doi: 10.1038/nature05690

SUPPLEMENTARY MATERIAL FOR:

GENOME-WIDE ANALYSIS OF GENETIC ALTERATIONS IN ACUTE LYMPHOBLASTIC LEUKEMIA

Charles G. Mullighan, Salil Goorha, Ina Radtke, Christopher B. Miller, Elaine Coustan-Smith, James D. Dalton, Kevin Girtman, Susan Mathew, Jing Ma, Stanley B. Pounds, Xiaoping Su, Ching-Hon Pui, Mary V. Relling, William E. Evans, Sheila A. Shurtleff and James R. Downing

LIST OF SUPPLEMENTARY TABLES

LIST OF SUPPLEMENTARY FIGURES

SUPPLEMENTARY METHODS

PATIENTS AND SAMPLES

Two hundred and forty two patients with acute lymphoblastic leukaemia (ALL) treated at St Jude Children's Research Hospital (SJCRH) between 1993 and 2005 were studied. These included precursor-B ALL with high hyperdiploidy (greater than 50 chromosomes on karyotyping, HD>50, n=39), *TCF3-PBX1* positive (*E2A-PBX1*, n=17), *ETV6-RUNX1* positive (*TEL-AML1*, n=47), *BCR-ABL1* positive (n=9), *MLL* rearranged (n=11), low hyperdiploidy (47-50 chromosomes, HD47-50, n=23), hypodiploidy (n=10), B-ALL with pseudodiploidy, near haploid karyotype, normal cytogenetics or non-recurring cytogenetic abnormalities (n=36), and T-lineage ALL (T-ALL, n=50) (Supplementary Table 1).

Informed consent for the use of leukaemic cells for research was obtained from patients, parents or guardians in accordance with the Declaration of Helsinki, and study approval was obtained from the SJCRH institutional review board.

Mononuclear cells were purified from the diagnostic bone marrow or peripheral blood samples by density gradient centrifugation and were cryopreserved in liquid nitrogen. Diagnostic samples were characterized by conventional cytogenetics, reverse transcriptase–polymerase chain reaction (RT-PCR) assays for *BCR-ABL1*, *TCF3-PBX1 [E2A-PBX1]*, and *ETV6-RUNX1 [TEL-AML1]*, and for the presence of *MLL* chimeric fusion genes by at least two of the following methods: cytogenetics, 11q23 fluorescence in situ hybridization (FISH), or RT-PCR for t(9;11)[*MLL-AF9*], t(11;19)[*MLL-ENL*], t(11;19)[*MLL-ELL*], t(10;11)[*MLL-AF10*], or t(4;11)[*MLL-AF4*]. Remission samples were obtained from peripheral blood collected into EDTA vacutainer tubes at a median time of 200 days (range <1-3009) from diagnosis.

DNA was extracted from diagnostic leukaemia samples using the DNA Blood Mini Kit (Qiagen, Valencia, CA). DNA was quantified using either a Nanodrop Spectrophotometer or PicoGreen. Quality was assessed using the Nanodrop and by agarose gel electrophoresis. RNA was extracted from diagnostic samples using TriZOL (Invitrogen, Carlsbad, CA). Cells for FISH analysis were stored in Carnoy's fixative.

CASES EXAMINED BY SNP ARRAY, GENOMIC SEQUENCING, AND METHYLATION ANALYSIS

Supplementary Table 1. ALL cases studied by SNP array analysis

Corresponding Affymetrix gene expression array chip identifiers from previously published studies[39, 40], cases in which *EBF1* sequencing and methylation analysis of *EBF1* and *PAX5* exon 1a and 1b promoter CpG islands were performed, and blast percentage of the diagnostic samples used for DNA extraction for SNP microarray analysis are also listed. The primary SNP microarray data have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE5511. The data are also available at http://www.stjuderesearch.org/data/ALL-SNP1/. "E2A-PBX1" is equivalent to "TCF3-PBX1", and "TEL-AML1" to "ETV6-RUNX1". **EBF1* sequencing was performed in the 8 *EBF1*-deleted cases and an additional 106 B-progenitor B-ALL cases without *EBF1* deletion. ‡Blast percentages are those obtained from diagnostic marrow aspirates; DNA was extracted on samples post density gradient centrifugation, in which blast counts are higher. †Corresponding remission samples not available for these cases. For B-progenitor ALL, 173 of 192 cases had diagnostic blast counts greater than 80%, mean 92.3% (range 39-100%).

Case	Affymetrix HG-U133A chip[39]	Affymetrix U95A chip[40]	EBF1 sequencing*	Methylation analysis	Blast %‡
Hyperdip>50-SNP-#1	JD-ALD485-v5-U133A	00-0712-U95Av2			95
Hyperdip>50-SNP-#2	JD0070-ALL-v5-U133A	17798-U95A			71
Hyperdip>50-SNP-#3	JD-ALD510-v5-U133A	01-0116-U95Av2		Yes	63
Hyperdip>50-SNP-#4	JD0017-ALL-v5-U133A				90
Hyperdip>50-SNP-#5	JD0020-ALL-v5-U133A				77
Hyperdip>50-SNP-#6	JD0023-ALL-v5-U133A			Yes	82
Hyperdip>50-SNP-#7	JD-ALD611-v5-U133A			Yes	93
Hyperdip>50-SNP-#8	JD-ALD612-v5-U133A			Yes	84
Hyperdip>50-SNP-#9	JD0041-ALL-v5-U133A				96
Hyperdip>50-SNP-#10	JD0077-ALL-v5-U133A			Yes	94
Hyperdip>50-SNP-#11	JD0111-ALL-v5-U133A			Yes	98
Hyperdip>50-SNP-#12	JD0097-ALL-v5-U133A			Yes	98
Hyperdip>50-SNP-#13	JD0117-ALL-v5-U133A				93
Hyperdip>50-SNP-#14	JD0120-ALL-v5-U133A		Yes		100
Hyperdip>50-SNP-#15	JD0121-ALL-v5-U133A		Yes	Yes	84
Hyperdip>50-SNP-#16	JD0127-ALL-v5-U133A				77
Hyperdip>50-SNP-#17	JD0151-ALL-v5-U133A				91
Hyperdip>50-SNP-#18	JD0168-B-ALL-v5-U133A				95
Hyperdip>50-SNP-#19	JD0178-ALL-v5-U133A				99
Hyperdip>50-SNP-#20					92
Hyperdip>50-SNP-#21	JD0196-ALL-v5-U133A			Yes	90
Hyperdip>50-SNP-#22	JD0219-ALL-v5-U133A				99
Hyperdip>50-SNP-#23	JD0222-ALL-v5-U133A		Yes	Yes	79
Hyperdip>50-SNP-#24	JD-ALD085-v5-U133A	94-0901-U95A			97

Case	Affymetrix HG-U133A chip[39]	Affymetrix U95A chip[40]	EBF1 sequencing*	Methylation analysis	Blast %‡
Hyperdip>50-SNP-#25†		94-0968-U95A			88
Hyperdip>50-SNP-#26		95-0149-U95A			92
Hyperdip>50-SNP-#27	JD-ALD013-v5-U133A	95-0572-U95A			98
Hyperdip>50-SNP-#28		95-0480-U95A			76
Hyperdip>50-SNP-#29	JD-ALD112-v5-U133A	95-0893-U95A			95
Hyperdip>50-SNP-#30	JD-ALD163-v5-U133A	95-1300-U95A			96
Hyperdip>50-SNP-#31		96-0372-U95A			98
Hyperdip>50-SNP-#32		96-0258-U95A			98
Hyperdip>50-SNP-#33		96-1101-U95A			91
Hyperdip>50-SNP-#34		96-1000-U95Av2		Yes	98
Hyperdip>50-SNP-#35		96-1259-U95A			95
Hyperdip>50-SNP-#36		97-0273B-U95A		Yes	98
Hyperdip>50-SNP-#37		97-0605-U95A			95
Hyperdip>50-SNP-#38		98-0396-U95A			90
Hyperdip>50-SNP-#39		98-0550-U95Av2		Yes	96
E2A-PBX1-SNP-#1	JD0004-ALL-v5-U133A	01-0063-U95Av2			98
E2A-PBX1-SNP-#2	JD0015-ALL-v5-U133A	18464-U95Av2			99
E2A-PBX1-SNP-#3	JD0036-ALL-v5-U133A				94
E2A-PBX1-SNP-#4	JD0042-ALL-v5-U133A			Yes	68
E2A-PBX1-SNP-#5†	JD0083-ALL-v5-U133A			Yes	98
E2A-PBX1-SNP-#6	JD0099-ALL-v5-U133A				97
E2A-PBX1-SNP-#7	JD0104-ALL-v5-U133A		Yes		98
E2A-PBX1-SNP-#8				Yes	95
E2A-PBX1-SNP-#9	JD0203-ALL-v5-U133A				92
E2A-PBX1-SNP-#10	JD-ALD019-v5-U133A	95-0887-U95A	Yes	Yes	98
E2A-PBX1-SNP-#11	JD-ALD025-v5-U133A	95-1205B-U95A		Yes	97
E2A-PBX1-SNP-#12	JD-ALD437-v5-U133A	96-0446-U95A			97
E2A-PBX1-SNP-#13	JD-ALD034-v5-U133A	96-0750-U95A	Yes	Yes	95
E2A-PBX1-SNP-#14	JD-ALD041-v5-U133A	97-0052B-U95A			95
E2A-PBX1-SNP-#15	JD-ALD071-v5-U133A	98-0328-U95A		Yes	98
E2A-PBX1-SNP-#16	JD-ALD073-v5-U133A	98-0515-U95A		Yes	94
E2A-PBX1-SNP-#17	JD-ALD079-v5-U133A	98-1225-U95A			99
TEL-AML1-SNP-#1	JD0002-ALL-v5-U133A	00-0921-U95Av2	Yes	Yes	88
TEL-AML1-SNP-#2	JD0066-ALL-v5-U133A	17650Dial-U95A	Yes		100
TEL-AML1-SNP-#3	JD0056-ALL-v5-U133A	17701-U95A	Yes	Yes	96
TEL-AML1-SNP-#4	JD-ALD493-v5-U133A	00-1177-U95Av2	Yes		91
TEL-AML1-SNP-#5	JD0058-ALL-v5-U133A	17797-U95A	Yes	Yes	53
TEL-AML1-SNP-#6	JD0059-ALL-v5-U133A	17889-U95A	Yes		98
TEL-AML1-SNP-#7	JD0005-ALL-v5-U133A	17948-U95A	Yes		99
TEL-AML1-SNP-#8†	JD0009-ALL-v5-U133A	18054-U95A	Yes		86
TEL-AML1-SNP-#9	JD0033-ALL-v5-U133A	18284-U95A	Yes		91
TEL-AML1-SNP-#10	JD0014-ARD-v5-U133A	18463-U95Av2	Yes	Yes	96
TEL-AML1-SNP-#11	JD0016-ARD-v5-U133A		Yes		98
TEL-AML1-SNP-#12†	JD0018-ALL-v5-U133A		Yes		98
TEL-AML1-SNP-#13	JD0048-ALL-v5-U133A		Yes		76
TEL-AML1-SNP-#14	JD0085-ALL-v5-U133A		Yes		92
TEL-AML1-SNP-#15	JD0101-ALL-v5-U133A		Yes		99

Case	Affymetrix HG-U133A chip[39]	Affymetrix U95A chip[40]	EBF1 sequencing*	Methylation analysis	Blast %‡
TEL-AML1-SNP-#16	JD0118-ALL-v5-U133A		Yes		97
TEL-AML1-SNP-#17	JD0107-ALL-v5-U133A		Yes	Yes	93
TEL-AML1-SNP-#18	JD0109-ALL-v5-U133A		Yes		94
TEL-AML1-SNP-#19	JD0123-ALL-v5-U133A		Yes		99
TEL-AML1-SNP-#20	JD0139-ALL-v5-U133A		Yes		96
TEL-AML1-SNP-#21	JD0149-ALL-v5-U133A		Yes		98
TEL-AML1-SNP-#22			Yes		97
TEL-AML1-SNP-#23			Yes		98
TEL-AML1-SNP-#24	JD0175-ALL-v5-U133A		Yes	Yes	66
TEL-AML1-SNP-#25	JD0193-ALL-v5-U133A		Yes		93
TEL-AML1-SNP-#26	JD0201-ALL-v5-U133A		Yes		84
TEL-AML1-SNP-#27			Yes	Yes	97
TEL-AML1-SNP-#28	JD0212-ALL-v5-U133A		Yes	Yes	98
TEL-AML1-SNP-#29	JD0221-ALL-v5-U133A		Yes		79
TEL-AML1-SNP-#30		94-0746-U95A	Yes		83
TEL-AML1-SNP-#31	JD-ALD004-v5-U133A	94-1106-U95A	Yes		95
TEL-AML1-SNP-#32	JD-ALD005-v5-U133A	94-1116-U95A	Yes	Yes	98
TEL-AML1-SNP-#33	JD-ALD006-v5-U133A	94-1118-U95A	Yes		99
TEL-AML1-SNP-#34	JD-ALD096-v5-U133A	95-0225-U95A	Yes		95
TEL-AML1-SNP-#35		95-0318-U95A	Yes	Yes	93
TEL-AML1-SNP-#36	JD-ALD108-v5-U133A	95-0709-U95A	Yes		94
TEL-AML1-SNP-#37	JD-ALD109-v5-U133A	95-0724-U95A	Yes		92
TEL-AML1-SNP-#38		95-1008-U95A	Yes	Yes	97
TEL-AML1-SNP-#39		95-1149-U95Av2	Yes		96
TEL-AML1-SNP-#40		96-0482-U95A	Yes		95
TEL-AML1-SNP-#41		96-0313-U95A	Yes		98
TEL-AML1-SNP-#42		96-1341B-U95Av2	Yes		84
TEL-AML1-SNP-#43		97-0127A-U95A	Yes	Yes	99
TEL-AML1-SNP-#44	JD-ALD054-v5-U133A	97-0984-U95A	Yes		99
TEL-AML1-SNP-#45		97-0497-U95A	Yes		94
TEL-AML1-SNP-#46		98-0746-U95Av2	Yes		88
TEL-AML1-SNP-#47		98-1103-U95A	Yes		99
MLL-SNP-#1	JD0080-ALL-v5-U133A			Yes	97
MLL-SNP-#2	JD0084-ALL-v5-U133A			Yes	95
MLL-SNP-#3				Yes	100
MLL-SNP-#4	JD0124-ALL-v5-U133A			Yes	94
MLL-SNP-#5	JD-ALD009-v5-U133A	95-0213-U95A		Yes	95
MLL-SNP-#6	JD-ALD433-v5-U133A	95-0892-U95A			98
MLL-SNP-#7	JD-ALD180-v5-U133A	96-1038-U95A	Yes	Yes	91
MLL-SNP-#8	JD-ALD057-v5-U133A	97-1174b-U95A	Yes		99
MLL-SNP-#9	JD-ALD052-v5-U133A	97-0861B-U95A	Yes	Yes	95
MLL-SNP-#10	JD-ALD294-v5-U133A	97-1663-U95A			97
MLL-SNP-#11	JD-ALD078-v5-U133A	98-0780-U95Av2			99
BCR-ABL-SNP-#1	JD-ALD494-v5-U133A	00-1178-U95Av2	Yes		79
BCR-ABL-SNP-#2	JD-ALD613-v5-U133A		Yes	Yes	93
BCR-ABL-SNP-#3	JD0102-ALL-v5-U133A		Yes		77

Case	Affymetrix HG-U133A chip[39]	Affymetrix U95A chip[40]	EBF1 sequencing*	Methylation analysis	Blast %‡
BCR-ABL-SNP-#4	JD0129-ALL-v5-U133A		Yes	Yes	95
BCR-ABL-SNP-#5	JD0154-ALL-v5-U133A		Yes	Yes	78
BCR-ABL-SNP-#6	JD0192-ALL-v5-U133A		Yes		94
BCR-ABL-SNP-#7†	JD0206-ALL-v5-U133A		Yes	Yes	79
BCR-ABL-SNP-#8	JD-ALD008-v5-U133A	95-0184-U95A	Yes	Yes	92
BCR-ABL-SNP-#9	JD-ALD035-v5-U133A	96-1324-U95A	Yes		95
Hyperdip47-50-SNP-#1	JD0064-ALL-v5-U133A	17462B-U95A		Yes	94
Hyperdip47-50-SNP-#2	JD-ALD509-v5-U133A	00-1196-U95Av2	Yes		94
Hyperdip47-50-SNP-#3	JD0062-ALL-v5-U133A	18056-U95A	Yes		94
Hyperdip47-50-SNP-#4	JD-ALD554-v5-U133A	18333-U95Av2	Yes		98
Hyperdip47-50-SNP-#5	JD0098-ALL-v5-U133A		Yes	Yes	97
Hyperdip47-50-SNP-#6	JD0112-ALL-v5-U133A		Yes	Yes	91
Hyperdip47-50-SNP-#7	JD0108-ALL-v5-U133A		Yes	Yes	98
Hyperdip47-50-SNP-#8	JD0132-ALL-v5-U133A		Yes	Yes	95
Hyperdip47-50-SNP-#9	JD0133-ALL-v5-U133A		Yes	Yes	97
Hyperdip47-50-SNP-#10†	JD0137-ALL-v5-U133A		Yes	Yes	96
Hyperdip47-50-SNP-#11	JD0138-ALL-v5-U133A		Yes	Yes	85
Hyperdip47-50-SNP-#12	JD0150-ALL-v5-U133A		Yes		97
Hyperdip47-50-SNP-#13	JD0157-ALL-v5-U133A		Yes		93
Hyperdip47-50-SNP-#14	JD0181-ALL-v5-U133A		Yes		94
Hyperdip47-50-SNP-#15			Yes		88
Hyperdip47-50-SNP-#16		95-0431-U95A	Yes	Yes	94
Hyperdip47-50-SNP-#17		96-0379-U95A	Yes		99
Hyperdip47-50-SNP-#18		96-1566-U95A	Yes		94
Hyperdip47-50-SNP-#19		97-0684-U95A	Yes		97
Hyperdip47-50-SNP-#20		97-1715-U95Av2	Yes		99
Hyperdip47-50-SNP-#21		97-1238-U95A	Yes	Yes	93
Hyperdip47-50-SNP-#22		98-0250-U95A	Yes	Yes	94
Hyperdip47-50-SNP-#23		98-0678C-U95Av2	Yes		97
Hypodip-SNP-#1	JD0057-ALL-v5-U133A	17790-U95A	Yes		94
Hypodip-SNP-#2	JD-ALD536-v5-U133A	01-0564-U95A	Yes	Yes	96
Hypodip-SNP-#3	JD0025-ALL-v5-U133A		Yes		88
Hypodip-SNP-#4	JD0037-ALL-v5-U133A		Yes	Yes	98
Hypodip-SNP-#5	JD0087-ALL-v5-U133A		Yes	Yes	93
Hypodip-SNP-#6	JD0095-ALL-v5-U133A		Yes	Yes	94
Hypodip-SNP-#7		95-1137-U95A	Yes	Yes	93
Hypodip-SNP-#8		96-0251-U95Av2	Yes	Yes	91
Hypodip-SNP-#9	JD-ALD196-v5-U133A	97-0530-U95A	Yes		99
Hypodip-SNP-#10		98-0179-U95A	Yes	Yes	98
Other-SNP-#1	JD0065-ALL-v5-U133A		Yes		42
Other-SNP-#2†	JD0116-ALL-v5-U133A		Yes	Yes	85
Other-SNP-#3	JD0122-ALL-v5-U133A		Yes	Yes	99
Other-SNP-#4	JD0131-ALL-v5-U133A		Yes		97
Other-SNP-#5	JD0166-ALL-v5-U133A		Yes		57
Other-SNP-#6	JD0202-ALL-v5-U133A				62
Other-SNP-#7	JD0226-ALL-v5-U133A		Yes		93

Case	Affymetrix HG-U133A chip[39]	Affymetrix U95A chip[40]	EBF1 sequencing*	Methylation analysis	Blast %‡
Other-SNP-#8†	JD-ALD340-v5-U133A	93-0483-U95A	Yes		98
Other-SNP-#9†	JD-ALD363-v5-U133A	94-0306-U95A	Yes		98
Other-SNP-#10		95-0546-U95A	Yes	Yes	93
Other-SNP-#11		96-0272-U95Av2	Yes	Yes	96
Other-SNP-#12	JD-ALD279-v5-U133A	96-0739-U95Av2			94
Other-SNP-#13		96-0818-U95A	Yes	Yes	97
Other-SNP-#14	JD-ALD194-v5-U133A	97-0468-U95A	Yes	Yes	98
Other-SNP-#15	JD-ALD066-v5-U133A	97-1527-U95A	Yes	Yes	100
Other-SNP-#16		98-0303-U95A	Yes	Yes	84
Pseudodip-SNP-#1	JD0001-ALL-v5-U133A	00-0586-U95Av2			95
Pseudodip-SNP-#2	JD0071-ALL-v5-U133A	17818-U95A		Yes	100
Pseudodip-SNP-#3	JD0012-ALL-v5-U133A	01-0670-U95Av2			97
Pseudodip-SNP-#4	JD0032-ALL-v5-U133A	18249A-U95A			90
Pseudodip-SNP-#5	JD0021-ALL-v5-U133A		Yes		39
Pseudodip-SNP-#6	JD-ALD610-v5-U133A		Yes	Yes	98
Pseudodip-SNP-#7	JD0103-ALL-v5-U133A				98
Pseudodip-SNP-#8					95
Pseudodip-SNP-#9	JD0173-ALL-v5-U133A			Yes	98
Pseudodip-SNP-#10	JD0185B-ALL-v5-U133A				99
Pseudodip-SNP-#11	JD0188-ALL-v5-U133A				92
Pseudodip-SNP-#12	JD0225-ALL-v5-U133A				96
Pseudodip-SNP-#13		95-0119-U95A		Yes	97
Pseudodip-SNP-#14		95-0142-U95A		Yes	93
Pseudodip-SNP-#15		95-1230-U95A			95
Pseudodip-SNP-#16	JD-ALD164-v5-U133A	95-1354-U95A			90
Pseudodip-SNP-#17		97-0808-U95A			97
Pseudodip-SNP-#18		97-0921-U95A			92
Pseudodip-SNP-#19		97-1913-U95A			95
Pseudodip-SNP-#20		97-1523-U95A		Yes	100
T-ALL-SNP-#1	JD0067-ALL-v5-U133A	17679-U95A			94
T-ALL-SNP-#2	JD0055-ALL-v5-U133A	17695-U95A			98
T-ALL-SNP-#3	JD0006-ALL-v5-U133A				50
T-ALL-SNP-#4	JD0007-ALL-v5-U133A				93
T-ALL-SNP-#5	JD-ALD535-v5-U133A	18098A-U95A			91
T-ALL-SNP-#6	JD-ALD542-v5-U133A	18175-U95A		Yes	95
T-ALL-SNP-#7	JD0013-ALL-v5-U133A	18383-U95Av2		Yes	87
T-ALL-SNP-#8	JD0022-ALL-v5-U133A			Yes	85
T-ALL-SNP-#9	JD0024-ALL-v5-U133A				23
T-ALL-SNP-#10†	JD0052-ALL-v5-U133A			Yes	63
T-ALL-SNP-#11	JD0046-ALL-v5-U133A			Yes	88
T-ALL-SNP-#12	JD0045-ALL-v5-U133A			Yes	87
T-ALL-SNP-#13	JD0038-ALL-v5-U133A				83
T-ALL-SNP-#14	JD0040-ALL-v5-U133A			Yes	94
T-ALL-SNP-#15	JD0078-ALL-v5-U133A			Yes	100
T-ALL-SNP-#16	JD0079-ALL-v5-U133A			Yes	96
T-ALL-SNP-#17	JD0096-ALL-v5-U133A				91
T-ALL-SNP-#18	JD0114-ALL-v5-U133A			Yes	66

