayuki Akagi[d], Norihiko Kawamata[a], Ryoko Okamoto[a], Nils H. Thoennissen[a], Motohiro Kato[e], Masashi Sanada[e], Wolf-Karsten Hofmann[b], Seishi Ogawa[e], Glenn M. Marshall[f], Richard B. Lock[c], and H. Phillip Koeffler[a,g]

[a]*Division of Hematology and Oncology, Cedars Sinai Medical Center, University of California, Los Angeles, School of Medicine, Los Angeles, CA, United States;* [b]*Department of Hematology and Oncology, Medical Faculty Mannheim of the University of Heidelberg, Heidelberg, Germany;* [c]*Children's Cancer Institute Australia for Medical Research, Lowy Cancer Research Centre, University of New South Wales, Sydney, Australia;* [d]*Department of Stem Cell Biology, Graduate School of Medical Science, Kanazawa University, Ishikawa, Japan;* [e]*Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan;* [f]*Kids Cancer Centre, Sydney Children's Hospital, Randwick, Australia;* [g]*National University of Singapore, Singapore, Singapore*

(Received 2 April 2014; revised 18 August 2014; accepted 19 September 2014)

The use of genome-wide copy-number analysis and massive parallel sequencing has revolutionized the understanding of the clonal architecture of pediatric acute lymphoblastic leukemia (ALL) by demonstrating that this disease is composed of highly variable clonal ancestries following the rules of Darwinian selection. The current study aimed to analyze the molecular composition of childhood ALL biopsies and patient-derived xenografts with particular emphasis on mechanisms associated with acquired chemoresistance. Genomic DNA from seven primary pediatric ALL patient samples, 29 serially passaged xenografts, and six in vivo selected chemoresistant xenografts were analyzed with 250K single-nucleotide polymorphism arrays. Copy-number analysis of non–drug-selected xenografts confirmed a highly variable molecular pattern of variegated subclones. Whereas primary patient samples from initial diagnosis displayed a mean of 5.7 copy-number alterations per sample, serially passaged xenografts contained a mean of 8.2 and chemoresistant xenografts a mean of 10.5 copy-number alterations per sample, respectively. Resistance to cytarabine was explained by a new homozygous deletion of the *DCK* gene, whereas methotrexate resistance was associated with monoallelic deletion of *FPGS* and mutation of the remaining allele. This study demonstrates that selecting for chemoresistance in xenografted human ALL cells can reveal novel mechanisms associated with drug resistance.

Current treatment of pediatric acute lymphoblastic leukemia (ALL) consists of empirically optimized combinations of antileukemic drugs such as glucocorticoids, vincristine (VCR), asparaginase, methotrexate (MTX), mercaptopurine, cytarabine (ARA-C), cyclophosphamide, and others [1], which achieve five-year survival rates of 80% to 90% [2]. However, deaths due to relapse, primary treatment failure, or disease progression still occur in 10% to 15% of cases [2]. The elucidation of biological mechanisms for such relapses would improve treatment outcome for pediatric ALL.

In this context, xenograft models of ALL in immunodeficient (NOD/SCID) mice have proven to be valuable tools to study this disease in vivo [3–5]. Primary ALL cells derived from patient biopsies have minimal capacity for proliferation using currently available in vitro cell culture techniques. Xenograft models allow massive cell expansion and facilitate in vivo preclinical testing of experimental

RBL and HPK contributed equally to this work.

Offprint requests to: Dr. Daniel Nowak, Department of Hematology and Oncology, Medical Faculty Mannheim of the University of Heidelberg, Pettenkofer Strasse 22, Mannheim 68169, Germany; E-mail: daniel.nowak@medma.uni-heidelberg.de

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.exphem.2014.09.007.

0301-472X/Copyright © 2015 ISEH - International Society for Experimental Hematology. Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.exphem.2014.09.007