Case	Affymetrix HG-U133A chip[39]	Affymetrix U95A chip[40]	EBF1 sequencing*	Methylation analysis	Blast %‡
T-ALL-SNP-#19	JD0115-ALL-v5-U133A				95
T-ALL-SNP-#20	JD0105-ALL-v5-U133A			Yes	95
T-ALL-SNP-#21	JD0119-ALL-v5-U133A				95
T-ALL-SNP-#22	JD0134-ALL-v5-U133A			Yes	98
T-ALL-SNP-#23	JD0152-ALL-v5-U133A			Yes	87
T-ALL-SNP-#24	JD0155-ALL-v5-U133A				94
T-ALL-SNP-#25	JD0169-ALL-v5-U133A				90
T-ALL-SNP-#26	JD0194-ALL-v5-U133A				85
T-ALL-SNP-#27	JD0207-ALL-v5-U133A			Yes	85
T-ALL-SNP-#28†	JD0214-ALL-v5-U133A			Yes	98
T-ALL-SNP-#29†		94-0792-U95A			86
T-ALL-SNP-#30		94-1039-U95A			83
T-ALL-SNP-#31		94-1078-U95A			94
T-ALL-SNP-#32		95-0580A-U95A			82
T-ALL-SNP-#33		95-0923B-U95Av2			98
T-ALL-SNP-#34	JD-ALD111-v5-U133A	95-0789-U95A			92
T-ALL-SNP-#35	JD-ALD167-v5-U133A	96-0129-U95A			92
T-ALL-SNP-#36		96-0627-U95A			100
T-ALL-SNP-#37	JD-ALD436-v5-U133A	96-0427-U95A			94
T-ALL-SNP-#38		96-1200-U95A		Yes	98
T-ALL-SNP-#39		96-1503-U95A			95
T-ALL-SNP-#40		97-0051-U95Av2			95
T-ALL-SNP-#41		97-0130-U95Av2		Yes	95
T-ALL-SNP-#42		97-0355-U95Av2			92
T-ALL-SNP-#43	JD-ALD047-v5-U133A	97-0485b-U95A			99
T-ALL-SNP-#44	JD-ALD049-v5-U133A	97-0576-U95A		Yes	98
T-ALL-SNP-#45		97-1213-U95A			81
T-ALL-SNP-#46	JD-ALD207-v5-U133A	97-1562-U95A			98
T-ALL-SNP-#47		98-0820-U95A			89
T-ALL-SNP-#48		98-0914B-U95Av2			90
T-ALL-SNP-#49		98-0661-U95A		Yes	76
T-ALL-SNP-#50		98-0772-U95A			78

AFFYMETRIX MAPPING 100K AND 500K SINGLE NUCLEOTIDE POLYMORPHISM ARRAYS

Samples were genotyped with Affymetrix 50K GeneChip Human Mapping 50K Hind 240, 50K Xba 240 and 250K Sty arrays (Affymetrix, Santa Clara, CA). DNA was restriction enzyme digested, PCR-amplified, purified, labeled, fragmented and hybridized to the arrays according to the manufacturer's instructions. Briefly, 250ng of DNA was digested with *Xba*I, *Hind*III or *Sty*I (New England Biolabs, Boston, MA). Digested DNA was adaptor-ligated and PCR-amplified using AmpliTaq Gold (Applied Biosystems, Foster City, CA) in four 100µl PCR reactions for each enzyme-digested sample. PCR products from each set of four reactions were pooled, concentrated and fragmented using DNase I. Fragmented PCR products were then labeled, denatured and hybridized to the arrays. Arrays were then washed using Affymetrix fluidics stations, and scanned using the Gene Chip Scanner 3000. CEL files were generated using either Affymetrix GeneChip Operating Software v 3.0 or Affymetrix GeneChip Genotyping Analysis Software (GTYPE) v 4.0. SNP calls were generated using GTYPE. Array SNP call rates are summarized in Supplementary Table 2. Affymetrix CEL and GTYPE-generated SNP call text files for the 242 blast samples, corresponding remission samples in those cases with *PAX5* or *EBF1* mutations, and the 62 remission samples used as diploid reference samples for copy number analysis have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE5511. The data are also available at http://www.stjuderesearch.org/data/ALL-SNP1/.

Supplementary Table 2. SNP array quality control data

	Median SNP call % (range)
Blast 50k Hind 240	94.13 (80.56-98.43)
Blast 50k Xba 240	97.55 (85.06-99.46)
Blast 250k Sty	92.30 (74.04-98.87)

ANALYSIS OF AFFYMETRIX SNP ARRAY DATA.

SNP array data was analyzed using two methods: (1) dChipSNP[41, 42] (www.dchip.org) and circular binary segmentation (DNA Copy)[43].

1. Analysis using dChipSNP

1.1 Data processing

Copy number analyses of 100K (50K Hind, 50K Xba) and 250K Sty SNP array data were performed using dChipSNP. Affymetrix CEL files and corresponding SNP genotype call files generated by Affymetrix GTYPE v 4.0 were read in by dChipSNP. In our initial analyses, we implemented the standard dChipSNP normalization and model-based expression algorithms. Probe intensity data for each array were normalized to a baseline array with median signal intensity using the "invariant set" model [44]. Model-based expression was performed using the perfect-match/mismatch (PM/MM) model to summarize signal intensities for each probe set. For copy number inference, raw copy number was calculated by comparing the signal intensity of each SNP probe set for each tumor sample against a diploid reference set of samples obtained from peripheral blood or bone marrow samples of acute leukaemia cases in remission. Raw copy number data is inherently noisy, and we observed that using reference samples run in the same laboratory and in the same batches as the tumour samples markedly reduced this noise. Consequently, we used a set of 62 leukaemia remission samples as a reference set for copy number estimation for all samples. To perform LOH analysis, and to exclude inherited (germline) copy number variants as the basis of copy number changes detected in the tumour samples, we also used a reference set of matched remission samples for 228 of the ALL cases studies. Remission samples were not available for the remaining 14 cases. Raw and median-smoothed copy number was examined. Hidden Markov Model (HMM) estimates of copy number change were not used as they were insensitive to very focal lesions. For ALL cases with corresponding remission samples, loss-of-heterozygosity was analysed using a Hidden Markov Model in dChipSNP and an inferred LOH call threshold of 0.5.

1.2 Initial analyses

Our initial analyses suggested that copy number of grossly aneuploid samples was not correctly estimated following the standard invariant-set normalization approach in dChipSNP. This was most evident for duplicated chromosomes in samples with high hyperdiploidy and deleted chromosomes in samples with near haploid karyotype. As a fixed amount of genomic DNA is processed and hybridized to arrays for each sample, the absolute mass of DNA from diploid chromosomes in an aneuploid sample will not be the same as the mass of DNA from the same chromosomes in a sample with no numerical chromosomal abnormalities. As dChipSNP may use probe sets from non-diploid regions to perform invariant set normalization, we hypothesized that this may adversely influence the normalization process, and result in false estimates of copy number in these aneuploid samples.

1.3 Karyotype-guided normalization ("Cytonormalization")

We developed a karyotype-guided normalization method to improve the accuracy of copy number estimation by dChipSNP. For each sample, we identified probes located on non-sex chromosomes with no cytogenetically detected aberrations as an initial internal reference set. We used dChip's PM/MM model to compute unnormalized summary signals for all probes. Next, for each array, we computed the unitized-ranks of the summary signals among all probe sets. We then identified unitized-ranks that met the box plot definition of outliers among the internal reference probes. Then, the internal reference set was refined by excluding any probes with an outlying unitized-rank. This step was included to provide additional assurance that most signals in the refined reference set were representative of two copies by excluding any signals that might correspond to small lesions. We then computed the unitized-ranks across the refined reference set. Thus, for the refined reference set, we had two sets of unitized ranks: one computed across the entire array and one computed within the refined reference set. We then plotted the within-set unitized-ranks against the across-array unitized-ranks for the refined reference sets. We then fit a linear interpolation through these points. The linear interpolation then defined a mapping of across-array unitized-ranks to within-reference unitized ranks. We applied this mapping to the across-array unitized-ranks of all probes. The transformed signals of the refined reference set were thus equally spaced between 0 and 1 and the signals of other probes were mapped to values in the unit-interval in a manner that maintained their initial ordering. Finally, the transformed signals were then mapped to a log-normal distribution with $\mu = 5$ and $\sigma = 1$. These transformed signals were then imported as normalized data into dChipSNP for inference of copy number. This normalization algorithm was applied to each of the arrays (50K Hind, 50K Xba, and 250K Sty) individually. Cytonormalized data was then combined for downstream analysis in dChipSNP. The method was implemented using S-PLUS software, Windows version 6.2 (Insightful Corp., Seattle, WA).

2. Analysis using circular binary segmentation (DNAcopy)

The circular binary segmentation (CBS) algorithm uses array-CGH data to split a chromosome into contiguous regions of equal copy number, and uses a permutation reference distribution to define copy number abnormalities at a defined significance level [43]. CBS is based on the model that gains or losses of DNA are discrete and occur in contiguous regions of chromosomes covering multiple markers up to whole chromosomes, and that array-CGH data is noisy with some markers not reflecting true copy numbers in the sample. CBS has been applied to array-CGH and representational oligonucleotide microarray analysis (ROMA) data[43], and compares favorably to alternative algorithms in sensitivity and specificity of detecting copy number abnormalities in BAC array-CGH data[45]. CBS also provides a comprehensive list of copy number abnormalities for each sample, and thus may detect lesions that may be missed by visual inspection of high-density data.

To our knowledge, this is the first reported application of CBS to SNP array data. For each autosomal SNP in each array for the 242 blast samples, we calculated a $\log_2$ ratio of blast signal/median signal of all 62 reference samples using the combined 350K cytonormalized signal intensity values. We then applied the CBS algorithm to the above

$\log_2$ ratio data to identify copy number alterations for each sample. This algorithm recursively splits chromosomes into sub-segments based on a maximum *t* statistic. To decide whether or not to split at each stage, we did 10,000 permutations to obtain the reference distribution to estimate the significance of the maximum *t* statistic. The segmentations were performed using a *p*-value threshold of 0.01 (significance level $\alpha = 0.01$). We used an R[46] version of the CBS algorithm implemented in the "DNAcopy" package of Bioconductor[47]. We then used the following criteria to obtain candidate genomic lesions (gain or loss): (1) mean $\log_2$ ratios of the segment ≥ 0.2 or ≤ -0.2; (2) more than 2 SNPs within a segment; (3) segment size > 0 (this was to exclude the cases where one genomic location may have more than one SNPs due to using combined data from multiple sub-arrays).

3. Generation of lesion lists and lesion frequencies in B-progenitor and T-lineage ALL

Results of the CBS algorithm for combined 350K array data were compared to those of dChipSNP. All lesions previously identified by dChipSNP were also identified by the CBS algorithm, except two cases of focal hemizygous *PAX5* deletion involving less than 10 SNPs, which were subsequently verified by genomic real-time PCR. To exclude calls of genomic gain or loss arising from inherited genomic copy number variations (CNVs), the dChipSNP and CBS algorithms were also applied to 228 samples with corresponding germline DNA samples, using the paired normal sample as the copy number reference for each tumor sample. Copy number variations identified from this analysis were excluded. For the remaining 14 samples lacking paired remission samples, apparent gains or losses that were also identified in the pool of reference samples were assumed to be inherited CNVs and were excluded. Final lesion listings and tallies exclude DNA gains and losses arising from B- and T- cell antigen receptor gene rearrangements at 2p11.2 (*IGKL*), 7p14.1 (*TRGV*), 7q34 (*TRBV*), 14q11.2 (*TRAV*, *TRDV*, *TRDJ*, *TRDC*, and *TRAJ*), 14q32.33 (*IGHV*) and 22q11.22 (*IGLL*).

4. Detection of abnormalities in genes encoding regulators of B cell development

Using published data and pathway analysis software (Ingenuity, Ingenuity Systems, Redwood City, CA), we catalogued genes with established roles in normal B cell development and interrogated our dataset for evidence of genomic copy number change (Supplementary Table 12). We also examined targets of these genes, particularly components of the pre-B and B cell antigen receptor complexes. SNP coverage is not uniform across the genome, and for a number of the genes the coverage is insufficient to allow the accurate assessment of copy number changes using the SNP microarray (Supplementary Table 12, Supplementary Figure 11).

FLUORESCENCE IN-SITU HYBRIDIZATION (FISH) ANALYSIS

Dual-color FISH was performed on archived bone marrow cells obtained at presentation, treated with Carnoy's fixative and dried onto slides. Probes were derived from bacterial artificial chromosome (BAC) clones (Children's Hospital Oakland Research Institute, Oakland, CA.; Invitrogen; Open Biosystems, Huntsville, AL). BACs used were RP11-

586B19 (*EBF1* deletion), RP11-160B5 (*EBF1* deletion, case Hypodip-SNP-#5), RP11-96F2 (*PAX5* deletion), CTD-2535J16 (*PAX5* deletion), RP11-614P24 and RP11-1136K1 (*PAX5-ETV6* fusion), RP11-652D9 and RP11-79P21 (*PAX5-FOXP1* fusion), RP11-652D9 and RP11-73L12 (*PAX5-ZNF521* fusion) RP11-652D9 (*PAX5* amplification in case Hypodip-SNP-#7), CTC-600N23 (*IKZF3* deletion). BACs labeled for control probes were CTD-2194L12 (5p13.2), RP11-235C23 (9q31.2) and RP11-4F24 (17p13.3).

BAC clone identity was verified by T7 and SP6 BAC-end sequencing and by hybridization of fluorescently labeled BAC DNA with normal human metaphase preparations. BACs were labeled with either fluorescein isothiocyante or rhodamine fluorochromes. Target probes were paired with control probes from the opposite chromosomal arm where possible. All probe mixtures were diluted 1:50 in DenHyb buffer (Insitus Biotechnologies, Albuquerque, NM) and co-denatured with the target cells on a hotplate at 90°C for 1 minute. The slides were incubated overnight at 37°C on a slide moat and then washed in 50% formamide/1xSSC at 25°C for 5 minutes. Nuclei and metaphases were counterstained with DAPI (200ng/ml) (Insitus Bio.) for viewing on either an Olympus BX60 or a Nikon Eclipse E800 fluorescence microscope equipped with a 100 watt mercury lamp; FITC, Rhodamine, and DAPI filters; a 100X PlanApo (1.40) oil objective; and a COHU CCD or Photometrics SenSys camera. Images were captured and processed with an exposure time ranging from 0.5-2 seconds for each fluorochrome using Cytovision v3.6 software from Applied Imaging (San Jose, CA). Images were captured and enhanced using Applied Imaging's MacProbe v4.3 software.

FLUORESCENCE ACTIVATED CELL SORTING

Sorting of leukaemic blasts according to level of CD10 expression was performed using a BD FACSAria Cell-Sorting System and a phycoerythrin-labeled anti-CD10 antibody (BD Pharmingen, BD, Franklin Lakes, NJ).

GENOMIC SEQUENCING

All 16 exons of *EBF1* were sequenced using genomic DNA from 8 blast samples with *EBF1* deletions and an additional 106 B-progenitor ALL cases without *EBF1* deletions (Supplementary Table 1). DNA was amplified using Accuprime *Pfx* DNA polymerase (Invitrogen) or Accuprime GC-rich DNA polymerase (Invitrogen). Exon sequences were determined aligning the reference *EBF1* mRNA and DNA sequences (Genbank accessions NM_024007.2 and NC_000005.8). Primers for *EBF1* sequencing are listed in Supplementary Table 3. The coding regions of all 11 exons of *PAX5* (exons 1A, 1B, and 2-10) and the promoter region of exon 1B were sequenced in all 242 blast samples. Remission samples for patients with blast samples harboring *PAX5* mutations were also sequenced. *PAX5* PCR and sequencing primers are listed in Supplementary Table 4, and primers for *IKZF1* (Ikaros) in Supplementary Table 5. Primers were designed using Primer 3 [48]. PCR amplification was performed according to the manufacturer's instructions using Eppendorf Mastercyclers (Eppendorf North America, Westbury, NY). Thermal cycling parameters for Accuprime *Pfx* DNA polymerase were 95°C for 2 minutes followed by 35 cycles of 95°C for 15 seconds, 60°C (or annealing temperature as

indicated in Supplementary Tables 3-5) for 30 seconds, and 68°C for 60 seconds. Thermal cycling parameters for Accuprime GC-rich DNA polymerase were 95°C for 3 minutes followed by 35 cycles of 95°C for 30 seconds, 60°C (or annealing temperature as indicated in Supplementary Tables 3-5) for 30 seconds, and 72°C for 60 seconds, with a final extension of 10 minutes at 72°C. PCR products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI). PCR products were sequenced directly with primers indicated in Supplementary Tables 3-5, using Big Dye Terminator (v.3.1) chemistry on 3730xl DNA Analyzers (Applied Biosystems, Foster City, CA). Mutations detected by direct sequencing were confirmed by cloning PCR products into either pCR2.1-TOPO (Invitrogen) or pGEM-T Easy (Promega, Madison, WI) vectors, and sequencing multiple clones in both directions using M13 primers.

Supplementary Table 3. Primers used for *EBF1* sequencing

PCR reactions used Accuprime Pfx DNA polymerase with annealing temperature of 60°C unless otherwise indicated. Sequencing primers are the same as PCR primers unless otherwise indicated. *Accuprime GC-rich DNA polymerase. †Thermal cycling annealing temperature 56°C.

Exon	Primer	Sequence (5'-3')	Sequencing primer	Sequence (5'-3')
1	EBF 1_altF*	gggggaggagattttccac		
	EBF 1_altR*	gctcctcgcagacagctc		
2	EBF 2AF*	atgttttgggggaagaggag		
	EBF 2AR*	ggacccccacatagaagtgt		
3	EBF 3AF	aggccctggaaagaggatag		
	EBF 3AR	gataaggctcttgggccact		
4	EBF 4AF	agccccttcgtgatatgtg		
	EBF 4AR	cccgctggcttttagagtta		
5	EBF 5AF	gagatgtgcatcatggcaag	EBF 5BF	tgctactcctttccattcttagc
	EBF 5AR	tgatgcctgacactttactacca	EBF 5BR	ttgctaattaacgggagagacc
6	EBF 6AF	ccacctccttctttgagcag		
	EBF 6AR	agccacatgcattcctaacc		
7	EBF 7AF	tgccaacagtgtttgtgtca		
	EBF 7AR	cctcagctgctcttcacctc		
8	EBF 8AF	atgccaaagacccactcatc		
	EBF 8AR	gtggtcccacaagaaaggtg		
9	EBF 9AF	gtcctaatgtggctgggaaa	EBF 9BF	tggttctaacagcagactccttc
	EBF 9AR	atgggccaaagaaaggaagt		
10	EBF 10AF	gccatgtaggaaaacccaga		
	EBF 10AR	gccattcaacatggacacct		
11	EBF 11AF	cccaccgtggtataaagcag		
	EBF 11AR	cagtgttgcctggcacataa		
12	EBF 12AF	tctcgcagtaccaatgatgc		
	EBF 12AR	gattcagtggaaagggcaat		
13	EBF 13AF	tgcagctgaccattgaaaac		
	EBF 13AR	tgcatcctttctagccctct		
14	EBF 14AF*	cacagagctagcaccaccaa		
	EBF 14AR*	tgccttacaggagggaaaga		
15	EBF 15AF	aaggtttgtgctggaagcat		
	EBF 15AR	aaggctgataccctcagcaa		
16	EBF 16AF†	cattccagagaagatgagagca	EBF 16BF	tgaacgcaaagaaaccattg
	EBF 16AR†	tcccttgtatagagctttacgg	EBF 16BR	ttgggaggtacaactttaaccaa

Supplementary Table 4. Primer used for *PAX5* sequencing

PCR reactions used Accuprime Pfx DNA polymerase with annealing temperature of 60°C unless otherwise indicated. Sequencing primers are the same as PCR primers unless otherwise indicated. *Accuprime GC-rich DNA polymerase. †Thermal cycling annealing temperature 58°C.

Exon	Primer	Primer sequence (5'-3')	Sequencing primer	Sequence (5'-3')
1A	PAX5 1cF*	gtctgccccttcccgtag		
	PAX5 1cR*	cctcctcctccagggtca		
1B	PAX5 e1B 1F*	gagcgtgattggcaggttag		
	PAX5 e1B 1R*	cgaagttgcaaagaacttcctc		
1B	PAX5 e1B 2F*	gcagcgggtctcagtgtt		
	PAX5 e1B 2R*	aggcgggaaatggtgcta		
1B	PAX5 e1B 3F*	ctcaaagctgctccttcctg		
	PAX5 e1B 3R*	tccttccggccttagtacct		
1B	PAX5 e1B 4F*	tgcatccatgcatagtaagtagg		
	PAX5 e1B 4R*	gctctcaacctcttcctcca		
2	PAX5 2F†	cagcggtgcttctcctatgt		
	PAX5 2R†	gctctgcgtgtgaaacaaaa		
3	PAX5 3bF	ggccagagtagcccgttatt	PAX5 3F	cccgttattttgttgccaat
	PAX5 3bR	cagatcttcaggaaaggcaca		
4	PAX5 4F	ctgtgcatagctggttgagg		
	PAX5 4R	cgtgtgctgaagtgttttatgc		
5	PAX5 5F*	gggtcagtccttctcagtgc		
	PAX5 5R*	actcgctcctctgcaggtaa		
6	PAX5 6bF*	ttggggtcaggtcctcttc		
	PAX5 6bR*	tctctgagcagaacctggtg	PAX5 6R	tctgagcagaacctggtgtg
7	PAX5 7F	agctcagaacgtggagttgg		
	PAX5 7R	caccaagaagccactcttcc		
8	PAX5 8F†	cgtgacaaatgtgcagaagc		
	PAX5 8R†	ttctcagaagcgtagaggtcac		
9	PAX5 9bF†	acagctgcccactccataat	PAX5 9F	actcacggaagaggcaaatg
	PAX5 9bR†	tcctaacccaccaaagcatc	PAX5 9R	acccacctcagtgaccagac
10	PAX5 10F*	gactgagtgaggggaggaaa		
	PAX5 10R*	agtcagacagctggaggacag		

Supplementary Table 5. Primers used for *IKZF1* (Ikaros) sequencing

PCR reactions used Accuprime Pfx DNA polymerase with annealing temperature of 60°C unless otherwise indicated. Sequencing primers are the same as PCR primers unless otherwise indicated. *Accuprime GC-rich DNA polymerase. †Thermal cycling annealing temperature 58°C. The reference genomic sequence of *IKZF1* (Genbank accession NC_000007.12) lacks corresponding sequence for the first 3 exons of the reference mRNA (Genbank accession NM_006060.2). The complete genomic sequence was obtained from the University of California Santa Cruz genome browser. Alignments were performed using the SIM4 module of Vector NTI 10 Advance (Invitrogen), resulting in 8 exons, 7 of which are coding. ‡Exon numbering corresponds to Molnar *et al.* [49] and Sun *et al.* [50]. Exon 0 is untranslated.