Table 1. Patient/xenograft demographics

Patient/xenograft ID	ALL subtype	Age at diagnosis (months)	Sex	Disease status at biopsy	Risk category	WBC at biopsy ($\times 10^9$/L)	ANZCCSG study	Cytogenetics of original patient biopsy [Number of analyzed metaphases]
ALL3	Pre-B	154	F	Diagnosis	High	93.9	6	46,XX,del(11q23)
ALL17	Common	107	F	Diagnosis	High	97.3	7	46,XX,-20,+21[18]/47,XX,-20,+21, +mar[4]
ALL25	Common	149	M	Diagnosis	Standard	4.54	8	46,XY,t(1;19)(q23;p13)
ALL26	Common	43	F	Diagnosis	Standard	90.0	8	46,XX,t(12;21)(p12;q22)
ALL27	T-ALL	104	M	Diagnosis	High	526	8	46,XY
ALL28	Common	20	M	Diagnosis	High	15.0	8	Hyperdiploid
ALL29	T-ALL	59	M	Diagnosis	High	86.3	7	46,XY
ALL30	T-ALL	90	M	Diagnosis	Very High	886	8	46,XY
ALL31	T-ALL	123	M	Diagnosis	Very High	212	8	46,XY,del(6)(q21),del(11) (q23)[4]/46, XY[14]

ANZCCSG = Australia and New Zealand Children's Cancer Study Group; *Common* = $CD10^+$ pre-B ALL; *pre-B* = B-cell precursor ALL; *T-ALL* = T-lineage ALL; *WBC* = white blood cell count.

therapies. Moreover, the process of serial xenotransplantation has enabled the identification and study of rare subsets of leukemia cells, which possess repopulation capacity, giving support for the tumor stem cell concept in leukemia and solid tumors [6,7].

With regard to the molecular pathogenesis of pediatric and adult ALL, substantially new understanding has been gained by the analysis of primary ALL samples with high density single-nucleotide polymorphism (SNP) arrays. Single-nucleotide polymorphism arrays allow a detailed genome-wide mapping for cryptic copy-number alterations (CNAs) and loss of heterozygosity such as copy-number neutral loss of heterozygosity (CNLOH), adding to the discovery of several important and prognostically relevant genomic lesions such as microdeletions of *IKZF1* and *PAX5* in ALL [8–12].

The combination of xenotransplantation models with high throughput genomic analysis, including SNP arrays, has recently led to the discovery that the subclonal architecture of ALL and acute myeloid leukemia cells is complex and variegated with linear or branching evolutionary histories [13–15].

In light of this, the aim of the current study was to assess the molecular composition of childhood ALL samples xenografted in immunodeficient mice with and without chemotherapeutic drug selection pressure. We explored the possibility of using this model to determine molecular mechanisms of chemoresistance when ALL xenografts were exposed in vivo to various chemotherapeutic agents used in the standard treatment of pediatric ALL. Single-nucleotide polymorphism arrays were used to perform a genome-wide copy-number analysis to compare several parameters: (1) xenografts compared with their matched primary biopsy, (2) xenografts subjected to up to four serial passages in comparison with their primary diagnostic sample, and (3) xenografts rendered chemoresistant in vivo against single chemotherapeutic compounds versus their passage-matched control cells. Our findings indicate that ALL continues to clonally evolve following the process of xenografting and that these cells had the capacity to quickly give rise to clones carrying molecular lesions associated with chemoresistance, which were sometimes not detectable in diagnostic samples by current methods.

Materials and methods

Development of xenografts, selection of in vivo–derived drug-resistant sublines, and study of in vivo drug responses

All experimental studies were approved by the Human Research Ethics Committee and the Animal Care and Ethics Committee of the University of New South Wales. Procedures by which xenografts from childhood ALL biopsies in immune-deficient NOD/SCID (NOD.CB17-Prkdc^{scid}/SzJ) mice were established have been described in detail previously [5,16]. Briefly, to establish initial xenografts, 2.5–10 $\times$ 10^6 mononuclear cells from bone marrow or peripheral blood biopsies were injected into the tail veins of 5- to 8-week-old female NOD/SCID mice. On the day of inoculation, mice received 250 cG of total body irradiation at a dose rate of 325 cG/min by parallel opposed 4 MV x-rays. Patient and xenograft characteristics are presented in Table 1 and have in part been described previously [16–18]. Continuous xenografts were established by inoculating 5 $\times$ 10^6 spleen-derived cells from highly engrafted mice into the tail veins of secondary recipient mice. For all mouse experiments involving patient-derived xenografts, the number of mice that successfully engrafted compared with the number inoculated exceeded 90%.