Exon‡	Primer	Sequence (5' - 3')	Sequencing primer	Sequence (5' - 3')
0	IKAROS e0 F†	caatgcgagtgagcaacttc		
	IKAROS e0 R†	cgacaccagggtctaccaac		
1	IKAROS e1 cF†	gacccagggccatttaattt	IKAROS e1 F	gccagtctgatactccagca
	IKAROS e1 R†	ccatgagcataccaagcact		
2	IKAROS e2 bF†	actggctccacccagtacct		
	IKAROS e2 bR†	cccatcctgctgatctttgt		
3	IKAROS e3 F	gctctccacacctatttgattg	IKAROS e3dF	ttgctgctgtgttgttttgttgag
	IKAROS e3 cR	aaccaatcgcttgcaacaac		
4	IKAROS e4 bF†	aaggagctggcaggtttagtc		
	IKAROS e4 bR†	ggttagccagcaaggacaca		
5	IKAROS e5 F†	cttggccaccaacgttttta	IKAROS e5cF	ggtaataattgtattgcatgc
	IKAROS e5 R†	ctctgctcctaaggctgcat		
6	IKAROS e6 F†	gcctgtctggaagtgttgct		
	IKAROS e6 R†	ccctttcttccaccctcaac		
7	IKAROS e7 1F*	tccccggttgtagatttcag		
	IKAROS e7 1R*	cgatgtggttggtcaggtag	IKAROS e7 1bR	ctgctcctcgttgttgctct
7	IKAROS e7 2F*	ctgctctccaaggccaagt		
	IKAROS e7 2R*	tccagtccagtctatgctgct	IKAROS e7 2bR	ctggtccagtccagtctatgc

MODELLING OF *PAX5* PAIRED DOMAIN MUTATIONS

A structural view of the PAX5 paired domain was generated by PyMOL v0.99 (http://pymol.sourceforge.net/) using the coordinates of the X-ray structure of PAX5 interacting with ETS1 on DNA, and PAX6 deposited with the Brookhaven Data Bank (PDB: 1K78 and 6pax; http://www.rcsb.org/pdb/)[51, 52]. The protein sequences of the paired domains of PAX6 and PAX5 are 70.1% identical.

CELL CULTURE

The human pre-B ALL cell line Kasumi-2 and REH, the Burkitt lymphoma cell line Raji, and the T-lineage ALL cell lines Jurkat and MOLT-4 (all obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen, DSMZ, Braunschweig, Germany) were grown in RPMI-1640 containing 100 units/ml penicillin, 100 µg/ml streptomycin, 2 mM glutamine and 10% fetal bovine serum (20% for MOLT-4). 293T cells were maintained in DMEM containing 100 units/ml penicillin, 100 µg/ml streptomycin, 2 mM glutamine and 10% fetal bovine serum.

QUANTITATIVE RT-PCR AND GENOMIC PCR

PAX5 gene expression of ALL samples was quantitated using Taqman® Assays-on-demand Hs00277134_m1 (specific for *PAX5* exons 4-5) and Hs00172001_m1 (*PAX5* exons 7-8) (Applied Biosystems, Foster City, CA). RNA was extracted using TriZOL (Invitrogen), and reverse transcribed using random hexamer primers and Superscript III Reverse Transcriptase (Invitrogen). Taqman® assays were performed using a 7500 Real-Time PCR system and 7500 System Software (Applied Biosystems, Foster City, CA), using the 7500 universal cycling conditions: 50°C for 2 minutes, followed by 95°C for 10 minutes, then 40 cycles of 95°C for 1 minute and 60°C for 1 minute.

Standard curves for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and *PAX5* gene expression were generated from 10-fold serial dilutions of the human t(12;21) *[ETV6-RUNX1]* REH cell line (American Type Culture Collection, Manassas, VA), which expressed high levels of *PAX5* as determined by gene expression profiling using Affymetrix U133A chips (Affymetrix, Santa Clara, CA) (data not shown). These normalized ratios were compared to each other for differences in overall levels of *PAX5* expression.

Primers for genomic quantitative PCR were designed using Primer Express 3.0 (Applied Biosystems, Foster City, CA), and are listed in Supplementary Table 6. Taqman® RNase P primers (Applied Biosystems, Foster City, CA) were used for control amplification. 200ng of leukaemic blast DNA or control human DNA was amplified using the same conditions as those described for RNA real-time PCR. Standard curves for each *PAX5* exon and RNase P were generated using normal human DNA. Assays were performed in duplicate. *PAX5* exon-specific copy numbers values were normalized by dividing the value obtained for *PAX5* by the paired value obtained for RNase P for each sample.

Cutoffs of 0.7 and 0.3 were used to identify hemizygous and homozygous *PAX5* deletions, respectively.

Supplementary Table 6. Primers used for *PAX5* genomic real-time PCR

Exon	Primer	Sequence (5'-3')	Amplicon Size (bp)
3	PAX5 e3 Taqman F	ggttcctcatggctaagcttctt	
	PAX5 e3 Taqman R	tcagttccatcaacaggtgagg	243
	PAX5 e3 Taqman probe*	FAM-cgccacacccaaagtggtggaaaaaa-TAMRA	
6	PAX5 e6 Taqman F	tgtctttctttagcaacgtgtataacc	
	PAX5 e6 Taqman R	gtgatgcacgcccacca	263
	PAX5 e6 Taqman probe*	FAM-acggccactcgcttccggg-TAMRA	
8	PAX5 e8 Taqman F	gggcacattgccgttca	
	PAX5 e8 Taqman R	agtttgcactgtcggcgtc	244
	PAX5 e6 Taqman probe*	FAM-ccccgctggacagggcagc-TAMRA	

DETECTION OF PAX5 EXPRESSION IN LEUKAEMIC BLASTS BY FLOW CYTOMETRY

We characterized PAX5 expression in bone marrow mononuclear cells obtained at diagnosis from 16 patients with B lineage ALL. After density gradient separation, cells were labeled with anti-CD34 (of IgG2a class) conjugated to allophycocyanin (APC; Miltenyi Biotech, Auburn, CA) and anti-human CD19 (of IgM class; Research Diagnostics, Concord, MA) followed by goat-anti mouse IgM conjugated to phycoerythrin (PE; Jackson ImmunoResearch Laboratories, West Grove, PA). After cell permeabilisation with 8E, a paraformaldehyde-based reagent developed in our laboratory, cells were labeled with anti-human PAX5 (of IgG1 class; BD Transduction Labs, San Jose, CA) followed by goat anti-mouse IgG1 conjugated to fluorescein isothiocyanate (FITC; Southern Biotechnology Associates, Birmingham, AL). In parallel tests, anti-CD19 was omitted and cells were also labeled with a goat F(ab')2-anti-human IgM PE (Southern Biotechnology Associates, Birmingham, AL) after permeabilisation. Isotype-matched non-reactive antibodies were used as controls. We used the cell lines Raji and Kasumi-2 as positive controls for PAX5 staining; Molt-4 and Jurkat were used as negative controls. We analyzed 20,000 cells for each antibody combination using a FACSCalibur flow cytometer (BD Biosciences) and either CellQuestPro (BD Biosciences) or FlowJo (Treestar Inc, Ashland, OR) software.

To show the B-cell specificity of the PAX5 antibody, we examined PAX5 expression in peripheral blood mononuclear cells from a healthy donor. After density gradient separation, cells were labeled with anti-CD3 (of IgG2a class) conjugated to Pacific Blue (Invitrogen) and goat F(ab')2-anti -human IgM PE. After cell permeabilisation with 8E, cells were labeled with the anti-human PAX5 antibody followed by goat-anti-mouse IgG1 FITC. In a parallel tube, we used an isotype-matched non-reactive control antibody (DakoCytomation, Carpinteria, CA) in place of Pax-5 followed by goat-anti-mouse IgG1

FITC (Southern Biotechnology Associates). We analysed 10,000 cells using an LSR II flow cytometer (BD Biosciences) and CellQuestPro. The PAX5 antibody stained B-lymphocytes, but not T-lymphocytes (Supplementary Figure 19).

To demonstrate the PAX5-specificity of the anti-PAX5 antibody, we synthesized 25 12mer overlapping peptides corresponding to the immunogen used to raise the PAX5 antibody (residues 151-306). Peptides were synthesized using an Aapptec 396 Multiple Organic Synthesizer (Aapptec, Louisville, KY). Peptide purity and quality was assessed by high performance liquid chromatography (HPLC) and matrix assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry. Peptide sequences are available upon request. PAX5, CD79A (Dako Cytomation) or mouse IgG1 antibody were preincubated with 15ng of either the PAX5 peptide pool or a non-PAX5 negative control peptide (corresponding to an internal sequence of SMAC) and then used to stain either peripheral blood mononuclear cells (PBMNCs) obtained from a normal donor or the Raji cell line. Data for PBMNCs are shown in Supplementary Figure 20. The PAX5 peptide pool specifically abolished staining of B lymphocytes by PAX5 (Supplementary Figure 20a) but not CD79A (Supplementary Figure 20c). No blocking was seen with the control peptide (Supplementary Figure 20b,d) and no reactivity with mouse IgG1 antibody was seen (Supplementary Figure 20e,f). These results demonstrate the PAX5 specificity of the anti-PAX5 antibody used for quantitation of leukaemic blast intracellular PAX5 levels.

CLONING OF *PAX5* AND *EBF1* WILD-TYPE AND MUTANT ALLELES

Total RNA was extracted from leukaemia blast samples using TriZOL (Invitrogen) and 1μg of total RNA was reverse transcribed using Superscript III (Invitrogen). The coding region of the exon 1a isoform of *PAX5* was amplified using primers C282 and C302, the exon 1b isoform was amplified using primers C317 and C302, and the entire coding region of EBF1 was amplified using primers C503 and C504 (Supplementary Table 7) using 1μl of cDNA and the Advantage 2 PCR Kit (Clontech, Mountain View, CA). Thermal cycling conditions were 5 cycles of 94°C for 30 sec and 72°C 3 min, followed by 5 cycles of 94°C for 30 sec, 70°C for 30 sec and 72°C 3 min, followed by 25 cycles of 94°C for 30 sec, 68°C for 30 sec and 72°C 3 min. PCR products were sequenced directly and after cloning into either pCR2.1-TOPO (Invitrogen) or pGEM-T Easy (Promega, Madison, WI) vectors.

IDENTIFICATION, DETECTION AND CLONING OF *PAX5* TRANSLOCATIONS

The *PAX5-ETV6 [PAX5-TEL]* translocation was detected using primers PAX5ex3-F1 and ETV6ex3-R1 as previously described[53] (Supplementary Table 7). The coding region of *PAX5-ETV6* mRNA was amplified using primers C282 and C283. 3' rapid amplification of cDNA ends (RACE) was performed using the BD Smart RACE™ amplification kit (Clontech, Mountain View, CA) according to the manufacturer's instructions, using the supplied universal primer mix and gene specific primers C294 (*PAX5* exon 3) and C300 (exon 1a). Nested PCR was performed using the supplied nested universal primer, and

nested gene-specific primer C295. PCR products were gel-purified and sequenced directly and after cloning into pCR2.1-TOPO (Invitrogen). Following identification of *PAX5-FOXP1* and *PAX5-ZNF521 [PAX-EVI3]* fusions by RACE, translocation-specific RT-PCR was performed using primers C303 and C304 for *PAX5-FOXP1*, and C303 and C309 for *PAX5-ZNF521*. The coding region of the *PAX5-FOXP1* mRNA was amplified using primers C334 and C335, and *PAX5-ZNF521* with primers C326 and C327, using Phusion™ High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) and the following thermal cycling parameters: 98°C for 60 seconds, followed by 35 cycles of 98°C for 10 seconds, 72°C for 90 sec (*PAX5-FOXP1*) or 3 minutes (*PAX5-ZNF521)* followed by a final extension step of 72°C for 10 min.

Supplementary Table 7. Sequences of primers used for RT-PCR and cloning

FOXP1 exon nomenclature was derived by aligning the reference *FOXP1* mRNA sequence (Genbank accession NM_032682) against the corresponding genomic sequence (Genbank accession NC_000003.10). *ZNF521* exon nomenclature was derived by aligning the reference *ZNF521* mRNA sequence (Genbank accession NM_015461.1) against the corresponding genomic sequence (NC_000018.1). Alignments were performed using the SIM4 module of Vector NTI 10 Advance (Invitrogen).

Primer	Specificity	Sequence (5' – 3')
PAX5ex3-F1	*PAX5* exon 3	ccatgtttgcctgggagatcag
ETV6ex3-R1	*ETV6* exon 3	cctctttggtcagcagcaggag
C282	*PAX5* exon 1a	agtggaaacttttccctgtcca
C283	*ETV6* exon 8	gggtgagggtgaggaattacag
C294	*PAX5* exon 3	cgtgcctagcgtcagttccatcaaca
C295	*PAX5* exon 4	aagtacagcagccacccaaccaacca
C300	*PAX5* exon 1a	ccctgtccattccatcaagtcctgaa
C302	*PAX5* exon 10	ataggtgccatcagtgtttggtg
C303	*PAX5* exons 5/6	gcgcaagagagacgaaggtatt
C304	*FOXP1* exon 8	tgtcatcatagccactgacacg
C309	*ZNF521* exon 4	gtcctcatttggggagcattc
C317	*PAX5* exon 1b	gaagctccagcagtgttttctg
C326	*PAX5* exon 1	ggcttgaattattccgacctgtgagc
C327	*ZNF521* exon 8	ctgaatagggcccaagtccacttgtct
C334	*FOXP1* exon 4/5	ttcaggggtaagacgtgacctttgagg
C335	*FOXP1* exon 21	cccaaatggggttcttgatggcacta
C503	*EBF1* exon 1	gggggaggagattttccacaagaaaagg
C505	*EBF1* exon 16	cctgcacttgcagatccctcttcca

REPORTER ASSAYS

The coding regions of the exon 1a isoforms of wild type PAX5 and PAX5 mutations, deletions and fusions were cloned into the *Xho*I site of the MSCV-IRES-mRFP (MIR) vector by blunt-ended ligation. This vector was created by replacing the green fluorescent protein (GFP) cassette of the MSCV-IRES-GFP vector[54] with a monomeric red fluorescent protein cassette[55] (kindly provided by Martine Roussel of St Jude Children's Research Hospital, Memphis, TN). Twenty-four hours after plating, $2x10^5$ 293T cells were transfected with 1 μg of wild-type MIR-PAX5, mutant MIR-PAX5 or MIR without PAX5 plasmid DNA, 1 μg of *luc*-CD19 reporter plasmid DNA (kindly provided by Meinrad Busslinger, Vienna, Austria)[56], and 0.1 μg of pRL-TK *Renilla* luciferase plasmid DNA (Promega) using FuGENE 6 (Roche Diagnostics, Alameda, CA). Forty-eight hours post-transfection, cell lysis and measurement of firefly and *Renilla* luciferase activity was performed using the Dual-Luciferase® Reporter Assay System (Promega) according to the manufacturer's instructions. Transfections were performed in triplicate. The firefly luciferase activity was normalized according to corresponding *Renilla* luciferase activity, and luciferase activity was reported as mean (± s.e.m.) relative to the *luc*-CD19/PAX5 WT transfection. For competition assays, in which increasing amounts of either *PAX5-ETV6* or *PAX5-FOXP1* vector was transfected with fixed amounts of *PAX5* wild type vector, "empty" MSCV-IRES-mRFP vector was also used to maintain a constant mass of expression vector in each experiment.

ELECTROPHORETIC MOBILITY SHIFT ASSAYS

Nuclear extracts of 293T cells transfected with wild type *PAX5* or *PAX5* mutant alleles were prepared using the method of Andrews and Faller[57]. Equivalent expression of each PAX5 variant was confirmed by western blotting of nuclear extracts (data not shown). Two micrograms of protein was incubated with [γ-^{32}P]ATP end-labeled double-stranded oligonucleotides containing a CD19 promoter PAX5 binding site[58] using the Gel Shift Assay System (Promega). A mutated CD19 binding site was used as a non-specific competitor. Complexes were supershifted using a PAX5 N-terminus specific rabbit polyclonal antibody (Chemicon, Temecula, CA) and were resolved using 6% DNA retardation gels (Invitrogen).

WESTERN BLOTTING

Four million leukaemic blasts were washed with PBS and lysed with 4x LDS sample buffer (Invitrogen). For nuclear extracts, 10μl of nuclear lysate was blotted. Lysates were separated using NuPAGE 10% Bis-Tris gels, transferred to nitrocellulose membranes, and after blocking were incubated with N-terminal (Chemicon) or C-terminal PAX5 antibodies (Santa Cruz Biotechnology, Santa Cruz, CA), C-terminal Actin (Santa Cruz), and as a control for blotting of nuclear extracts, N-terminal DEK antibody (BD Biosciences). Nuclear extracts of 558LμM cells were incubated with PCNA antibody (Santa Cruz). Following incubation with secondary antibody, blots were developed using the SuperSignal West Femto Max Sensitivity Substrate chemiluminescent reagent

(Pierce, Rockford, IL) and exposed to film or captured using a ChemiDoc XRS gel imaging system (Bio-Rad Laboratories, Hercules, CA)

TRANSDUCTION AND ANALYSIS OF IgM EXPRESSION BY 558LμM CELLS

The terminally differentiated B-cell line 558LμM is a derivative of the J558L plasmacytoma cell line stably transfected with a construct expressing three of four components required for surface IgM expression: IgM heavy chain, immunoglobulin lambda light chain, and Ig-β. The cells do not express mb-1, hence the cells do not produce Ig-α and surface IgM is not expressed. mb-1 expression is dependent on Pax5, thus following transduction of 558LμM cells with ecotropic retroviruses expressing PAX5, surface IgM expression serves as a useful readout of the transactivating activity of PAX5 variant alleles.

558LμM cells were grown in RPMI 1640 media (Invitrogen) supplemented with 10% fetal bovine serum (Hyclone, Logan, UT), 2mM L-glutamine (Invitrogen), 50mg/ml gentamicin (Invitrogen), 0.3μg/ml Xanthine (Sigma, St Louis, MO), and 1μg/ml mycophenolic acid (Sigma) as previously described[59]. The Phoenix packaging system was used to generate ecotropic retrovirus expressing PAX5 cloned into either the MSCV-IRES-mRFP or MSCV-IRES-YFP constructs, as described above. 500,000 558LμM cells were transduced in six-well dishes with 3 ml of retroviral supernatant and 8μg/ml polybrene. One day post transduction cells were transferred to 10ml of fresh media. Cells were harvested for flow cytometric analysis and fluorescence activated cell sorting (FACS) day 3 post transduction.

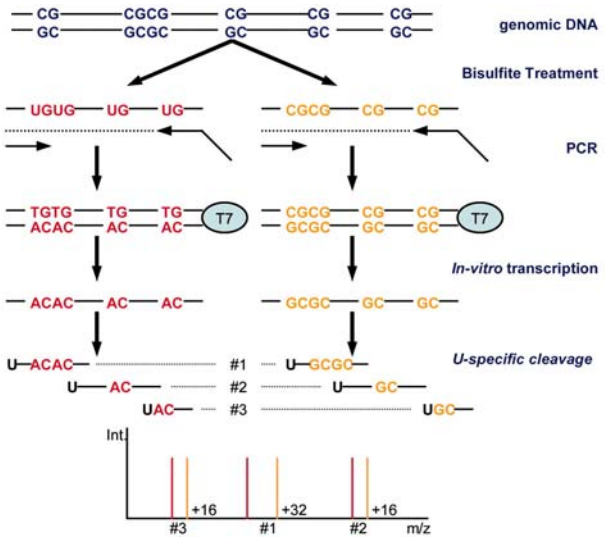
Samples were stained with fluorescein isothiocyanate conjugated anti-mouse IgM (Southern Biotechnology Associates, Birmingham, AL) and a rat IgG2aκ isotype control antibody (BD Pharmingen) on a FACSVantage SE (BD Biosciences). 100,000 viable events were collected for analysis. IgM expression was analysed in the RFP or YFP positive populations. Two million RFP-positive cells were flow sorted for western blotting.

METHYLATION ANALYSIS

Methylation status of the promoter regions of *EBF1*, *PAX5* exons 1a and 1b, and *IKZF1* (Ikaros) was analyzed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) of PCR-amplified, bisulfite-modified leukaemic blast DNA, as previously described (Sequenom, San Diego, CA)[60, 61]. This method allows semiquantitative, high-throughput analysis of methylation status of multiple CpG units in each amplicon generated by base-specific cleavage. The principles of the procedure are detailed in Supplementary Figure 1.

Ninety-six samples were examined (Supplementary Table 1). PCR reactions were designed using annotated CpG island data obtained from the University of California Santa Cruz genome browser; details are shown in Supplementary Table 8. The *PAX5* gene contains several CpG islands, one of which (Chr 9:37024136-37028341) lies

upstream of the coding region of *PAX5* exon 1a. Two amplicons were used to examine this region: X014_PAX5 lies immediately upstream of the *PAX5* exon 1a coding region in the known regulatory region of the *PAX5* promoter[62, 63]; X019_PAX5 lies is further upstream in the same CpG island. Amplicon X021_PAX5 is located in the CpG island (Chr 9 37016223-37018014) upstream of the coding region of *PAX5* exon 1b. Methylation data for each amplicon was viewed in GeneMaths XT v 1.5 (Applied Maths, Austin, TX). To compare methylation levels between ALL subtypes, the mean methylation levels of CpG units in each amplicon was calculated for each patient. Mean methylation levels for each amplicon were then compared across ALL subtypes by one-way ANOVA, and Dunn's post-hoc test.



Supplementary Figure 1. Analysis of methylation by base-specific cleavage and MALDI-TOF MS

In the example shown, the PCR product is cleaved U specifically. A methylated template (yellow) carries a conserved cytosine, and, hence, the reverse transcript of the PCR product contains CG sequences. In an unmethylated template (indicated in red), the cytosine is converted to uracil. The reverse transcript of the PCR product therefore contains adenosines in the respective positions. The sequence changes from G to A yield 16-Da mass shifts. Cleavage product 1 has two methylation sites. Mass signals of the cleavage product will differ by 32 Da when both CpG sites are either methylated or nonmethylated. For cleavage products 2 and 3, mass shifts of 16 Da will be observed, because each contains only one methylation site. The spectrum can be analyzed for the presence/absence of mass signals to determine which CpGs in the template sequence are methylated, and the ratio of the peak areas of corresponding mass signals can be used to estimate the relative methylation. This assay enables the analysis of mixtures without cloning the PCR products. Figure reproduced and the legend adapted with permission from Fig 1. of Ehrich *et al.*[60] Copyright (2005) National Academy of Sciences, U.S.A.

Supplementary Table 8. Amplicons and primers used for methylation analysis

Amplicon name	Gene	Primers (5' to 3')	Genomic location of amplicon	CpGs (N)
X009_EBF	*EBF1*	ggtatattaattttaaagggaggagg	Chr 5: 158459928-158460250	30
		caaatccccaactcaaattcctaaa		
X014_PAX5	*PAX5*	gggagggaaggaaggttttagttt	Chr 9: 37024514-37024775	21
		ccctttcaaaaacacctacttaacc		
X019_PAX5	*PAX5*	ttaggtaggaggtaaagaagaggttt	Chr 9: 37027796-37028368	45
		aacccaccataaacaccaaattacc		
X021_PAX5	*PAX5*	gtgggtttgattttattgtttttt	Chr 9: 37017437-37017883	44
		attctctcccattacccaaatatcc		
X040_ZNFN1A1	*IKZF1* (Ikaros)	tgagtaattttaggaagttattgtgaaaga	Chr 7: 50121508-50121714	20
		aaactccctctaccctaccaaacttac		

GENE SET ENRICHMENT ANALYSIS

1. Gene expression profiling using Affymetrix HG-U133A arrays.

RNA isolation from diagnostic bone marrow or peripheral blood mononuclear cell suspensions, cDNA and cRNA synthesis, labeling, fragmentation, hybridization and scanning of Affymetrix HG-U133A oligonucleotide microarrays were performed as previously described[39, 40]. Gene expression signals were scaled to a target intensity of 500, and detection values were determined using the default settings of Affymetrix Microarray Suite 5.0. Probe sets lacking present calls for any sample were excluded, and signal intensities with values of less than 1 were set to 1. Signals were then $\log_2$ transformed for subsequent analysis.

We then identified HG-U133A probe sets associated with *PAX5* deletion in *ETV6-RUNX1* positive B-progenitor ALL. The *limma* (Linear Models for Microarray Analysis)[64] and empirical Bayes t-test implemented in Bioconductor[47] (www.bioconductor.org) was used to identify differentially expressed probe sets at a Benjamini-Hochberg false discovery rate (FDR)[65] of <0.3. This identified 42 genes differentially expressed in *PAX5*-deleted *v. PAX5*-wild-type cases. Thirteen genes were overexpressed in *PAX5*-deleted cases, and 29 genes were under expressed (Supplementary Table 21). These gene lists were used as PAX5-repressed and PAX5-stimulated gene sets in subsequent Gene Set Enrichment Analysis.

2. Overview of Gene Set Enrichment Analysis (GSEA).

GSEA[66] considers the genome-wide expression profiles of two classes of samples (here, *PAX5*-mutated and *PAX5*-wild-type). Genes are ranked based on correlation between expression and class distinction. GSEA then determines if the members of a gene set *S* are randomly distributed in the ranked gene list *L*, or primarily found at the top or bottom. An enrichment score *ES* is calculated that reflects the degree to which a gene set is overrepresented at the top or bottom of the entire ranked list *L*. The *ES* is a running sum, Kolmogorov-Smirnov like statistic calculated by walking down list *L* and increasing the statistic when a gene in *S* is encountered, and decreasing it when it is not. The magnitude of the increment depends on the strength of association with phenotype, and the *ES* is the maximum deviation from zero encountered in the random walk. The significance level of *ES* is calculated by phenotype-based permutation testing, and when a database of gene sets are evaluated, as in this analysis, the significance level is adjusted for multiple hypothesis testing by calculation of a false discover rate FDR.

3. Cross-subtype GSEA of PAX5-regulated genes in B-progenitor ALL

GSEA implemented in R (www.r-project.org) was used to assess enrichment of previously described functional gene sets, and the PAX5 regulated gene sets identified above, in B-progenitor ALL. Of 525 publicly available functional (GSEA C_2) gene sets, those with less than 10 or greater 500 genes were excluded, to avoid skewing of P and FDR values. To avoid the potential confounding effect of B-ALL subtype-specific gene expression on GSEA analysis, we first examined enrichment in B-ALLs without high

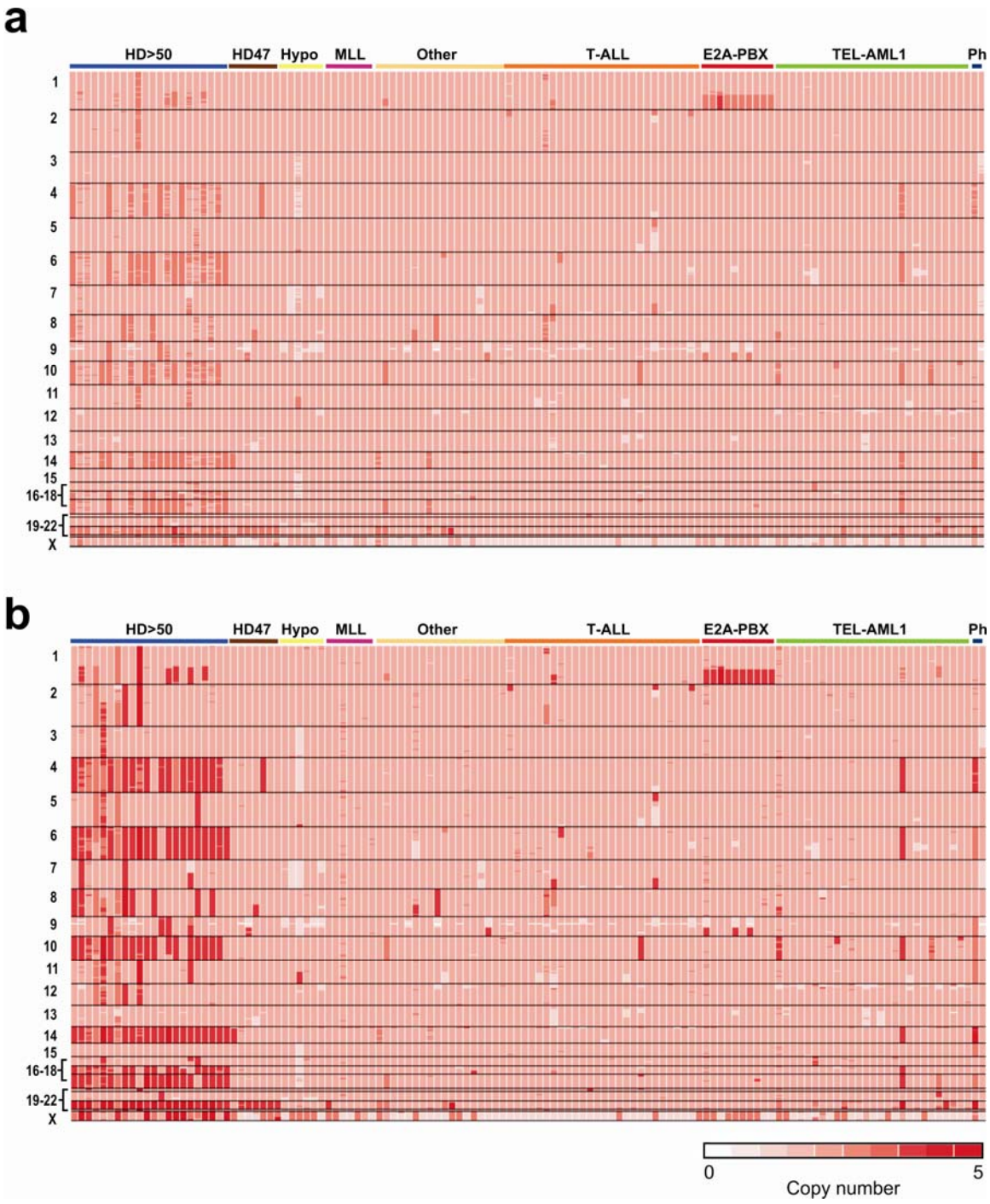
hyperdiploidy or *ETV6-RUNX1*, *TCF3-PBX1*, *BCR-ABL1* or *MLL* rearrangement. This group thus included samples with low hyperdiploidy, hypodiploidy, and normal or miscellaneous karyotype. We then examined enrichment in the entire non-*ETV6-RUNX1* B-ALL cohort. Significantly enriched gene sets after 1000 permutations at a FDR of <0.25 are reported.

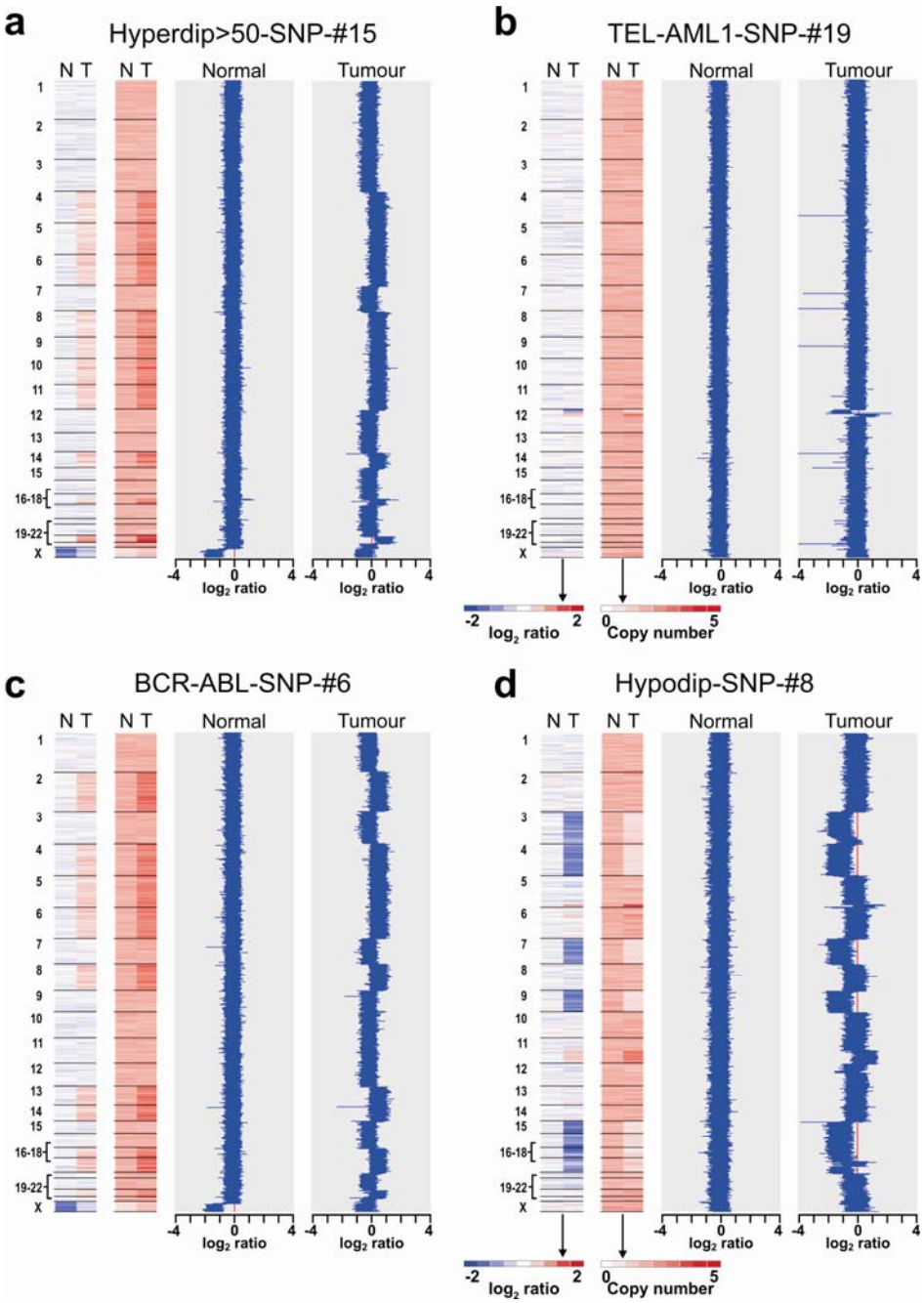
SUPPLEMENTARY RESULTS

DNA COPY NUMBER CHANGES IN PAEDIATRIC ALL

Supplementary Figure 2. Effect of karyotype-guided normalization on copy number inference

a shows dChipSNP Hidden Markov Model inferred copy number for a representative subset of 123 paediatric ALL cases using 50K Xba data and the default invariant set array normalization algorithm in dChipSNP. **b** shows the same cases following karyotype-guided normalization (cytnormalization) and subsequent copy number inference in dChipSNP.


a
HD>50
HD47
Hypo
MLL
Other
T-ALL
E2A-PBX
TEL-AML1
Ph
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16-18
19-22
X
b
HD>50
HD47
Hypo
MLL
Other
T-ALL
E2A-PBX
TEL-AML1
Ph
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16-18
19-22
X
0
5
Copy number

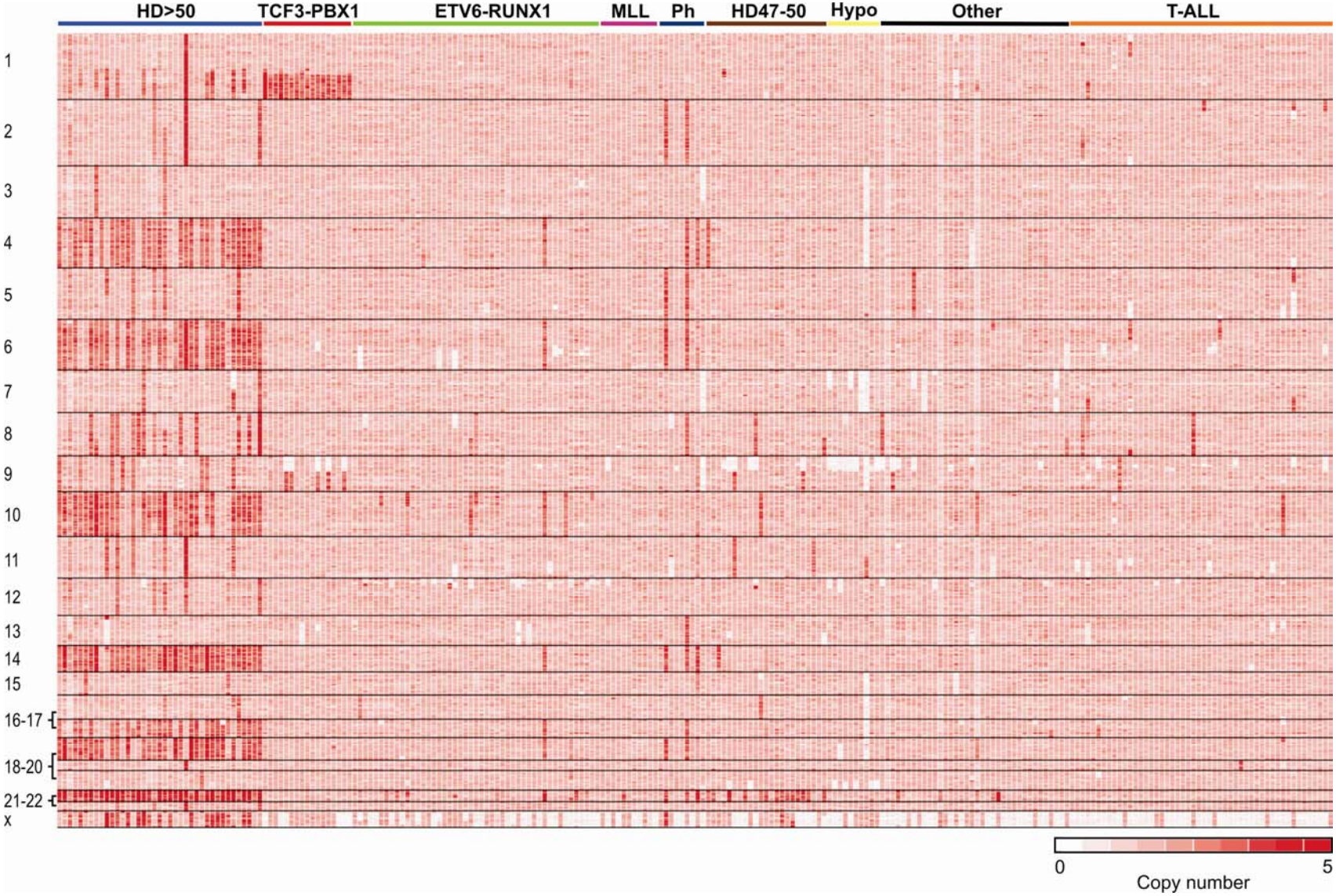


Supplementary Figure 3. Examples of paired tumor-germline copy number data demonstrating the somatic nature of copy number abnormalities

a-d, Median smoothed $\log_2$-ratio and copy number data is shown for four representative cases with a spectrum of focal to whole chromosomal copy number abnormalities, that are not evident in corresponding germline samples. N, normal (corresponding germline sample); T, tumour.

Supplementary Figure 4. DNA copy number changes in 242 paediatric ALL cases

A copy number heatmap generated by dChipSNP is shown where each case is represented by a column. Pink represents diploid copy number, white deletion and red amplification. Median smoothed copy number is shown. Abbreviations: HD>50, hyperdiploidy with greater than 50 chromosomes; Ph, *BCR-ABL1* positive ALL; Hypo, B-precursor ALL with hypodiploidy.


HD>50
TCF3-PBX1
ETV6-RUNX1
MLL
Ph
HD47-50
Hypo
Other
T-ALL
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16-17
18-20
21-22
x
0
5
Copy number

MULTIPLE NOVEL REGIONS OF GENOMIC DELETION AND AMPLIFICATION IN PAEDIATRIC ALL

Supplementary Table 9. Differences in frequency of genomic gains and losses between B-progenitor ALL subtypes

ANOVA post-hoc tests comparing differences in genomic lesion frequency between B-progenitor ALL subtypes. Bonferroni/Dunn P values adjusted for multiple comparisons are reported; *P values <0.0018 are significant.

Group		Amplifications		Deletions		All lesions	
		Mean difference	P	Mean difference	P	Mean difference	P
HD>50	TCF3-PBX1	7.976	<0.0001*	-0.528	0.6378	7.422	<0.0001*
HD>50	ETV6-RUNX1	8.67	<0.0001*	-4.432	<0.0001*	4.447	<0.0001*
HD>50	MLL	9.473	<0.0001*	0.681	0.061	10.128	<0.0001*
HD>50	BCR-ABL1	5.564	<0.0001*	-2.632	0.07	4.35	0.02
HD>50	HD47-50	7.868	<0.0001*	-1.889	0.06	5.998	<0.0001*
HD>50	Hypodiploid	8.464	<0.0001*	-4.41	0.0015*	4.028	0.02
HD>50	Other	8.425	<0.0001*	-3.049	0.0008*	5.545	<0.0001*
TCF3-PBX1	ETV6-RUNX1	0.695	0.35	-3.904	0.0004*	-2.975	0.04
TCF3-PBX1	MLL	1.497	0.14	1.209	0.42	2.706	0.16
TCF3-PBX1	BCR-ABL1	-2.142	0.03	-2.105	0.19	-3.072	0.13
TCF3-PBX1	HD47-50	-0.107	0.9	-1.361	0.27	-1.425	0.37
TCF3-PBX1	Hypodiploid	0.488	0.64	-3.882	0.01	-3.394	0.09
TCF3-PBX1	Other	0.449	0.56	-2.521	0.03	-1.877	0.2
ETV6-RUNX1	MLL	0.803	0.36	5.112	0.0001*	5.681	0.008
ETV6-RUNX1	BCR-ABL1	-3.106	0.0014*	1.779	0.2	-0.097	0.96
ETV6-RUNX1	HD47-50	-0.802	0.23	2.543	0.01	1.55	0.22
ETV6-RUNX1	Hypodiploid	-0.206	0.82	0.021	0.99	-0.419	0.8
ETV6-RUNX1	Other	-0.245	0.67	1.382	0.11	1.098	0.32
MLL	BCR-ABL1	3.909	0.0011*	-3.313	0.06	-5.778	0.01
MLL	HD47-50	-1.605	0.1	-2.569	0.07	-4.13	0.02
MLL	Hypodiploid	-1.009	0.38	-5.091	0.003	-6.1	0.005
MLL	Other	-1.048	0.25	-3.73	0.006	-4.583	0.008
BCR-ABL1	HD47-50	2.304	0.03	0.744	0.62	1.647	0.4
BCR-ABL1	Hypodiploid	2.9	0.02	-1.778	0.32	-0.322	0.89
BCR-ABL1	Other	2.861	0.004	-0.417	0.77	1.194	0.52
HD47-50	Hypodiploid	0.596	0.55	-2.522	0.09	-1.97	0.3
HD47-50	Other	0.557	0.43	-1.161	0.26	-0.453	0.73
Hypodiploid	Other	-0.039	0.97	1.361	0.32	1.517	0.39

Supplementary Table 10. Shared regions of deletion and amplification in paediatric ALL

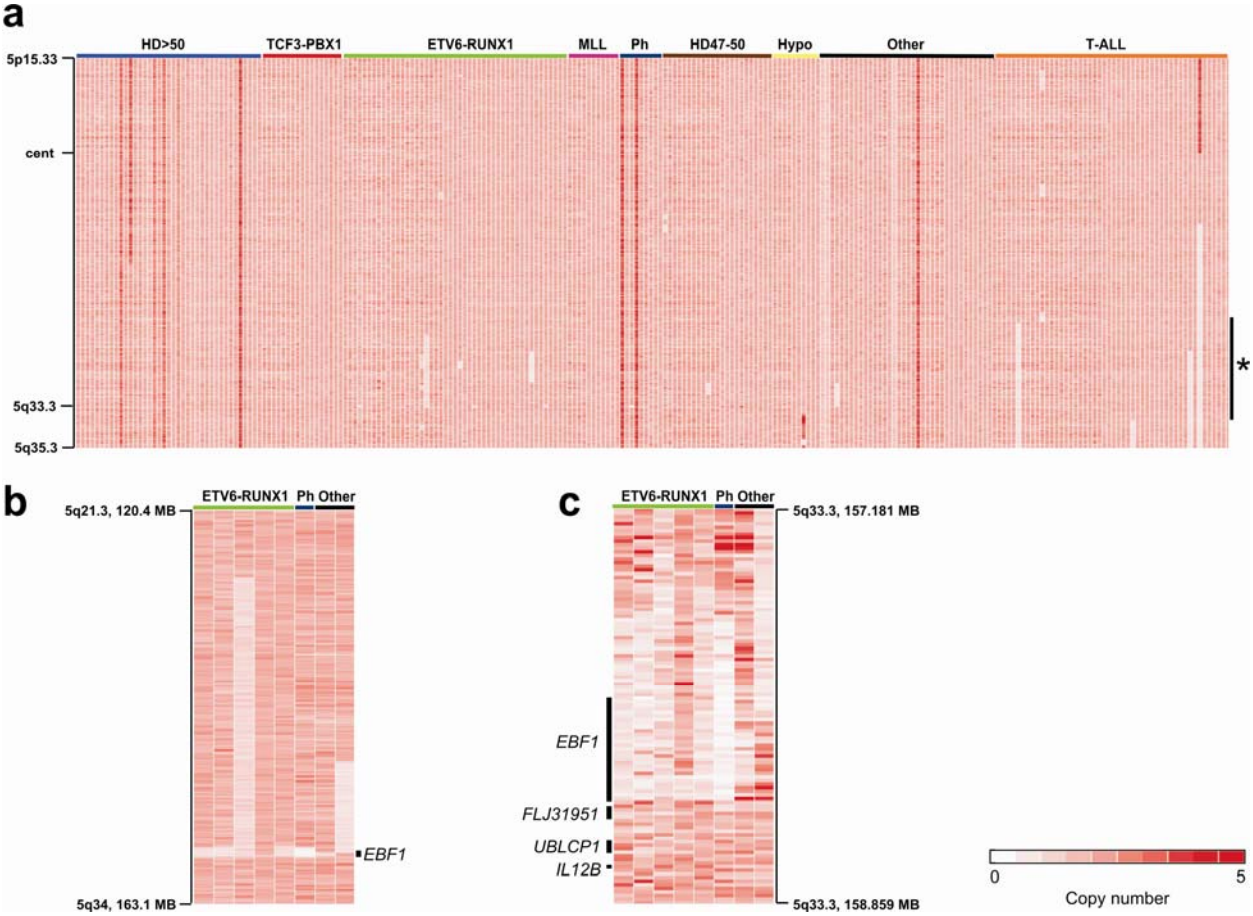
Summary statistics are shown for regions of genomic gain or loss affecting more than one ALL case. Start and end positions and genomic size of the minimally deleted regions (MDR) are also shown. Whole chromosomal gains are excluded. Genomic gains most commonly involved whole chromosomes or large regions adjacent to chromosomal breakpoints in cases with known translocations. Genes in each MDR are listed according to the NCBI Build 35.1 (hg17) human genome sequence. Copy number changes secondary to antigen receptor gene rearrangement at 2p11.2 (*IGKL*), 7p14.1 (*TRGV*), 7q34 (*TRBV*), 14q11.2 (*TRAV*, *TRDV*, *TRDJ*, *TRDC*, and *TRAJ*), 14q32.33 (*IGHV*) and 22q11.22 (*IGLL*) are not shown. Near haploid cases are excluded. *Region does not contain any known genes or micro-RNA encoding genes. †Region of deletion varies between cases, involving either *TBL1XR1* or regions immediately upstream or downstream of the gene; the smallest deletion is 0.118 Mb. ‡ Region of deletion varies between cases. ¶In all but one case the region of deletion extends telomerically to involve *NF1*. **All five deletions of *PAX5* in T-ALL were large, and in three cases were contiguous with deletions involving the *CDKN2A* at 9p21.3.