For in vivo selection of drug resistant sublines of xenografts ALL3 and ALL17, mice were treated with single agent vincristine (0.5 mg/kg weekly), cytarabine (100 mg/kg Monday through Friday every 3 weeks), methotrexate (5 mg/kg Monday through Friday every 2 weeks), or dexamethasone (DEX; 15 mg/kg Monday through Friday weekly). All drugs were administered by intraperitoneal (IP) injection. Leukemia engraftment and response to drug treatment were assessed by weekly enumeration of the proportion of human $CD45^+$ cells in the peripheral blood (%huCD45) [5]. Owing to previous experience showing that low level dissemination of xenograft cells into the peripheral blood is associated with high level infiltration of bone marrow and

Table 2. Experimental design and samples (n = 42) for analysis of genomic DNA with Affymetrix 250K arrays

	Xenografts				
Diagnosis biopsy	Passage 1	Passage 2	Passage 3	Passage 4	Relapse biopsy
ALL25-P-DX	ALL25-1	ALL25-2	ALL25-3	ALL25-4	ALL25-P-R
ALL26-P-DX	na	ALL26-2	ALL26-3	ALL26-4	na
ALL27-P-DX	ALL27-1	ALL27-2	ALL27-3	ALL27-4	na
ALL28-P-DX	ALL28-1	na	ALL28-3	na	na
ALL29-P-DX	ALL29-1	ALL29-2	ALL29-3	ALL29-4	na
ALL30-P-DX	ALL30-1	ALL30-2	ALL30-3	ALL30-4	na
na	ALL31a-1	ALL31a-2	ALL31a-3	ALL31a-4	na
Passage-matched control xenografts			In vivo drug-selected xenografts		
	ALL3-CTL			ALL3-VCR-R	
	ALL17-CTL1			ALL17-VCR-R1	
	ALL17-CTL2			ALL17-VCR-R2	
	ALL17-CTL3			ALL17-ARA-C-R	
				ALL17-MTX-R	
				ALL17-DEX-R	

na = Not available.

spleen [19], drug treatments were initiated when the %huCD45 reached 1% and continued until the %huCD45 increased through the drug treatments or until animals experienced leukemia- or drug-related morbidity. The number of emerging chemoresistant xenograft lines compared with the number of mice inoculated is depicted in Supplementary Table E1 (online only, available at www.exphem.org).

For confirmation of in vivo drug resistance, spleen-derived cells were engrafted into groups of three to four secondary recipient mice. Drug treatments were initiated when the %huCD45 reached 1% and consisted of vincristine (0.5 mg/kg IP × 4 weeks), cytarabine (100 mg/kg Monday through Friday IP every 3 weeks × 4), methotrexate (5 mg/kg Monday through Friday IP every 2 weeks × 4), or dexamethasone (15 mg/kg IP Monday through Friday weekly × 4). Individual mouse event-free survival (EFS) was calculated as the days from treatment initiation until the % huCD45 reached 25% or mice exhibited signs of leukemia-related morbidity. Event-free survival was represented graphically by Kaplan-Meier analysis, and survival curves were compared by the log rank test. The efficacy of drug treatment was evaluated by leukemia growth delay (LGD), calculated as the difference between the median EFS of vehicle control and drug-treated cohorts.

Extraction of genomic DNA

Genomic DNA (gDNA) was isolated with the DNeasy Blood and Tissue kit (QIAGEN, Valencia, CA) either from mononuclear cells of primary pediatric ALL biopsy samples obtained at diagnosis or relapse or from spleen-derived xenograft cells at a purity of >90% huCD45$^+$. A total of n = 42 gDNA samples were analyzed (Tables 1 and 2).

High density single-nucleotide polymorphism array analysis

Approximately 250 ng of gDNA from each sample was processed according to the genomic mapping 250K NspI protocol and hybridized to 250K NspI SNP arrays using the GeneChip Fluidics station 450 and GeneChip Scanner 3000 (Affymetrix, Santa Clara, CA) as described previously [20,21]. Data analysis of deletions, amplifications, and CNLOH was performed using the CNAG software (Copy Number Analyzer for Affymetrix GeneChip, Cancer Genomics Projekt, University of Tokyo, Japan) with nonmatched references as previously described [20,21]. Size, position, and location of genes contained in CNAs were determined with the University of California, Santa Clara, Genome Browser (http://genome.ucsc.edu/) and the Ensemble Genome Browser (http://www.ensembl.org/index.html). We excluded CNAs involving immunoreceptor and immunoglobulin genes from analysis. Genomic copy-number polymorphisms were excluded by comparison of the detected genomic lesions with the registered copy-number polymorphisms in these genome browsers.