Cytoband	Start (Mb)	End (Mb)	Size (Mb)	Hyperdiploid >50	TCF3-PBX1	ETV6-RUNX1	MLL	BCR-ABL1	Hyperdiploid 47-50	Hypo	Other	B-ALL N (%)	T-ALL N (%)	Gene(s) in region
				N=39	N=17	N=47	N=11	N=9	N=23	N=10	N=36			
Deletions														
1p33	47.440	47.479	0.039	0	0	0	0	0	0	0	0	0 (0)	3 (6.0)	*TAL1*
1q31.3	191.317	191.418	0.101	0	0	5	0	0	0	0	1	6 (3.13)	1 (2.0)	*TROVE2, GLRX2, CDC73, B3GALT2*
2p25.3	3.541	3.801	0.260	1	0	4	0	0	0	0	0	5 (2.60)	0 (0)	No annotated gene*
2p21	43.337	43.624	0.287	0	0	0	0	0	0	0	2	2 (1.04)	1 (2.0)	*THADA*
2q37.1-q37.3	232.465	qtel	10.265	1	0	2	0	0	0	0	0	3 (1.56)	1 (2.0)	124 genes
3p22.3	35.339	35.645	0.306	0	0	2	0	1	0	1	0	4 (2.08)	0 (0)	No annotated gene*
3p14.2	60.064	60.318	0.254	0	0	4	0	1	1	1	1	8 (4.17)	0 (0)	*FHIT*
3q13.2	113.538	113.686	0.148	0	0	7	0	3	1	1	1	13 (6.77)	0 (0)	*CD200*, *BTLA*
3q26.32	Various			0	0	6	0	0	1	0	0	7 (3.13)	0 (0)	*TBL1XR1*†
4q25	109.393	109.442	0.049	0	0	1	0	0	0	1	1	3 (1.56)	4 (8.0)	*LEF1*
4q31.21	144.540	144.613	0.073	0	0	1	0	0	1	1	2	5 (2.60)	0 (0)	No annotated gene*
4q31.23	150.055	150.200	0.145	0	0	5	0	0	0	1	0	6 (3.13)	1 (2.0)	None; telomeric to *NR3C2*
5q31.3	142.760	142.847	0.087	1	0	6	0	0	1	0	1	9 (4.69)	3 (6.0)	*NR3C1, LOC389335*
5q33.3	Various‡			0	0	5	0	1	0	1	1	8 (4.17)	3 (6.0)	*EBF1*
5q34	163.535	5qtel	17.265	0	0	0	0	0	0	0	0	0 (0)	4 (8.0)	172 genes
6p22.22	26.345	26.368	0.023	1	0	2	0	2	1	3	4	13 (6.77)	0 (0)	*HIST1H4F, HIST1H4G, HIST1H3F, HIST1H2BH*
6q16.2-3	99.852	102.492	2.640	1	1	7	0	0	0	0	1	10 (5.21)	5 (10)	16 genes including *CCNC*
6q21	109.347	109.435	0.088	0	0	7	0	0	0	0	4	11 (5.73)	4 (8.0)	*ARMC2*, *SESN1*

Cytoband	Start (Mb)	End (Mb)	Size (Mb)	Hyperdiploid >50	TCF3-PBX1	ETV6-RUNX1	MLL	BCR-ABL1	Hyperdiploid 47-50	Hypo	Other	B-ALL N (%)	T-ALL N (%)	Gene(s) in region
7p				1	0	1	0	0	0	1	2	5 (2.60)	0 (0)	All 7p
7p21.3	11.903	12.134	0.231	1	0	0	0	1	2	4	4	12 (6.25)	1 (2.0)	*KIAA0960*, *FLJ11273*
7p12.2	50.193	50.241	0.048	2	0	0	1	3	1	5	5	17 (8.85)	1 (2.0)	*IKZF1* (*ZNFN1A1*, Ikaros)
7q21.2	91.900	92.109	0.209	0	0	0	0	1	0	3	2	6 (3.13)	0 (0)	*LOC645862, GATAD1, ERVWE1, PEX1, DKFZP564O0523, LOC442710, MGC40405, CDK6*
8q12.1	60.195	60.289	0.094	0	0	4	0	1	0	0	2	7 (3.65)	0 (0)	Immediately 5' (telomeric) of *TOX*
9p				0	5	0	1	0	2	4	2	14 (7.29)	1 (2.0)	All 9p
9p21.3	20.504	20.637	0.133	3	5	4	1	1	7	9	9	39 (20.31)	13 (26.0)	*MLLT3* (*AF9*)
9p21.3	Various‡			9	6	12	1	4	10	10	13	65 (33.85)	36 (72.0)	*CDKN2A*
9p13.2	Various‡			3	7	13	1	4	7	10	12	57 (29.69)	5 (10)**	*PAX5*
9q22.32	96.112	96.173	0.061	0	0	3	0	1	0	1	0	5 (2.60)	2 (4.0)	*FAM22F, LOC728026*
10q23.31	89.666	89.728	0.062	0	0	0	0	0	0	0	0	0 (0)	3 (6.0)	*PTEN*
10q24.1	97.879	98.057	0.178	0	0	1	0	0	0	0	1	2 (1.04)	0 (0)	*BLNK*
10q25.1	111.772	111.850	0.078	1	0	2	0	3	0	0	3	9 (4.69)	0 (0)	*ADD3*
11p13	33.874	34.029	0.155	0	0	0	0	0	0	0	1	1 (0.52)	4 (8.0)	No gene; immediately 5' of *LMO2*
11p12	36.575	36.583	0.008	0	0	3	0	0	1	0	0	4 (2.08)	2 (4.0)	*RAG2*, *LOC119710*
11q13.1	63.721	63.781	0.060	0	0	3	0	1	0	0	0	4 (2.08)	2 (4.0)	*STIP1, URP2, DNAJC4, TRPT1, NUDT22, VEGFB, FKBP2, LOC728892, PPP1R14B, PLCB3*
11q23	107.119	109.887	2.768	0	0	2	0	1	0	0	1	4 (2.08)	6 (12.0)	20 genes including *RAB39, NPAT, ATM*
11q23.3	117.882	118.379	0.497	0	0	2	2	1	0	0	1	6 (3.13)	2 (4.0)	16 genes distal to *MLL* breakpoint, including 3' *MLL*
12p13.2*	Various‡	11.808	0.020	2	0	33	1	1	2	2	10	51 (26.56)	4 (8.0)	*ETV6*
12q21.33	90.786	91.039	0.253	0	0	6	0	2	3	1	1	13 (6.77)	0 (0)	3' of *BTG1*
13q14.11	40.453	40.484	0.031	1	2	4	0	0	1	1	0	9 (4.69)	2 (4.0)	*ELF1*
13q14.11	43.758	43.895	0.137	2	2	3	0	1	0	1	1	10 (5.21)	3 (6.0)	*C13orf21, LOC400128*
13q14.2	47.885	47.968	0.083	2	2	2	0	1	1	1	0	9 (4.69)	6 (12.0)	*RB1*
13q14.2-3	49.471	50.360	0.889	5	2	3	0	0	1	1	0	12 (6.25)	3 (6.0)	*DLEU2, RFP2, KCNRG, MIRN16-1, MIRN15A, DLEU1, FAM10A4, LOC647154, LOC730194, DLEU7*
14q24.2	72.289	72.423	0.134	0	0	1	0	0	1	0	1	3 (1.56)	0 (0)	*DPF3*
15q15.1	39.045	39.837	0.792	0	0	3	0	0	1	1	1	6 (3.13)	0 (0)	18 genes including *LTK* and *MIRN626*
16q22.1	66.116	66.423	0.307	0	0	1	0	0	0	1	0	2 (1.04)	3 (6.0)	*FAM65A, CTCF, RLTPR, ACD, PARD6A, C16orf48, LOC388284, GFOD2, RANBP10, TSNAXIP1, CENPT*
17p13.3-11.2	tel	18.837	18.837	1	0	1	0	0	0	1	2	5 (2.60)	2 (4.0)	383 genes
17q11.2	26.090	26.259	0.169	0	0	2	0	0	0	1	1	4 (2.08)	2 (4.0)	*LOC729690, SUZ12P, CRLF3, LOC646013, C17orf41, C17orf42, [NF1]*¶

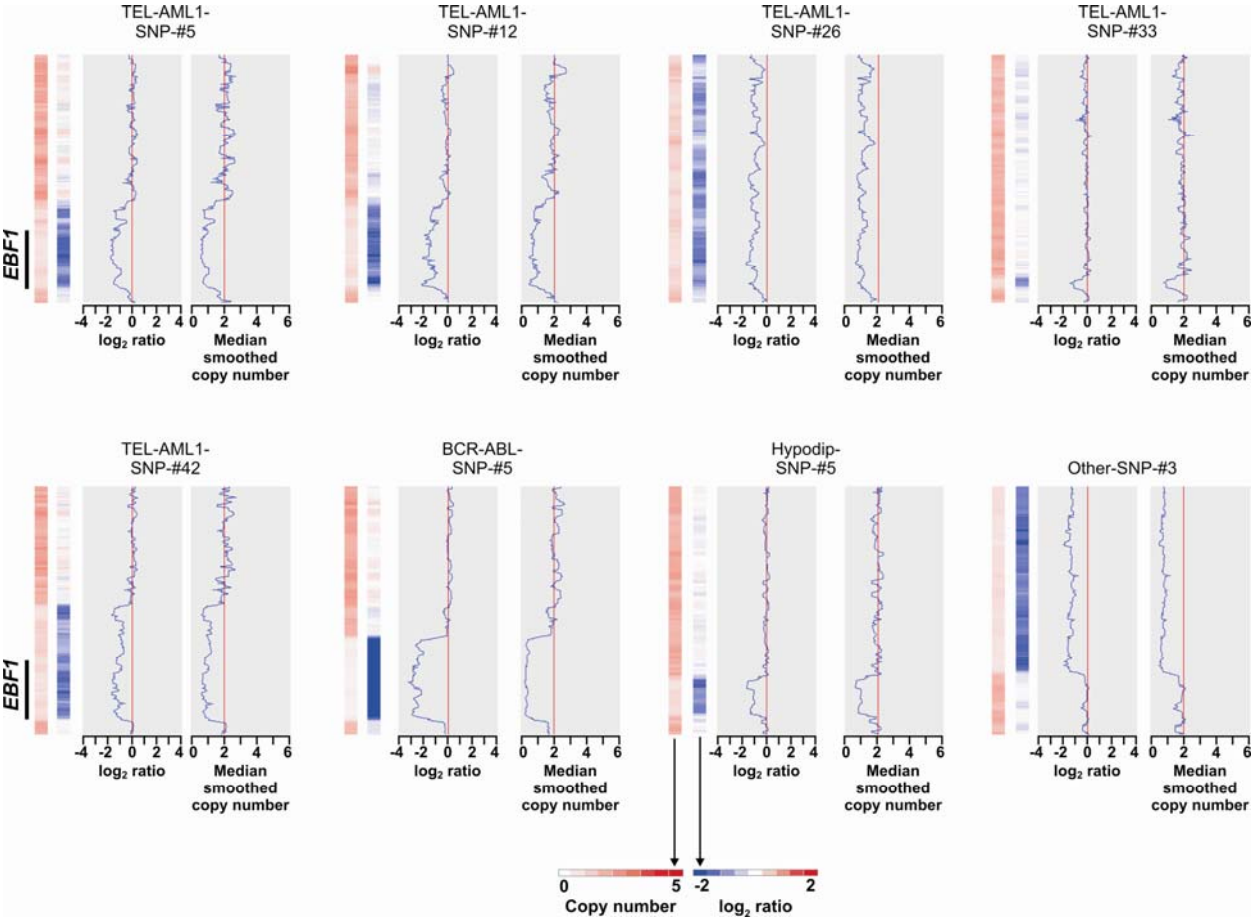
Cytoband	Start (Mb)	End (Mb)	Size (Mb)	Hyperdiploid >50	TCF3-PBX1	ETV6-RUNX1	MLL	BCR-ABL1	Hyperdiploid 47-50	Hypo	Other	B-ALL N (%)	T-ALL N (%)	Gene(s) in region
17q21.1	35.185	35.230	0.045	0	0	0	0	0	0	2	1	3 (1.56)	0 (0)	*IKZF3* (*ZNFN1A3*, Aiolos)
19p13.3	0.229	1.531	1.302	1	16	0	0	0	0	0	0	17 (8.85)	0 (0)	63 genes telomeric to *TCF3*; region may include *TCF3*
19q13.32	52.086	52.292	0.206	0	0	2	0	0	0	0	0	2 (1.04)	0 (0)	*GRLF1, LOC729514, NPAS1, TMEM160, C19orf7*
20p12.1	10.370	10.405	0.035	0	0	4	0	2	2	0	1	9 (4.69)	1 (2.0)	*C20orf94*
20q				2	0	1	0	1	2	5	0	11 (5.73)	0 (0)	
21q22.12	35.350	35.354	0.004	0	0	3	0	0	0	0	0	3 (1.56)	0 (0)	No gene, but immediately distal to *RUNX1*
21q22.2	38.706	38.729	0.023	0	0	0	0	0	1	0	4	5 (2.60)	0 (0)	*ERG*
Amplifications														
1q				6	3	0	0	0	0	0	0	9 (4.69)	0 (0)	All 1q
1q23.3-q44	161.491	qtel	81.326	0	16	0	0	0	0	0	0	16 (8.33)	0 (0)	719 genes telomeric of *PBX1*, including 3' region of *PBX1*
1q	Variable subregion			6	1	2	0	0	1	0	1	11 (5.73)	1 (2.0)	
2p25.3-2p22.3	2qtel	32.046	31.859	0	0	0	0	0	0	0	0	0 (0)	3 (6.0)	235 genes
6p25.3-p22.2	6ptel	26.216	26.216	1	0	0	0	0	0	0	1	2 (1.04)	2 (4.0)	190 genes
6q23.3	135.556	135.714	0.158	0	0	0	0	0	0	0	0	0 (0)	5 (10)	*MYB, MIRN548A2, AHI1*
9q				0	5	0	0	0	2	0	1	8 (4.17)	0 (0)	All 9q
9q34.12-q34.3	130.687	qtel	7.676	0	0	0	0	3	0	0	0	3 (1.56)	0 (0)	155 genes telomeric of *ABL1*, including 3' region of *ABL1*
10p				0	0	3	0	0	0	0	0	3 (1.56)	0 (0)	All 10p
21q22.11-q22.12	32.896	35.199	2.303	0	0	4	0	0	1	0	1	6 (3.125)	0 (0)	33 genes including *RUNX1*
22q11.1-q11.23	ptel	21.888	21.888	0	0	0	0	3	0	0	0	3 (1.56)	0 (0)	277 genes telomeric (5') of *BCR*, including 5' region of *BCR*

FOCAL DELETIONS OF *EBF1* IN ALL



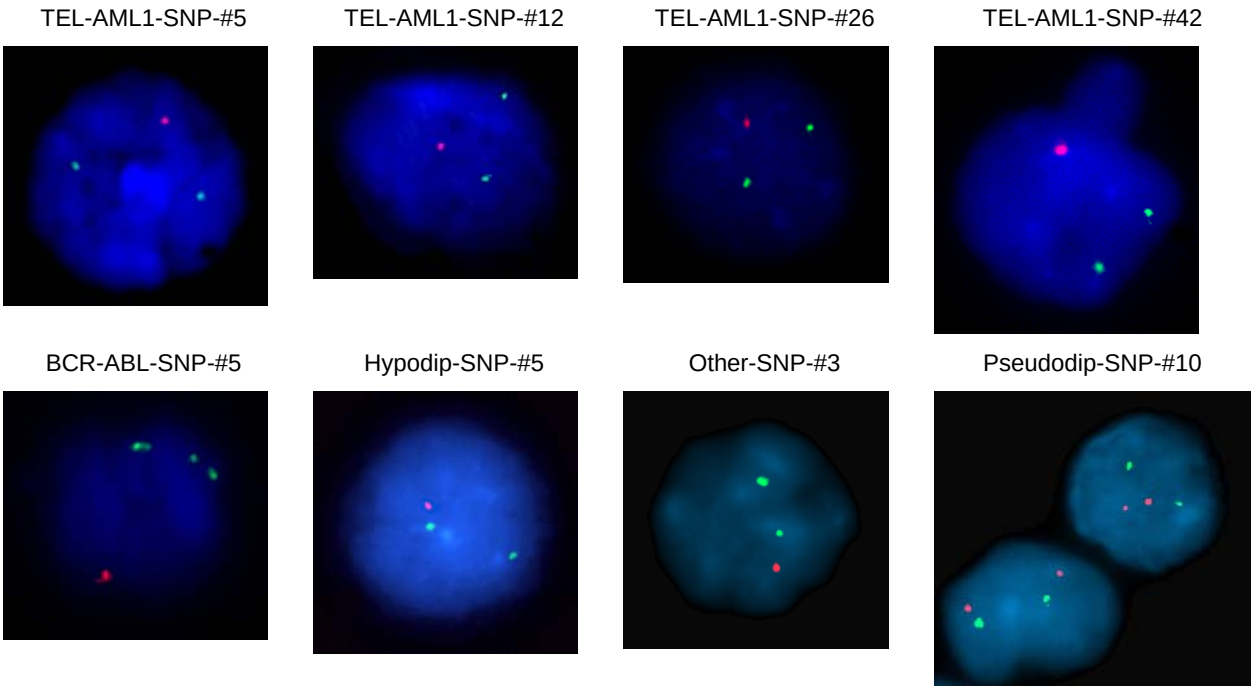
Supplementary Figure 5. Focal deletions involving *EBF1* in ALL

a, dChipSNP median-smoothed copy number heatmap showing chromosome 5 for 242 ALL cases. Cases are shown in columns, and SNPs arranged from 5p(tel) to 5q(tel) from top to bottom. * region shown in panel (b). **b,** Zoomed region of copy number heatmap showing extent of deletion in 8 EBF1-deleted cases. **c,** dChipSNP uninferred copy number in the same 8 cases in panel B. HD>50, hyperdiploidy with greater than 50 chromosomes; hypo, B-ALL with hypodiploidy; Ph, B-ALL with *BCR-ABL1*. Abbreviations: HD>50, hyperdiploidy with greater than 50 chromosomes; Ph, *BCR-ABL1* positive ALL; Hypo, B-precursor ALL with hypodiploidy.



Supplementary Figure 6. Copy number heatmaps and plots for B-ALL cases with *EBF1* deletions

For each case, two dChipSNP heat maps are shown: median smoothed copy number data (left panel, white-red scale) and smoothed log ratio data (centre panel, blue-red scale). The right panels for each patient depict median smoothed $\log_2$ ratio and copy number data.



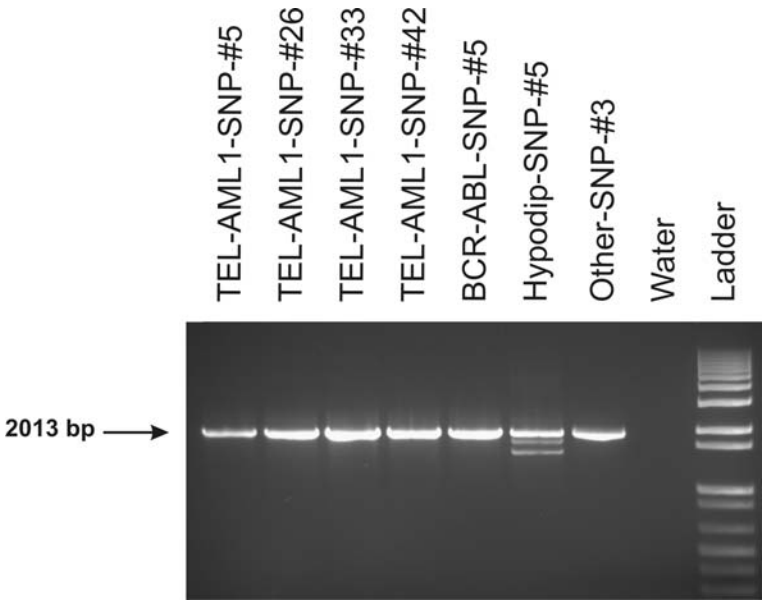
Supplementary Figure 7. Confirmation of *EBF1* deletions by FISH

The *EBF1*-specific probe is labelled with rhodamine (red) and control probe with fluorescein (green). Case Pseudodip-SNP-#10 has no *EBF1* deletion, and is included as a control.

Supplementary Table 11. FISH results for B-ALL cases with *EBF1* deletions

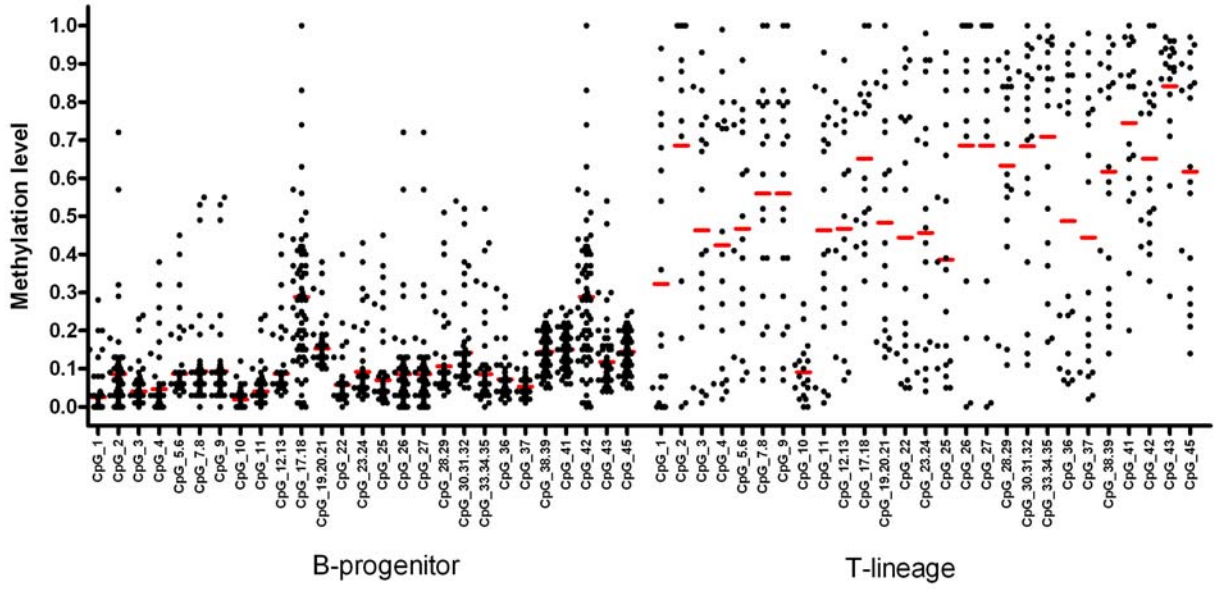
[1]Sort purity 94%. [2]Sort purity 92%.

				FISH results (N cells)			
Sample	Blast %	EBF1 deleted	FISH probe	Hemizygous deletion	Homozygous deletion	No deletion	Cells with deletion (%)
TEL-AML1-SNP-#12	99	Yes	RP11-586B19	92	3	5	95
BCR-ABL-SNP-#5	78	Yes	RP11-160B5	29	68	3	97
- $CD10^-$ sort fraction[1]		Yes	RP11-160B5	0	97	3	97
- $CD10^+$ sort fraction[2]		Yes	RP11-160B5	98	0	2	98
Hypodip-SNP-#5	93	Yes	RP11-160B5	84	0	16	84
Other-SNP-#3	99	Yes	RP11-45M18	94	0	6	94
TEL-AML1-SNP-#26	84	Yes	RP11-586B19	84	6	10	90
TEL-AML1-SNP-#5	93	Yes	RP11-586B19	90	0	10	90
TEL-AML1-SNP-#42	84	Yes	RP11-586B19	76	0	24	76
Control samples							
TEL-AML1-SNP-#39	96	No	RP11-45M18	1	0	99	1
TEL-AML1-SNP-#39	96	No	RP11-586B19	3	0	97	3
TEL-AML1-SNP-#22	97	No	RP11-45M18	2	0	98	2
TEL-AML1-SNP-#22	97	No	RP11-586B19	0	0	100	0
Pseudodip-SNP-#10	99	No	RP11-45M18	0	0	98	0
Pseudodip-SNP-#10	99	No	RP11-586B19	0	0	100	0
E2A-PBX1-SNP-#9	92	No	RP11-45M18	0	0	100	0
E2A-PBX1-SNP-#9	92	No	RP11-586B19	1	0	98	1
T-ALL-SNP-#28	98	No	RP11-45M18	0	1	99	1
T-ALL-SNP-#28	98	No	RP11-586B19	0	0	100	0



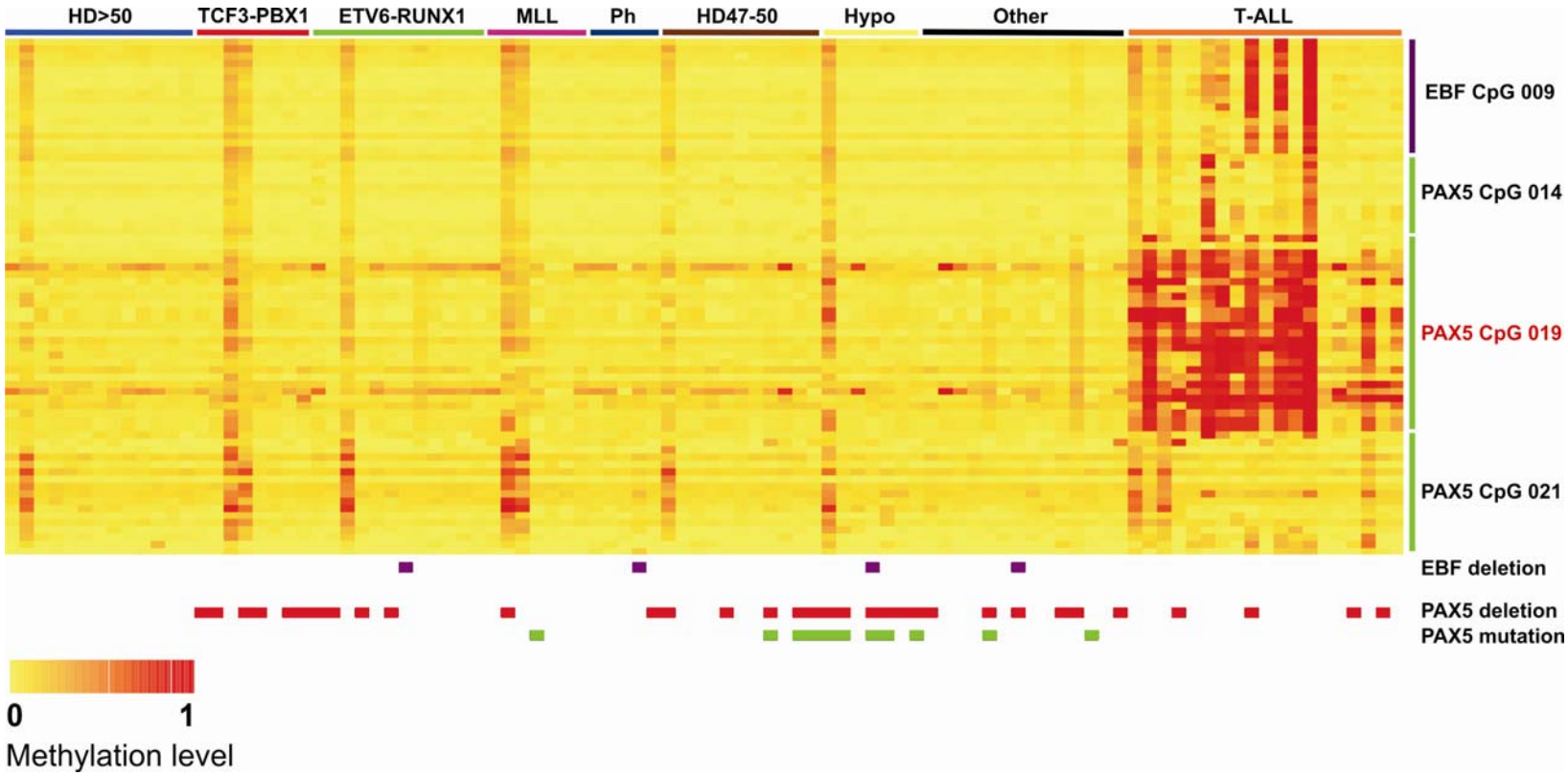
Supplementary Figure 8. Amplification of wild-type EBF1 in *EBF1*-deleted B-progenitor ALL

cDNA encompassing the entire coding region of EBF1 was amplified using primers C503 and C505 (Supplementary Table 7) in cases with hemizygous *EBF1* deletion.



Supplementary Figure 9. Representative data showing methylation levels of each base-specific cleavage product of the X019 PCR amplicon in the PAX5 exon 1A CpG island

Scatter plots depicting the degree of methylation of each base-specific cleavage fragment of the X019 PCR amplicon of the CpG island in the PAX5 exon 1A promoter. Samples are grouped according to lineage. Dots represent degree of methylation for each case, and red horizontal lines indicate means.



Supplementary Figure 10. Heatmap of methylation data of CpG islands in the *EBF1, PAX5* exon 1a (PAX5 CpG 014 and 019) and *PAX5* exon 1B (PAX5 CpG 021) promoters

Heatmap showing showing methylation data for each *EBF* and *PAX5* amplicon examined. Each column represents a sample (N=96). Relative level of methylation is shown for each CpG unit in each amplicon (shown in rows), with yellow representing no methylation, and red complete methylation. The most striking gene-specific methylation of *PAX5* and *EBF1* was observed in T-ALL cases. Several cases show global low-to-moderate hypermethylation of *EBF1, PAX5*, and numerous other genes (data not shown). The PAX5 exon 1A promoter CpG island amplicon 019, for which data is shown in scatter plot form in Supplementary Figure 9, is indicated in red.

A HIGH FREQUENCY OF MONO-ALLELIC *PAX5* DELETIONS IN B-ALL

Supplementary Table 12. List of genes encoding regulators of B cell differentiation, and genes encoding targets of B cell regulators, examined using the SNP microarrays

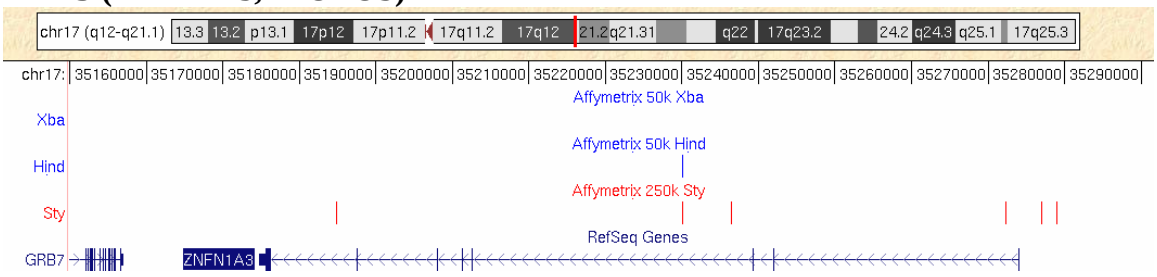
Coverage is indicated as poor if there are less than three Hind, Xba or Sty SNPs located within the genomic locus of each gene.

Gene symbol	Alternative symbol	Gene name	Chromosome	Size of genomic locus (bp)	SNP array coverage
BCL11A	EVI9	B-cell CLL/lymphoma 11A	2p16.1	102331	Good
BLNK	SLP-65	B-cell linker	10q23.2-q23.33	79864	Good
BLK		B lymphoid tyrosine kinase	8p23-p22	70587	Good
BTK		Bruton agammaglobulinemia tyrosine kinase	Xq21.33-q22	36741	Poor
CD19		CD19 molecule	16p11.2	7402	Poor
CD79A	IGA; MB-1	CD79a molecule, immunoglobulin-associated alpha	19q13.2	4249	Poor
CD79B	IGB; B29	CD79b molecule, immunoglobulin-associated beta	17q23	3606	Poor
CSF1R	C-FMS; M-CSF-R	colony stimulating factor 1 receptor	5q33-q35	60077	Good
DNTT	TDT	deoxynucleotidyltransferase, terminal	10q23-q24	34235	Good
EBF1	OLF1	early B-cell factor	5q34	401342	Good
FLT3	FLK2	fms-related tyrosine kinase 3	13q12	96951	Good
IGLL1	CD179b; lambda5; VPREB2	immunoglobulin lambda-like polypeptide 1	22q11.23	7181	Poor
IL7R	IL-7R-alpha	interleukin 7 receptor	5p13	19931	Good
IRF4	MUM1	interferon regulatory factor 4	6p25-p23	19434	Poor
IRF8	ICSBP	interferon consensus sequence binding protein 1	16q24.1	23436	Good
LEF1	TCF1 alpha	lymphoid enhancer-binding factor 1	4q23-q25	120878	Good
LYN		v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	8q13	130760	Good
NOTCH1		Notch homolog 1, translocation-associated (Drosophila)	9q34.3	51342	Poor
PAX5	BSAP	paired box gene 5	9p13	195946	Good
PLCG2		phospholipase C, gamma 2	16q24.1	178969	Good
RAG1		recombination activating gene 1	11p13	11733	Poor
RAG2		recombination activating gene 2	11p13	6292	Good
SOX4	EVI16	SRY (sex determining region Y)-box 4	6p22.3	4876	Poor
SPI1	PU.1	spleen focus forming virus (SFFV) proviral integration oncogene spi1	11p11.2	23687	Good
SYK		spleen tyrosine kinase	9q22	94408	Good
TCF3	E2A	transcription factor 3	19p13.3	40983	Poor
VPREB1	VPREB	pre-B lymphocyte gene 1	22q11.22	727	Poor
VPREB3		pre-B lymphocyte gene 3	22q11	1659	Poor
IKZF1	ZNFN1A1, IKAROS	zinc finger protein, subfamily 1A, 1 (Ikaros)	7p13-p11.1	123130	Good
IKZF2	ZNFN1A2, HELIOS	zinc finger protein, subfamily 1A, 2 (Helios)	2qter	143715	Good
IKZF3	ZNFN1A3, AIOLOS	zinc finger protein, subfamily 1A, 3 (Aiolos)	17q21	99241	Good

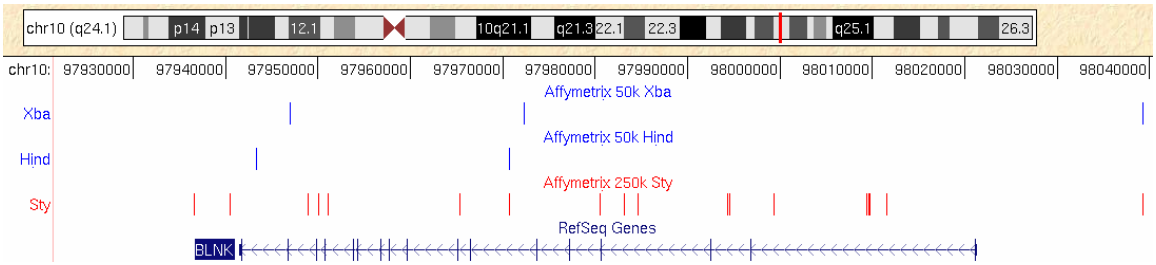
Supplementary Figure 11. SNP coverage for key genes in the B cell differentiation pathway

Plots were generated by loading a custom annotation file with the genomic position of each SNP interrogated by each array into the University of California Santa Cruz genome browser (http://www.genome.ucsc.edu/cgi-bin/hgGateway). Position of each probe set is represented by a vertical line below the array identifier.

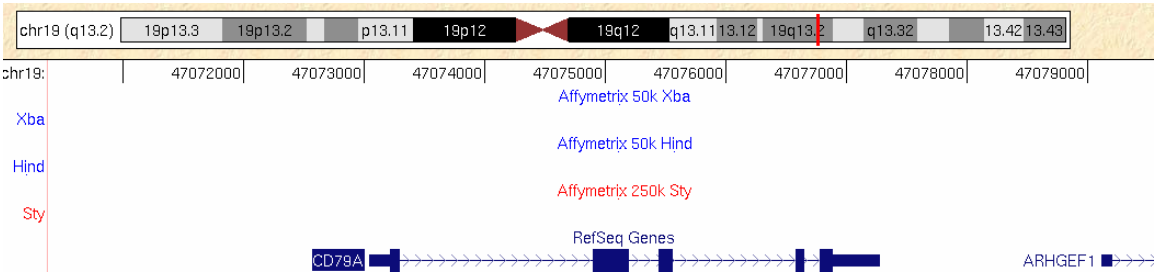
***IKZF3* (ZNFN1A3, AIOLOS)**



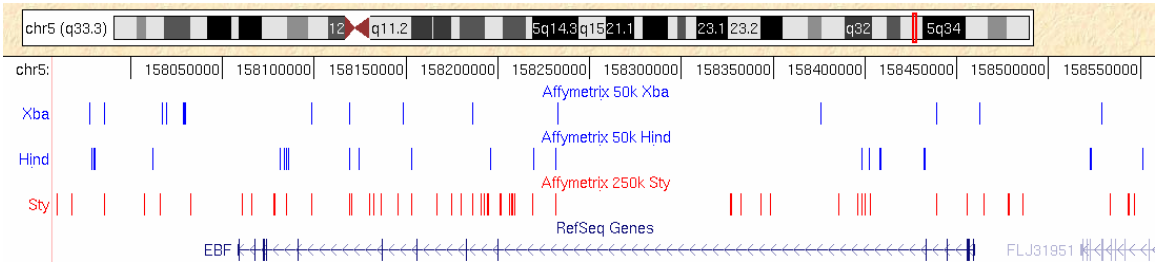
BLNK



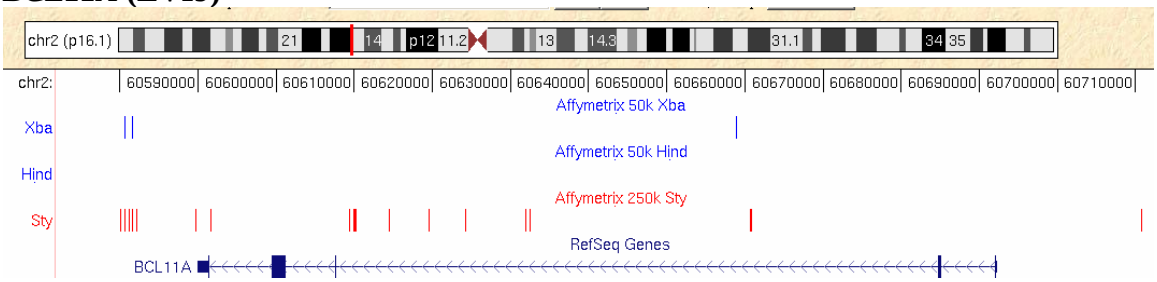
CD79A



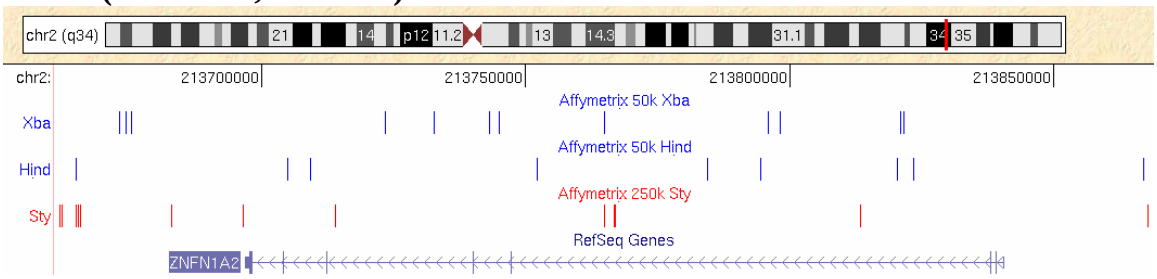
EBF1



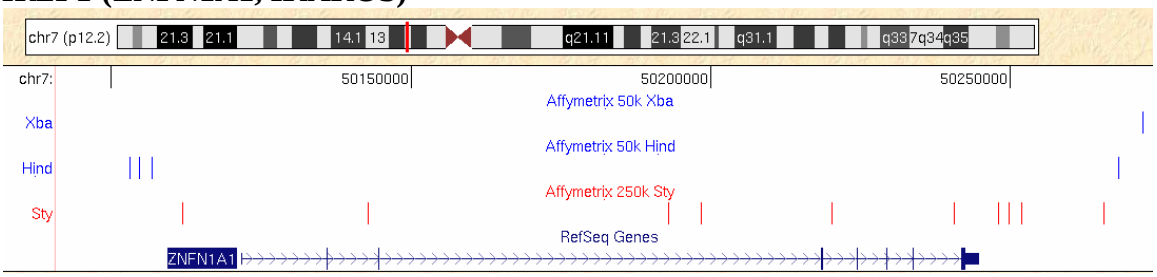
BCL11A (EVI9)



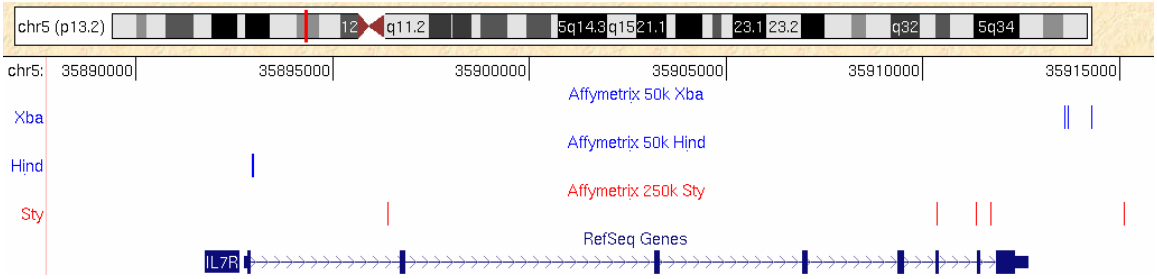
IKZF2 (ZNFN1A2, HELIOS)



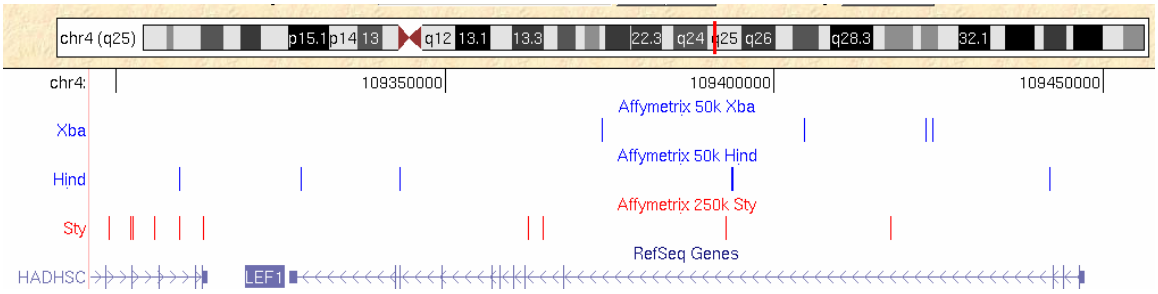
IKZF1 (ZNFN1A1, IKAROS)