Validation of copy number alterations and sequencing of mutations

For confirmation of genomic copy-number changes, Sybr-green quantitative real-time polymerase chain reaction (PCR) was performed on the gDNA of an in vivo cytarabine-selected subline (ALL17-ARA-C-R) and its corresponding passage-matched control sample (ALL17-CTL1) according to the $2^{-\Delta\Delta CT}$ method [22] on a Lightcycler 480 (Roche, Mannheim, Germany). The relative allele dosage of the putatively deleted genomic region in the drug-selected sample was compared with the same region

Table 3. Frequency of CNAs detected in the ALL samples from primary patient samples in comparison with xenografts and chemoresistant xenografts

	Primary patient samples (n = 7)		Xenografts (n = 29)		Chemoresistant xenografts (n = 6)	
Alterations	n	Mean per sample	n	Mean per sample	n	Mean per sample
Deletions	26	3.7	170	5.9	36	6
Duplications/ Amplifications	12	1.7	52	1.8	17	2.8
CNLOH	3	0.4	17	0.6	10	1.7
Total	41	**5.7**	239	**8.2**	63	**10.5**

Total mean values of CNAs are marked bold.

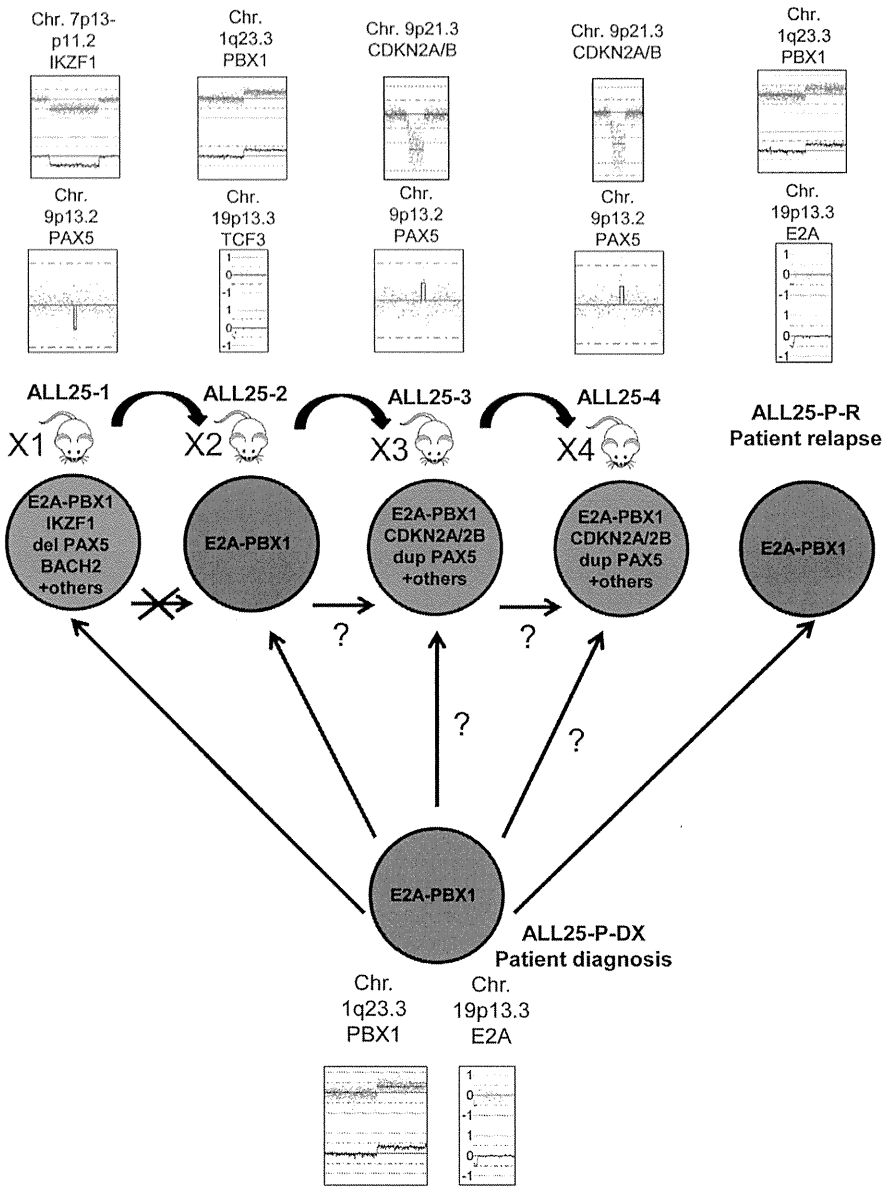

Figure 1. Heterogeneous clonal composition of serial ALL xenografts detected by SNP array genotyping. Boxes contain views from SNP array copy-number analysis of primary ALL patient samples from initial diagnosis (ALL25-P-DX), relapse (ALL25-P-R), and four serial ALL xenograft passages of ALL25 (ALL25-1–4). The copy number is visualized by red dots, which represent one SNP each (top of each box). The blue line represents a smoothed signal of copy number. In the diploid human genome, normal copy number = 2. Downward deviation of the copy-number signal indicates either a heterozygous or homozygous deletion; upward deviation indicates either duplication or amplification. An unbalanced t(1,19) translocation (*E2A-PBX1* fusion product) is present in all samples—patient and xenograft samples. Other genomic alterations are variable and allow the inference of different subclones. Red circle: predominating clone in the primary patient sample containing an unbalanced *E2A-PBX1* translocation. This clone is also found in the second-passage xenograft and the primary sample from relapse. Blue circle: predominating clone in xenograft passage 1. Apart from the unbalanced *E2A-PBX1* translocation, this clone contains additional lesions, such as deletions of *IKZF1*, *PAX5*, and *BACH2*, which could have evolved from the earlier (red) clone. Finding the red clone in xenograft passage 2 must mean that this clone persisted as a minor subclone in xenograft passage 1 and cannot have evolved from the predominating clone of that passage (crossed arrow). Green circle: predominating clone in xenograft passages 3 and 4, which contain additional lesions in the form of *CDKN2A/B* deletions and a duplication of *PAX5*.

in the control sample and with the allele dosage in the genomic regions flanking the deletion in both samples. All primer sequences are available on request. Sequencing of the *FPGS* gene was performed by direct Sanger sequencing of PCR-amplified genomic DNA on a 3130 Genetic Analyzer (Applied Biosystems, Waltham, MA). Deep sequencing of the identified *FPGS* mutation was carried out in a basic amplicon sequencing setup on a 454 GS Junior System (Roche).

Results

Single-nucleotide polymorphism array analysis of serial xenograft passages reveals heterogenic patterns of genomic alterations and indicates a diverse clonal architecture in acute lymphoblastic leukemia

To investigate the influence of serial passaging on DNA CNAs, we utilized a large panel of continuous xenografts that had previously been established in NOD/SCID mice for up to four passages [5] as well as patient biopsy samples obtained at diagnosis or relapse (Tables 1 and 2). A total of seven primary pediatric ALL patient samples obtained at diagnosis or relapse, 29 untreated xenografts, and six in vivo drug-selected xenograft sublines were interrogated for acquired CNAs (Table 2). A mean of 5.7 CNAs/sample from the primary patient samples, 8.2 CNAs/sample in the untreated ALL xenografts, and 10.5 CNAs/sample in the in vivo drug-selected xenografts were identified (Table 3). Most CNAs were heterozygous and homozygous deletions, but duplications, amplifications, and CNLOHs were also present and frequently affected common target genes such as *CDKN2A/CDKN2B*, *PAX5*, and *ETV6*. A complete and detailed documentation of all 343 identified CNAs is contained in Supplementary Tables E2–E4 (online only, available at www.exphem.org).

In the 29 untreated xenograft samples, seven sets had been serially passaged up to four times from the original diagnosis biopsy specimen. The CNAs of these seven sets showed that the xenografts established at first passage very frequently displayed genomic alterations different from those present in the original biopsy. Furthermore, serial transplantations following establishment of the initial xenograft revealed that their CNAs also diverged in a variegated fashion. In some of the xenograft lines with changing genomic profiles, specific genomic alterations were retained in all subsequent passages, while others were gained or lost in some xenografts.