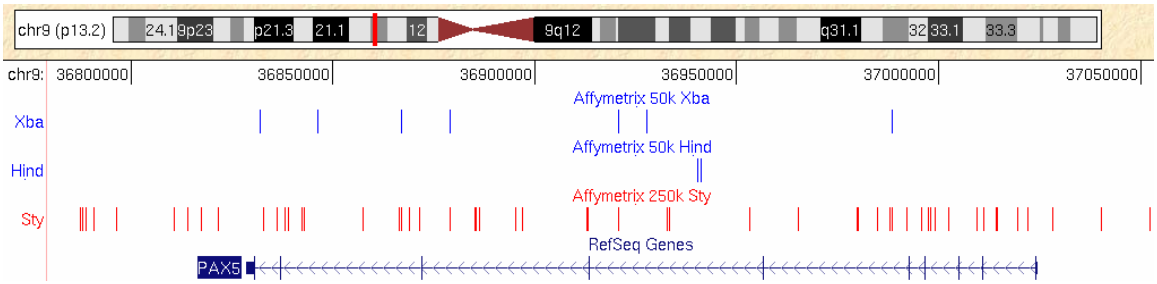
IL7R



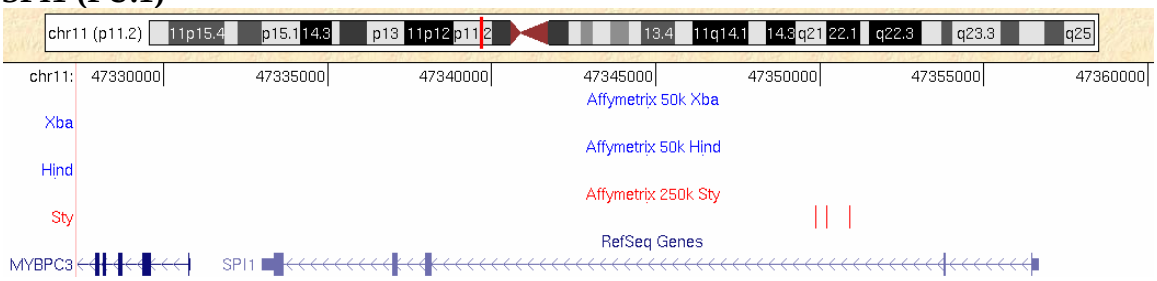
LEF1



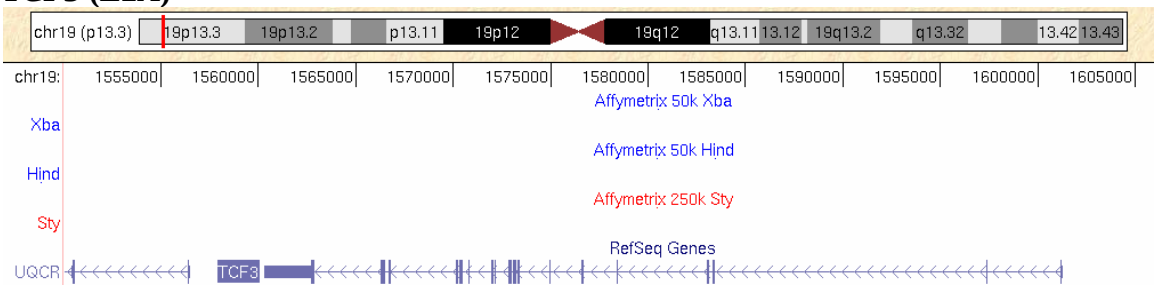
PAX5



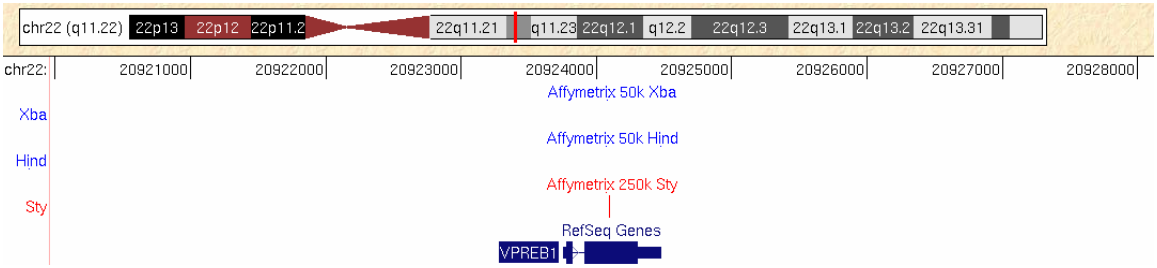
***SPI1* (PU.1)**

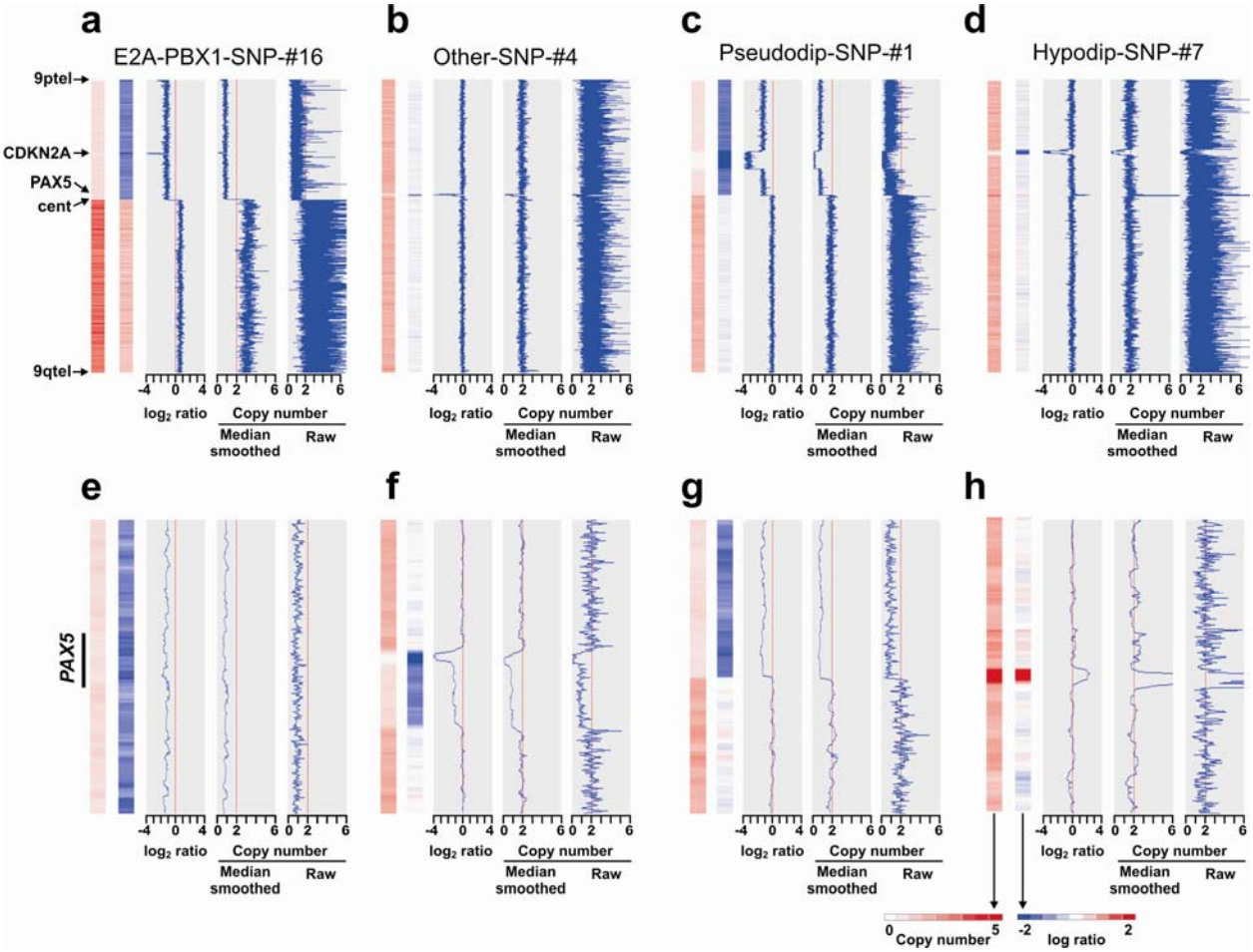


***TCF3* (E2A)**



VPREB1



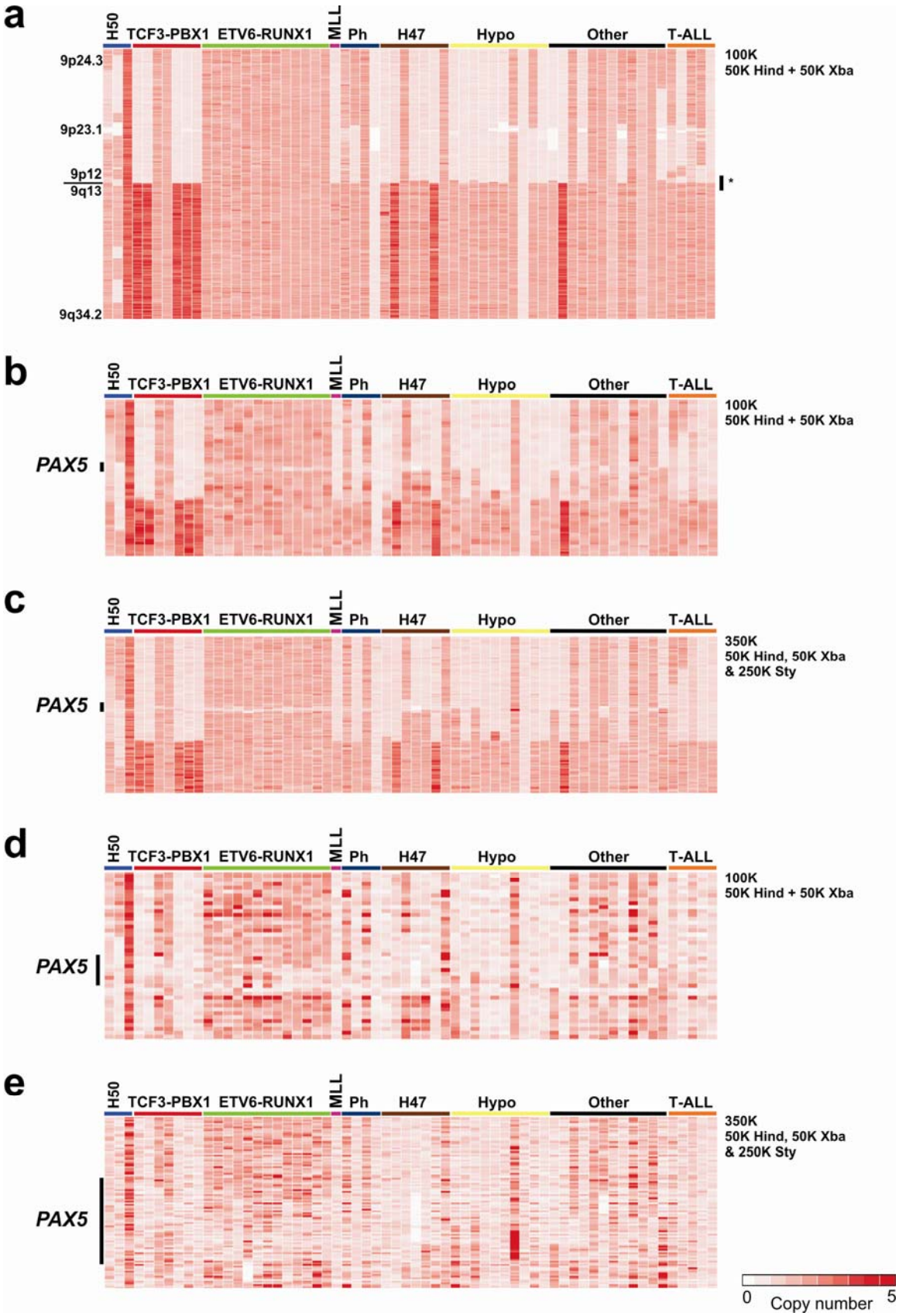


Supplementary Figure 12. Examples of chromosome 9 raw and smoothed copy number and $\log_2$ ratio data for *PAX5* deleted cases

Copy number heatmaps and $\log_2$ ratio and copy number plots are shown for four B-progenitor ALL cases with *PAX5* deletions. Panels **a-d** show all of chromosome 9; panels **e-h** show the region flanking *PAX5* for the same cases. For each case, two dChipSNP heat maps are shown: median smoothed copy number data (left panel, white-red scale) and smoothed log ratio data (panel second from left, blue-red scale). Two panels showing plots of copy number are shown; median smoothed (second from right) and raw (right).

Supplementary Figure 13. dChipSNP Copy number heatmaps of 62 ALL cases with *PAX5* deletion or amplification

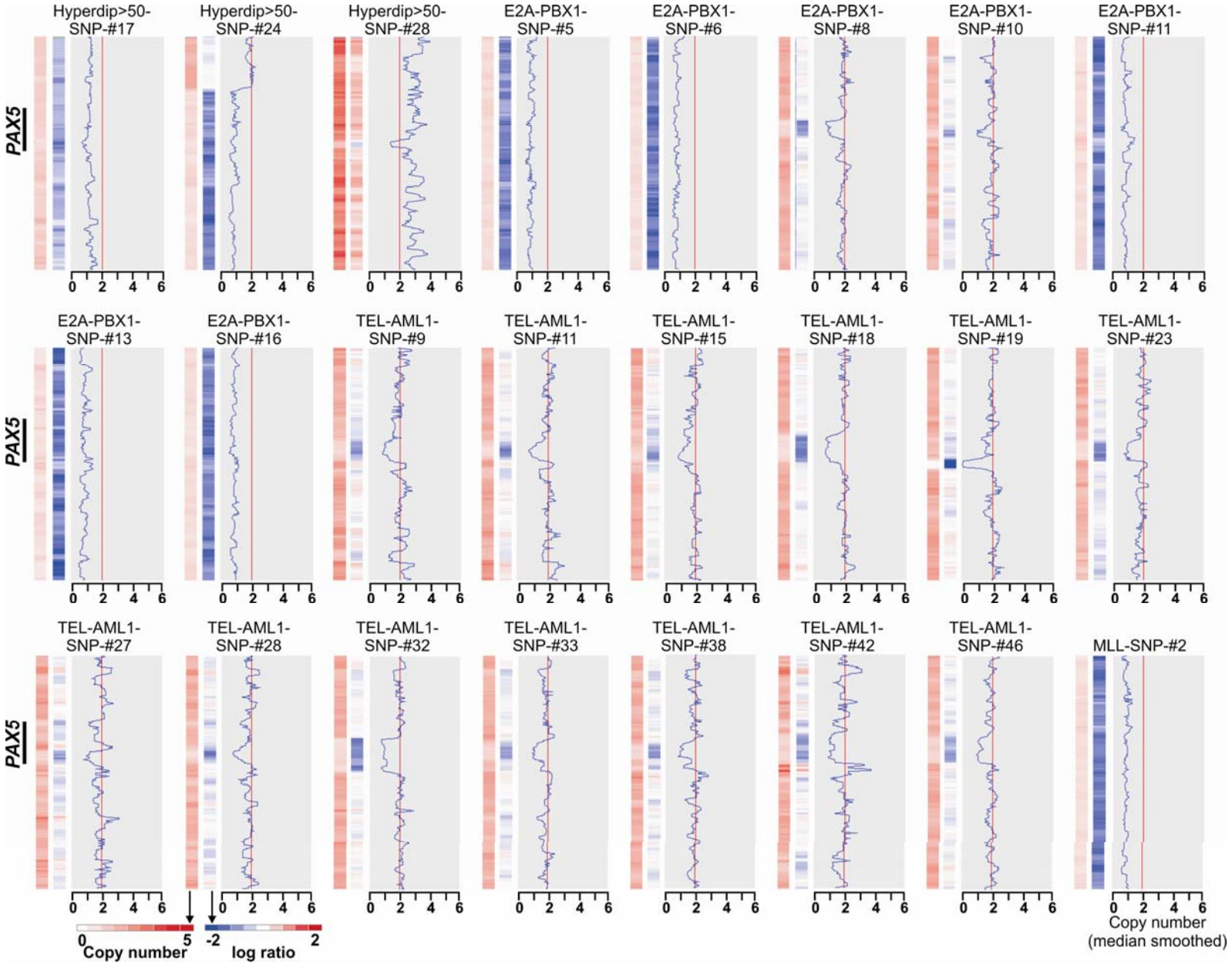
a, Chromosome 9 view of median-smoothed 100K (Hind & Xba) data of all 62 B-progenitor and T-ALL cases with copy number abnormalities involving *PAX5*. *indicates region shown in panel **b**. **b**, 100K median smoothed copy-number plots of the peri-centromeric region of chromosome 9 illustrating the different types of *PAX5* deletion: focal, broad, 9p extending to *PAX5* and all 9p extending to centromere. **c**, 350K (Hind, Xba & Sty) data of the same region as panel **b**, showing increased sensitivity of *PAX5* copy number abnormality detection using 350K data. **d**, Uninferred (raw) 100K copy number at *PAX5*. Several cases of PAX5 deletion and amplification involving few SNPs are evident in the *ETV6-RUNX1* group. **e**, Raw 350K copy number data illustrating higher density *PAX5* coverage and improved detection of *PAX5* copy number abnormalities. Abbreviations: H50, hyperdiploidy with greater than 50 chromosomes; Ph, *BCR-ABL1* positive ALL; Hypo, B-precursor ALL with hypodiploidy.


a
H50
TCF3-PBX1
ETV6-RUNX1
MLL
Ph
H47
Hypo
Other
T-ALL
9p24.3
9p23.1
9p12
9q13
9q34.2
100K
50K Hind + 50K Xba
b
PAX5
100K
50K Hind + 50K Xba
c
PAX5
350K
50K Hind, 50K Xba
& 250K Sty
d
PAX5
100K
50K Hind + 50K Xba
e
PAX5
350K
50K Hind, 50K Xba
& 250K Sty
0
Copy number
5

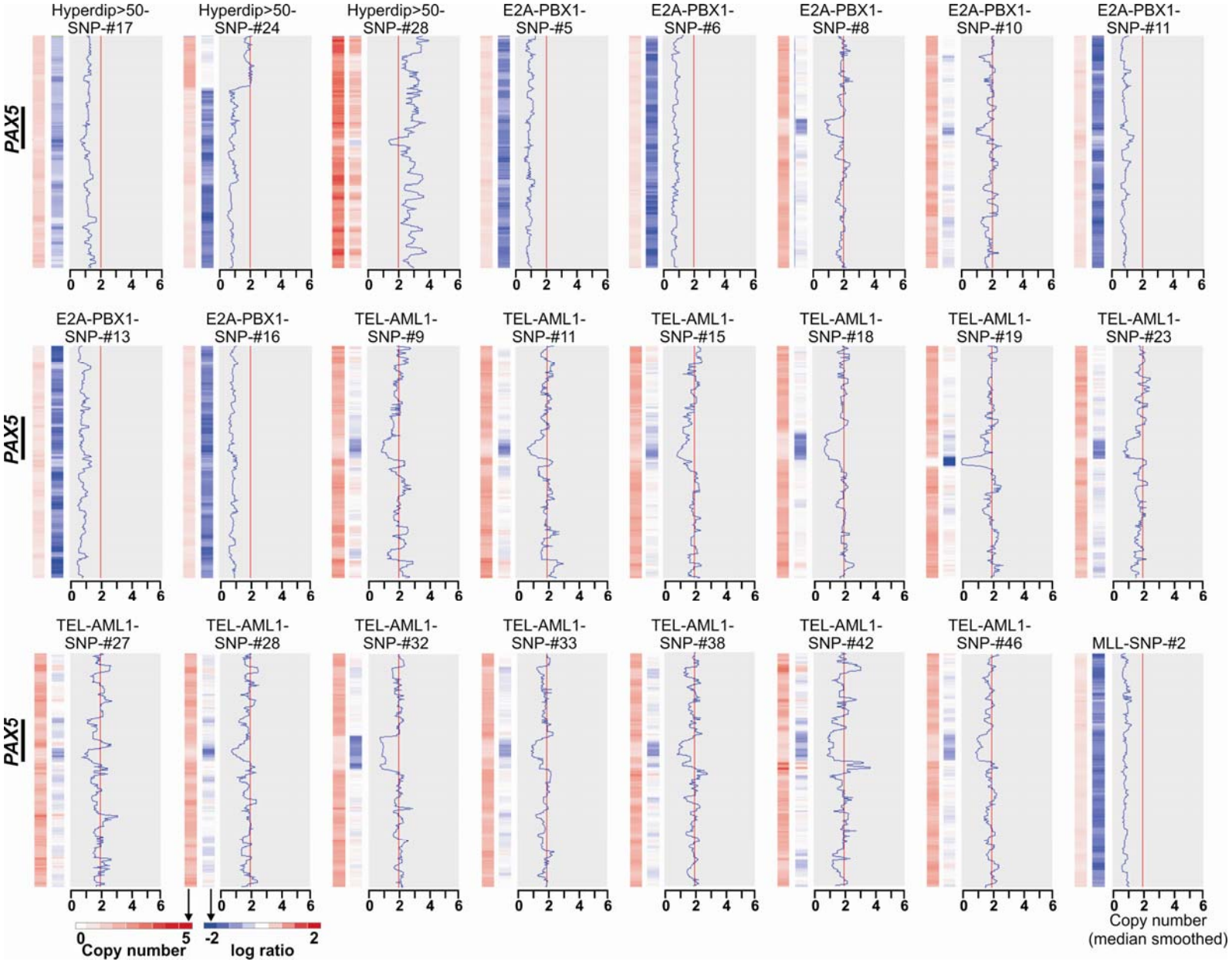
Supplementary Figure 14. Copy number heatmaps and plots for cases with *PAX5* copy number abnormalities

Three figures are shown per patient, from left to right: (1) white/red heat map of median smoothed copy number, with smoothing window of 10 SNPs; (2) blue (deletion) – red (amplification) heatmap of median smoothed log ratio copy number data, with smoothing window of 10 SNPs; and (3) corresponding plot of median smoothed copy number, on a scale of 0 – 6 copies. Normal (diploid) copy number is shown as a red line.

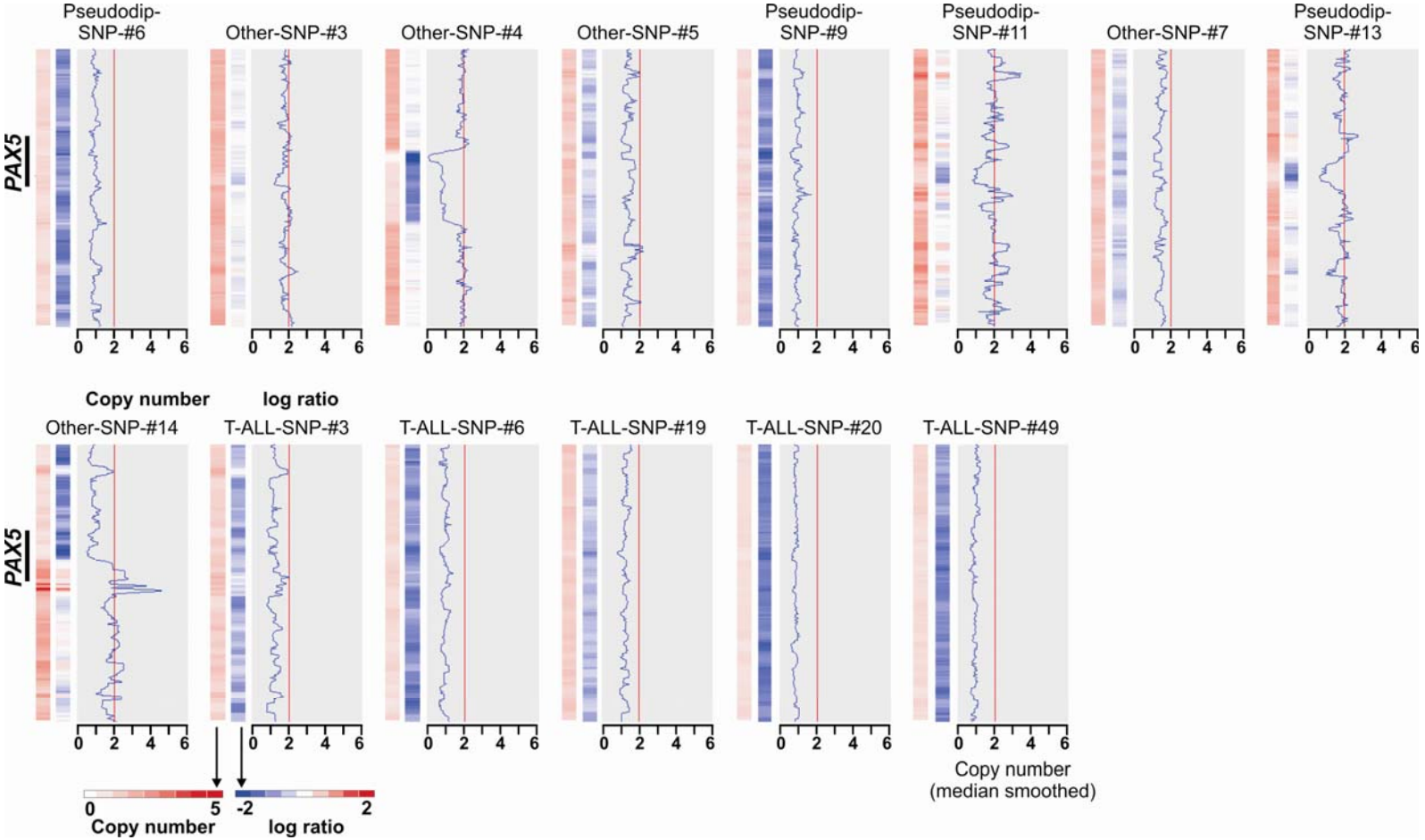
Supplementary Figure 14



Supplementary Figure 14

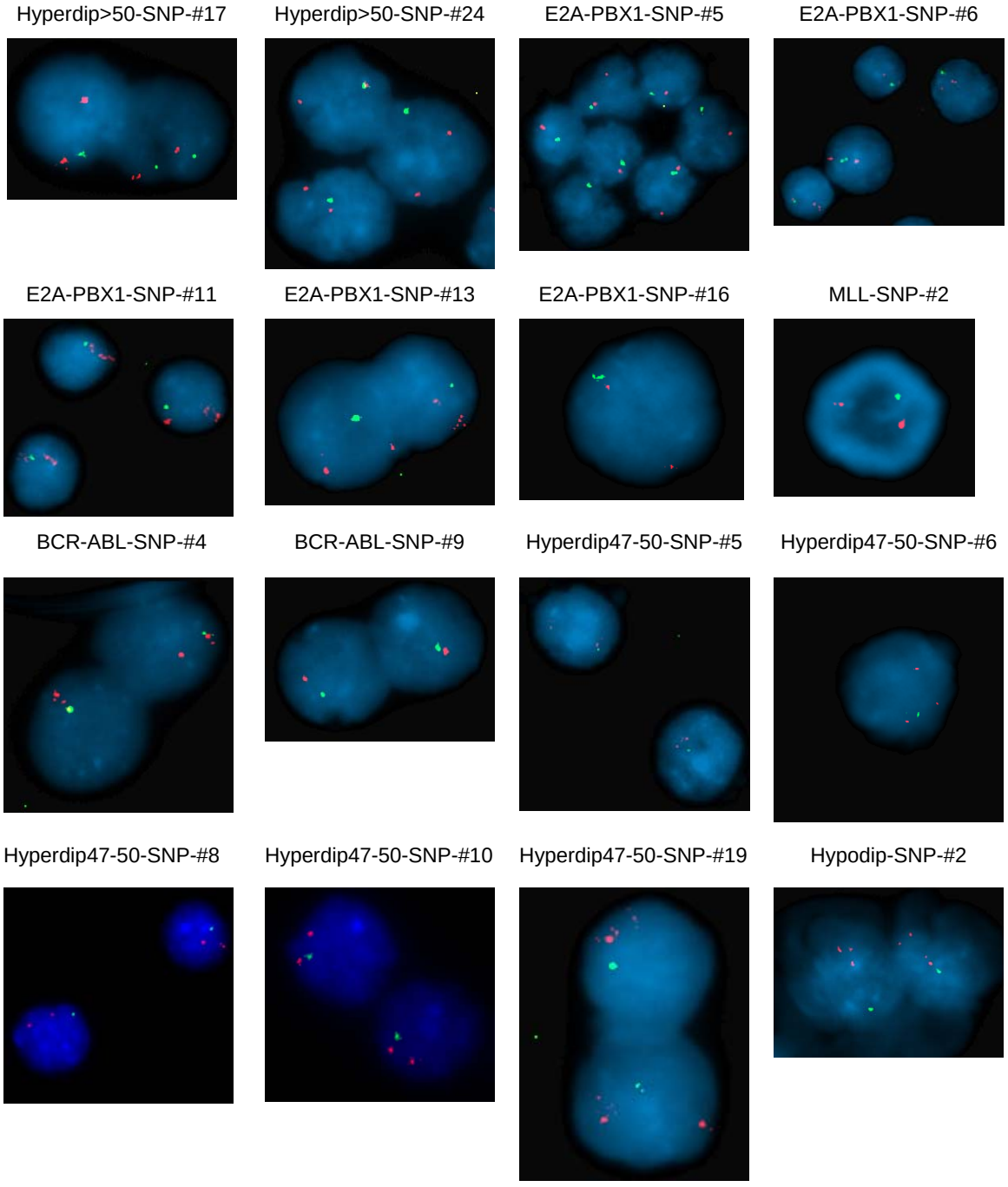


Supplementary Figure 14

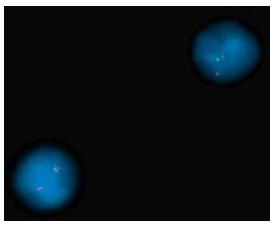


Supplementary Figure 15. Confirmation of *PAX5* deletions and amplification by FISH

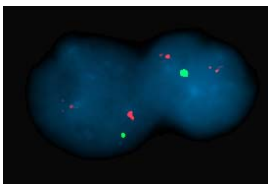
FISH confirms hemizygous *PAX5* deletion, and *PAX5* amplification in case Hypodip-SNP-#7. *PAX5* probe is labeled with FITC (green), and control probe rhodamine (red).