This level of complexity is highlighted in Figure 1, which provides an overview of different patterns of genomic alterations found in serial transplantations of the diagnosis sample from which xenograft ALL25 was established as well as the patient's biopsy obtained at relapse. While an unbalanced *E2A-PBX1* translocation was retained throughout all samples of this case (primary and xenografts), the cells of the first-passage xenograft harbored additional deletions of *IKZF1*, *PAX5*, and other regions. By contrast, the second-passage xenograft showed the identical genotype as the primary patient sample. From this, it can be hypothesized that the predominant clone detected in the first xenograft passage evolved from the predominant clone in the primary patient sample by acquisition of additional genomic deletions. However, the predominating clone in the second xenograft passage must have grown from the minor E2A-PBX1–only clone, which persisted in the first-passage xenograft. The third- and fourth-passage xenografts displayed a third clone, carrying a homozygous deletion of *CDKN2A/B* and additional deletions at other loci. It cannot be determined from these data whether this new clone was an evolved clone from the clone-engrafted second-passage xenograft or an outgrowth of a minor clone already present in the primary patient sample and then carried over in the serial transplantations. However, the data make clear that the analyzed ALL sample consists of a highly complex clonal architecture with multiple clonal populations with distinctly diverging molecular profiles. Interestingly, in this case, the only constant alteration was the unbalanced *E2A-PBX1* translocation, while other important recurrent molecular alterations typical for ALL, such as deletions of *IKZF1*, *PAX5*, and *CDKN2A/B*, occurred only intermittently.

As depicted for ALL25, almost all analyzed series of primary patient samples and their serially transplanted xenografts showed diverse genetic architectures. Similarly to ALL25, all cases displayed CNAs, which were retained in all samples, suggesting that these may be important driver lesions of these cells, while other lesions displayed great variability. Table 4 summarizes this observation, showing that constant lesions consisted of ALL-typical translocations, such as *E2A-PBX1* or *ETV6-RUNX1*, but also recurrent homozygous deletions of *CDKN2A/B*. Interestingly, other commonly affected genes in ALL, such as *IKZF1* and *PAX5*, were found in the variable set of lesions. Detailed CNA visualization in the sets of xenograft passages are shown in Supplementary Figures E1–E26 (online only, available at www.exphem.org).

Table 4. Summary of constant and variable CNAs in serial xenotransplants

Series	Constant CNAs	Variable CNAs
ALL25	unbalanced *E2A-PBX1* translocation	deletions of *IKZF1*, *CDKN2A/B*, *BACH2*, and others
ALL26	unbalanced *ETV6-RUNX1* translocation	deletions of *BTG1*, *CEP135*, *ZCCHC7*, and others; UPD Chr.12p
ALL27	homozygous deletion of *CDKN2A/B*, del5q35	dup2q, del3q, del4p, del5p, del8p, deletion of *IKZF1*, del10q, del17p, and others
ALL28	hyperdiploid	none
ALL29	homozygous deletion of *CDKN2A/B*	deletion of *ETV6*, del3q13.11, dup6p, dup8p22, and dup13q14.2
ALL30	deletion of *PTEN*	duplication of *IKZF1*, dup7p21.3, and others
ALL31a	homozygous deletion of *CDKN2A/B*	dup4p12 and del9q

Del = deletion; *dup* = duplication; *UPD* = uniparental disomy.

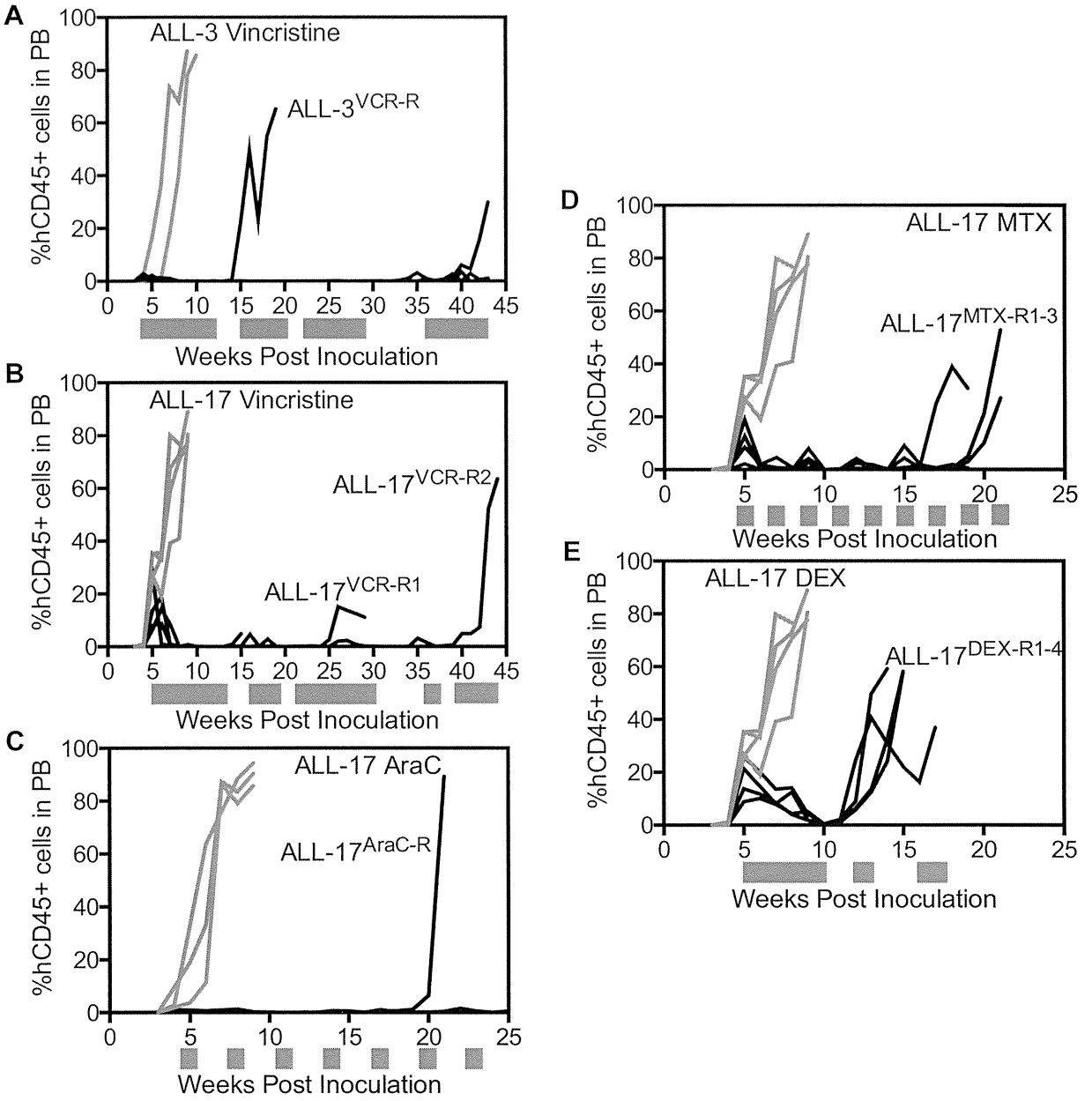

Figure 2. In vivo selection of drug-resistant xenograft sublines. Mice were inoculated with (**A**) second-passage ALL3 or (**B–E**) ALL17 xenograft cells and monitored for engraftment by weekly tail-vein bleeds. When the percentage of cells positive for human CD45 (%huCD45) reached 1%, treatments were initiated with vincristine, cytarabine, methotrexate, or dexamethasone, as detailed in Materials and methods. Mice were culled and spleen-derived cells harvested for subsequent analysis when the %huCD45 increased through the drug treatments or when animals experienced leukemia- or drug-related morbidity. Gray bars indicate drug treatment periods. Gray lines indicate passage-matched control mice; black lines indicate drug-treated mice.

In vivo selection of drug-resistant xenograft sublines and analysis of patterns of in vivo drug resistance

To assess the effects of in vivo selection with established chemotherapeutic drugs on DNA CNAs, drug-resistant sublines were generated by treating mice engrafted with second-passage ALL3 and ALL17 cells with single agent vincristine, cytarabine, methotrexate, or dexamethasone. Figure 2 depicts the emergence of the ALL3 and ALL17 sublines following exposure to single-agent drug treatments. The frequency with which drug-resistant lines emerged remained relatively consistent across all four drugs (mean = 41%, range = 40% to 50%; Supplementary Table E1, online only, available at www.exphem.org). However, drug-resistant sublines emerged at a markedly higher frequency for ALL17 (71%) compared with ALL3 (17%).

To confirm that the emergence of xenograft cells during in vivo drug treatments was due to the acquisition of cellular drug resistance rather than host-derived epiphenomena, spleen-derived cells from ALL3 and ALL17 drug-selected sublines were retransplanted into secondary recipient mice to test their in vivo drug responses. For each of the sublines tested, marked resistance to the selecting drug was observed (Table 5; Supplementary Figure E27; online only, available at www.exphem.org). The LGDs for vincristine and dexamethasone were approximately halved for ALL17-VCR-R1 and ALL17-DEX-R, respectively, and reduced by almost 90% for vincristine in ALL3-VCR-R, compared with their respective passage-matched control sublines. The acquisition of methotrexate resistance by ALL17-MTX-R was more profound, with an almost complete loss of in vivo drug sensitivity. Similarly, ALL17-ARA-C-R cells were profoundly resistant to cytarabine, with an approximately tenfold reduction in LGD compared with passage-matched control cells. These results confirmed that the in vivo emergence of drug-selected sublines is associated with profound levels of resistance to the selecting drug.

We next sought to determine whether the in vivo drug-selected sublines exhibited cross-resistance to established drugs used in the treatment of pediatric ALL. When treated with vincristine, cytarabine, methotrexate, or dexamethasone, the EFS of mice engrafted with ALL17-ARA-C-R, ALL17-MTX-R, and ALL17-DEX-R was not significantly different compared with passage-matched controls, indicating that resistance was only apparent to the selecting drug (Table 5).

Table 5. Median EFS and LGD of ALL17 and ALL3 sublines following in vivo exposure to VCR, ARA-C, MTX, and DEX

Xenograft	EFS (days post treatment initiation)					LGD (days)			
	Saline	VCR	ARA-C	MTX	DEX	VCR	ARA-C	MTX	DEX
ALL17-CTL1	11.0 (11.5,11.0,8.6)	51.8 (57.2,51.8,50.9)	102 (74.7,102,108)	54.9 (53.2,54.9,58.0)	25.9 (24.7,25.9,27.0)	40.8	91.0	43.9	14.9
ALL17-VCR-R1	15.2 (14.7,15.2,15.9)	**30.5*** (30.3,28.3,31.9,30.7)	>70.0 (>70,>70,>70)	62.8 (49.0,62.8,62.9)	18.6 (17.4,18.6,19.9)	**15.3**	>54.8	47.6	3.4
ALL17-ARA-C-R	7.9 (4.7,11.0)	51.7 (51.2,51.7,53.1)	**21.8*** (17.7,21.8,23.7)	58.4 (56.4,60.3)	22.2 (11.4,32.9)	43.8	**13.9**	50.5	14.3
ALL17-MTX-R	9.7 (7.8,9.7,11.3)	53.5 (48.7,53.5,57.9)	93.5 (93.1,93.7,98.4)	**9.2*** (9.1,9.2,19.5)	28.9 (28.3,28.9,28.9)	43.8	83.8	**−0.5**	19.2
ALL17-DEX-R	3.7 (2.4,4.9)	44.8 (41.9,44.8,52.3)	78.8 (77,80.5)	52.2 (44.8,59.6)	**10.4*** (3.2,10.4,10.5)	41.1	75.1	48.5	**6.7**
ALL3-CTL	5.3 (5.3,5.3,5.7)	75.6 (73.5,75.6,84.0)	nd	nd	nd	70.3	nd	nd	nd
ALL3-VCR-R	14.2 (7.7,14.2,19.2)	**22.8*** (20.9,22.8,23.0)	nd	nd	nd	**8.6**	nd	nd	nd

ALL17 = ALL17 xenograft; *ALL3* = ALL3 xenograft; *CTL* = untreated control; *nd* = not done; *R* = resistant.
Numbers in the EFS fields denote the median EFS, and the values below the median indicate the values of the performed experimental replicates for each condition. Median EFS values, which were significantly different from the passage-matched controls, and their corresponding LGD values are highlighted in bold.
*$p < 0.05$ compared with passage-matched control in log rank test.

Single-nucleotide polymorphism array analysis of in vivo drug-selected xenografts reveals genomic lesions involved in chemoresistance

To gain a greater understanding of mechanisms of chemoresistance in pediatric ALL, the in vivo drug-selected sublines were subjected to SNP array analysis and compared with their respective passage-matched controls. A total of seven newly acquired CNAs were identified in ALL3-VCR-R, ALL17-ARA-C-R, and ALL17-MTX-R (Table 6).

The vincristine-selected ALL3-VCR-R subline displayed a newly acquired partial deletion of *CDK6* on chromosome 7q21.2 (Table 6; Supplementary Figure E28A, online only, available at www.exphem.org) and a new trisomy 18 (Table 6; Supplementary Figure E28B, online only, available at www.exphem.org). In ALL17-ARA-C-R, we observed a homozygous deletion of part of the *DCK* gene (Fig. 3A and Table 6). This deletion was detected by three SNP probe sets of the array (Fig. 3B). Fine mapping and measurement of the allelic dosage by quantitative real-time PCR of genomic DNA confirmed a homozygous deletion of exons 3 and 4 of *DCK* (Fig. 3C).

The methotrexate-resistant xenograft subline ALL17-MTX-R displayed enlargement of an already-existing