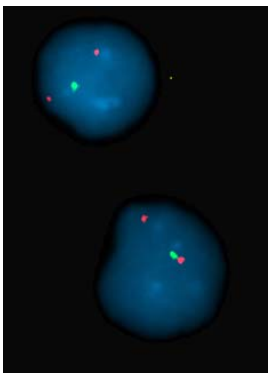
Hypodip-SNP-#4



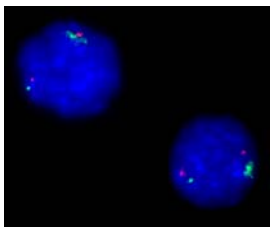
Hypodip-SNP-#5



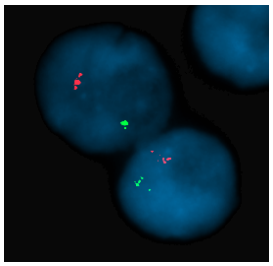
Hypodip-SNP-#6



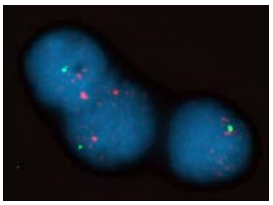
Hypodip-SNP-#7



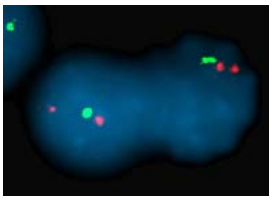
Hypodip-SNP-#8



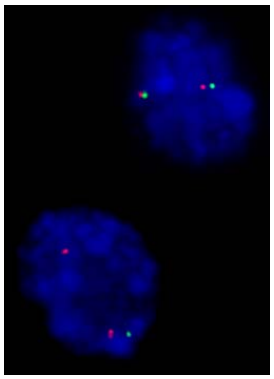
Hypodip-SNP-#9



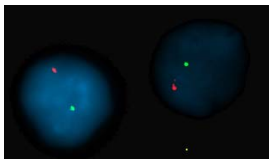
Hypodip-SNP-#10



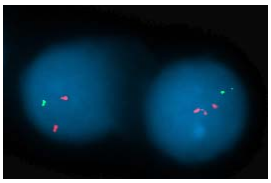
Other-SNP-#4



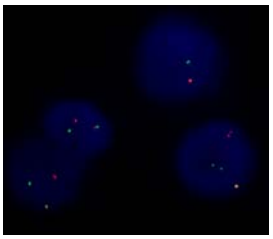
Other-SNP-#5



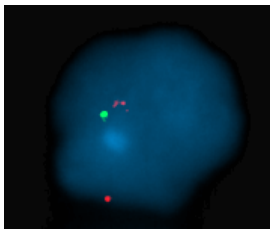
Other-SNP-#7



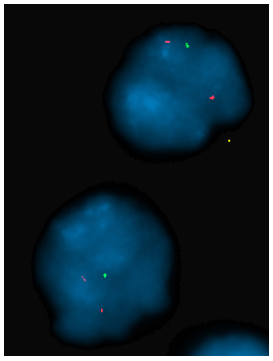
Pseudodip-SNP-#1



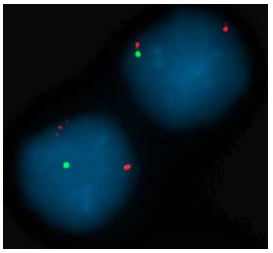
Pseudodip-SNP-#2



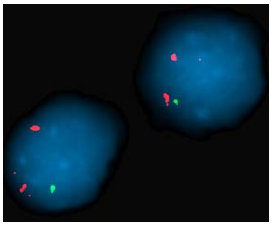
Pseudodip-SNP-#6



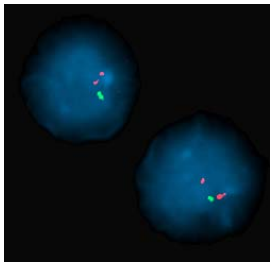
Pseudodip-SNP-#9



T-ALL-SNP-#3



T-ALL-SNP-#6



Supplementary Table 13. Quantitative FISH results for ALL cases with *PAX5* deletion, amplification and translocation

*Case known to have two populations of blasts on cytogenetic analysis

			FISH results (N of cells)					
Sample	Blast %	Region of PAX deletion	Monosomy 9	Hemizygous PAX5 deletion	Homozygous PAX5 deletion	Other PAX5 abnormality	normal cells	Cells with PAX5 deletion/fusion (%)
Hyperdip>50-SNP-#17	91	All gene	0	50	0		50	50*
Hyperdip>50-SNP-#24	97	All gene	0	96	0		4	96
E2A-PBX1-SNP-#5	98	All gene	34	64	0		2	98
E2A-PBX1-SNP-#6	97	All gene	0	97	0		3	97
E2A-PBX1-SNP-#11	97	All gene	0	0	100		0	100
E2A-PBX1-SNP-#13	95	All gene	0	100	0		0	100
E2A-PBX1-SNP-#16	94	All gene	0	93	0		7	93
MLL-SNP-#2	95	All gene	0	97	0		3	97
BCR-ABL-SNP-#4	95	All gene	0	94	0		6	94
BCR-ABL-SNP-#9	95	All gene	100	0	0		0	100
Hyperdip47-50-SNP-#5	97	All gene	0	88	0		12	88
Hyperdip47-50-SNP-#6	91	All gene	0	0	96		4	96
Hyperdip47-50-SNP-#8	95	e2-10		76			24	76
Hyperdip47-50-SNP-#9	97	All gene, homo e6-8		92			8	92
Hyperdip47-50-SNP-#10	96	e6-10				PAX5-ETV6 92	8	92
Hyperdip47-50-SNP-#19	97	All gene	0	84	0		0	100
Hypodip-SNP-#2	96	All gene	0	91	1		8	92
Hypodip-SNP-#3	88	e8-10				PAX5-ZNF521 75	25	75
Hypodip-SNP-#4	98	All gene	0	94	0		6	94
Hypodip-SNP-#5	93	All gene	0	85	1		14	86
Hypodip-SNP-#6	94	All gene	0	96	0		4	96
Hypodip-SNP-#7	93	e2-e5				Amplified 72	28	72
Hypodip-SNP-#8	91	All gene	74	0	0		26	74
Hypodip-SNP-#9	99	All gene	0	96	0		4	96
Hypodip-SNP-#10	98	All gene	0	92	0		8	92
Other-SNP-#4	97	All gene, homo e8		18			18	82
Other-SNP-#5	57	All gene	22	2	0		26	48
Other-SNP-#7	93	All gene	0	93	1		6	94
Other-SNP-#14	98	e8-tel				PAX5-FOXP1 87	13	87

			FISH results (N of cells)					
Sample	Blast %	Region of PAX deletion	Monosomy 9	Hemizygous PAX5 deletion	Homozygous PAX5 deletion	Other PAX5 abnormality	normal cells	Cells with PAX5 deletion/fusion (%)
Pseudodip-SNP-#1	95	e6 -tel				PAX5-ETV6, 97	3	97
Pseudodip-SNP-#2	100	All gene	0	100	0		0	100
Pseudodip-SNP-#6	98	All gene	0	94	0		6	94
Pseudodip-SNP-#9	98	All gene	0	95	0		5	95
T-ALL-SNP-#3	50	All gene	0	76	0		24	76
T-ALL-SNP-#6	95	All gene	0	89	2		9	91
T-ALL-SNP-#19	95	All gene	0	62	0		38	62
T-ALL-SNP-#20	95	All gene	0	96	2		2	98
T-ALL-SNP-#49	76	All gene	0	92	0		8	92
Control samples								
TEL-AML1-SNP-#22	97	None	0	2	0		98	2
E2A-PBX1-SNP-#9	92	None	0	1	0		99	1
Pseudodip-SNP-#10	99	None	0	6	0		94	6
T-ALL-SNP-#28	98	None	0	0	0		88	0
TEL-AML1-SNP-#25	93	None	0	2	1		97	3
TEL-AML1-SNP-#39	96	None	0	2	0		98	2

Supplementary Table 14. *PAX5* genomic quantitative PCR results

PAX5 target to control ratios of less than 0.7 and 0.3 were used as thresholds for hemizygous and homozygous deletion, respectively. The means of duplicate assays are shown. *Regions of *PAX5* deletion are defined by 350K SNP array analyses. e, exon.

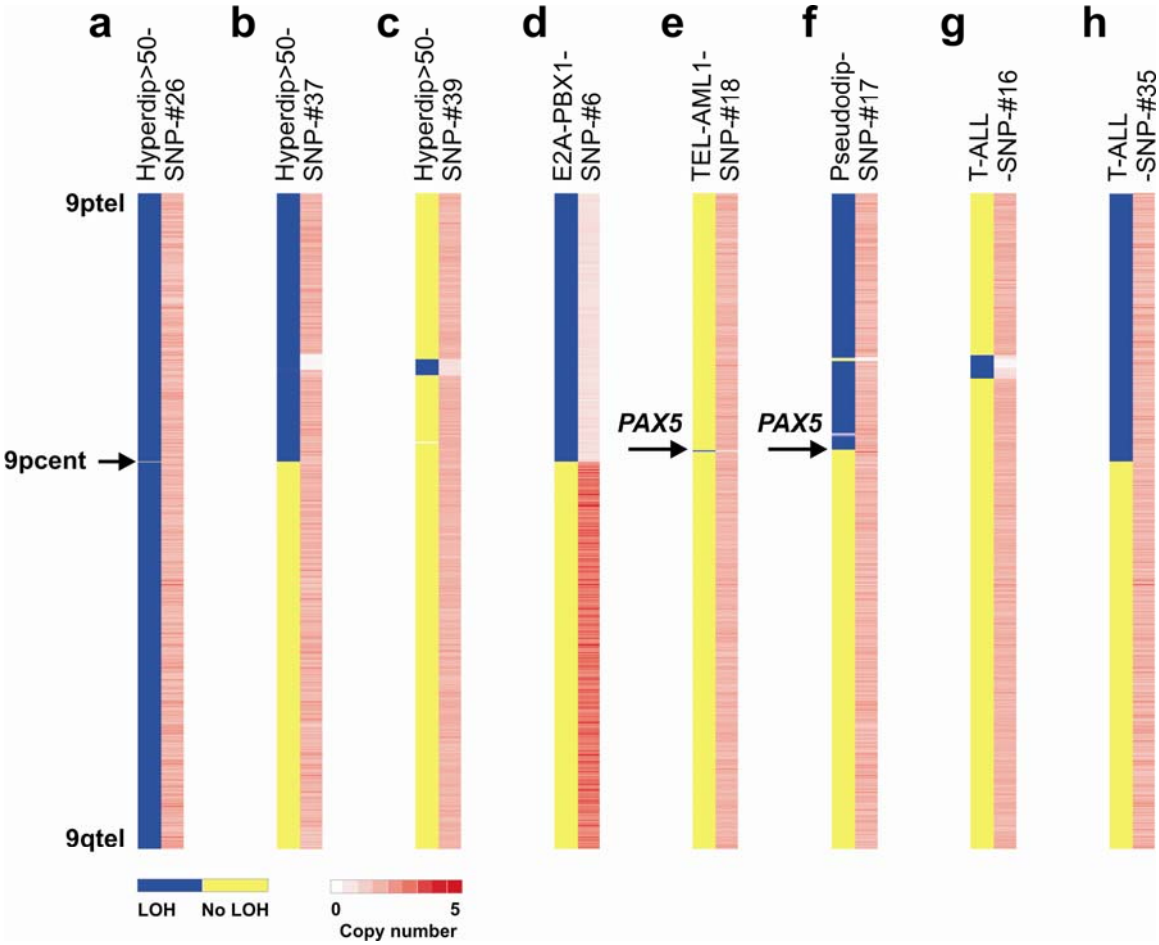
Sample	Blast %	Region of *PAX5* deletion*	*PAX5* exon 3 / Rnase P ratio	*PAX5* exon 6 / Rnase P ratio	*PAX5* exon 8 / Rnase P ratio
Hyperdip>50-SNP-#17	91	All gene	0.49	0.55	0.58
Hyperdip>50-SNP-#24	97	All gene	0.38	0.34	0.40
Hyperdip>50-SNP-#28	76	e2-e5	0.74	0.94	1.04
E2A-PBX1-SNP-#5	98	All gene	0.57	0.58	0.59
E2A-PBX1-SNP-#6	97	All gene	0.43	0.39	0.48
E2A-PBX1-SNP-#8	95	e6-e8	1.09	0.65	0.55
E2A-PBX1-SNP-#10	98	e6-7	1.10	0.68	0.95
E2A-PBX1-SNP-#11	97	All gene	0.65	0.66	0.68
E2A-PBX1-SNP-#13	95	All gene	0.17	0.15	0.19
E2A-PBX1-SNP-#16	94	All gene	0.66	0.65	0.68
TEL-AML1-SNP-#11	98	e2-e6	0.58	0.48	1.16
TEL-AML1-SNP-#15	99	e2-e3	0.64	0.97	1.03
TEL-AML1-SNP-#18	94	e2-e8	0.64	0.60	0.61
TEL-AML1-SNP-#19	99	5' to e1	1.07	1.10	1.08
TEL-AML1-SNP-#23	98	5'-e6	0.34	0.36	0.77
TEL-AML1-SNP-#27	97	e2-e5	0.62	0.89	0.98
TEL-AML1-SNP-#28	98	e2-e6	0.58	0.46	0.9
TEL-AML1-SNP-#32	98	5'-e8	0.58	0.59	0.58
TEL-AML1-SNP-#33	99	e2-e6	0.58	0.50	1.04
TEL-AML1-SNP-#38	97	e2-e6	0.58	0.56	0.98
TEL-AML1-SNP-#42	84	e2-e8	0.73	0.63	0.65
TEL-AML1-SNP-#46	88	e2-e8	0.68	0.67	0.44
TEL-AML1-SNP-#9	91	e2-e7	0.65	0.67	1.20
MLL-SNP-#2	95	All gene	0.62	0.62	0.66
BCR-ABL-SNP-#1	79	e2-e6	0.42	0.34	0.73
BCR-ABL-SNP-#4	95	All gene	0.64	0.61	0.52
BCR-ABL-SNP-#7	79	e2-e6	0.68	0.61	1.1
BCR-ABL-SNP-#9	95	All gene	0.20	0.19	0.22
Hyperdip47-50-SNP-#10	96	e6-10	1.26	0.47	0.45
Hyperdip47-50-SNP-#19	97	All gene	0.72	0.59	0.63
Hyperdip47-50-SNP-#20	99	e2-e8	0.58	0.59	0.65
Hyperdip47-50-SNP-#5	97	All gene	0.59	0.47	0.52
Hyperdip47-50-SNP-#6	91	All gene	0.59	0.46	0.49
Hyperdip47-50-SNP-#8	95	e2-10	0.66	0.46	0.47
Hyperdip47-50-SNP-#9	97	All gene, homozygous e6-8	0.38	0.03	0.04
Hypodip-SNP-#1	94	e7-distal	0.85	0.89	0.57
Hypodip-SNP-#10	98	All gene	0.45	0.44	0.49
Hypodip-SNP-#26	96	All gene	0.52	0.53	0.55
Hypodip-SNP-#3	88	e8-10	1.32	1.10	0.62
Hypodip-SNP-#4	98	All gene	0.65	0.58	0.65

Sample	Blast %	Region of *PAX5* deletion*	*PAX5* exon 3 / Rnase P ratio	*PAX5* exon 6 / Rnase P ratio	*PAX5* exon 8 / Rnase P ratio
Hypodip-SNP-#5	93	All gene	0.59	0.46	0.47
Hypodip-SNP-#6	94	All gene	0.57	0.48	0.49
Hypodip-SNP-#7	93	e2-e5 amplified	4.01	1.21	1.15
Hypodip-SNP-#8	91	All gene	0.56	0.55	0.64
Hypodip-SNP-#9	99	All gene	0.45	0.45	0.50
Other-SNP-#14	98	e8-tel	0.91	0.92	0.53
Other-SNP-#3	99	e2-e3	0.57	0.86	0.98
Other-SNP-#4	97	All gene, homozygous e8	0.58	0.43	0.05
Other-SNP-#5	57	All gene	0.53	0.51	0.49
Other-SNP-#7	93	All gene	0.67	0.55	0.74
Pseudodip-SNP-#1	95	e6 - tel	0.77	0.31	0.36
Pseudodip-SNP-#11	92	e2-e6	0.40	0.33	0.81
Pseudodip-SNP-#13	97	e2-e6	0.62	0.62	1.14
Pseudodip-SNP-#2	100	All gene	0.54	0.53	0.55
Pseudodip-SNP-#4	90	e2-e6	0.56	0.61	1.23
Pseudodip-SNP-#6	98	All gene	0.57	0.44	0.51
Pseudodip-SNP-#9	98	All gene	0.39	0.39	0.45
T-ALL-SNP-#19	95	All gene	0.45	0.36	0.45
T-ALL-SNP-#20	95	All gene	0.50	0.46	0.52
T-ALL-SNP-#3	50	All gene	0.59	0.50	0.56
T-ALL-SNP-#49	76	All gene	0.41	0.30	0.46
T-ALL-SNP-#6	95	All gene	0.57	0.63	0.59
Control samples					
TEL-AML1-SNP-#10	96	None	1.11	1.07	1.04
MLL-SNP-#1	97	None	0.96	0.93	1.05
BCR-ABL-SNP-#3	77	None	1.00	0.86	0.91
T-ALL-SNP-#1	94	None	1.04	0.97	1.16
T-ALL-SNP-#11	88	None	1.16	0.86	1.03
T-ALL-SNP-#14	94	None	1.04	0.93	0.95
Hyperdip47-50-SNP-#12	97	None	0.94	0.79	0.87

ANALYSIS OF LOSS-OF-HETEROZYGOSITY IN PAEDIATRIC ALL

Loss-of-heterozygosity analysis for 228 ALL samples with SNP array data for corresponding remission blood or marrow samples was also performed using dChipSNP. Several patterns of LOH were identified (Supplementary Figure 16), including (1) corresponding regions of LOH and deletion (Supplementary Figure 16c-e, g); (2) broad regions of predominantly copy-neutral LOH harbouring a focal region of homozygous deletion (Supplementary Figure 16b, f), and copy-neutral LOH with no focal copy number abnormalities (Supplementary Figure 16a, h). Most *PAX5* deletions were accompanied by corresponding regions of LOH, with the exception of the several cases of focal *PAX5* deletion. In these cases, the low number of informative (heterozygous) SNPs in the germline sample in the *PAX5*-deleted region precluded an inference of LOH by dChipSNP. Twenty cases with broad regions of copy-neutral LOH involving chromosome 9p or all of chromosome 9 were identified. In 18 cases, the region of LOH also harboured a focal homozygous deletion, most commonly at 9p21.3 (*CDKN2A*), suggesting deletion of one copy of *CDKN2A*, deletion of the wild-type 9p, and subsequent reduplication of the 9p harbouring the *CDKN2A* deletion. Importantly, only two cases with LOH involving chromosome 9 but no focal deletion in the region of LOH were identified (Supplementary Figure 16a, h), neither of which harboured *PAX5* mutations. No copy neutral LOH involving *EBF1* was observed.

A listing of all regions of copy-neutral LOH is provided in Supplementary Table 15.



Supplementary Figure 16. Chromosome 9 loss-of-heterozygosity in ALL

Representative paired LOH and copy number analyses in ALL. Each tumor sample is represented by a pair of columns with LOH on the left, and copy number on the right. **a**, copy-neutral LOH involving all of chromosome 9, with no detectable copy number abnormality; **b**, copy-neutral LOH of 9p, with focal homozygous deletion at 9p21.3 (harboring *CDKN2A*); **c**, 9p21.3 (*CDKN2A*) LOH and hemizygous deletion; **d**, 9p LOH and hemizygous 9p-, as well as 9q+; **e**, focal LOH and corresponding hemizygous deletion at 9p13.2 (*PAX5*); **f**, copy-neutral LOH extending from 9ptel to 9p13.2 (*PAX5)*, with focal homozygous 9p21.3 (*CDKN2A*) deletion; this case also is homozygous for the *PAX5* IVS9+1 mutation; **g**, focal 9p21.3 LOH and corresponding mixed hemizygous/homozygous deletion; and **h**, copy neutral 9p LOH with no detectable copy number abnormality in the region.

Supplementary Table 15. Regions of copy-neutral LOH in paediatric ALL

The table lists all regions of copy neutral loss-of-heterozygosity identified by dChipSNP using the Hidden Markov Model algorithm. Focal regions containing less than three SNPs showing LOH are excluded. The majority of regions contain a focal region of copy number change (most commonly focal homozygous deletions of *CDKN2A* at 9p21.3), or are adjacent to regions of copy number abnormality.

Case	Chr	Cytoband	Start	Stop	Size (Mb)	Comment
E2A-PBX1-SNP-#10	1	1q21.2-23.3	146.721	161.518	14.797	Adjacent to partial duplication of 1q from *PBX1*
Hyperdip>50-SNP-#12	1	1q43-qtel	236.025	245.353	9.328	Adjacent to partial duplication of 1q
T-ALL-SNP-#12	1	1p36.33-1p36.32	0.792	4.119	3.327	Adjacent to region of matching LOH and deletion
T-ALL-SNP-#4	1	1p36.22-1p12	0.531	120.009	119.478	
Hyperdip>50-SNP-#27	2	All chromosome 2				
Hyperdip>50-SNP-#35	2	All chromosome 2				
Hyperdip47-50-SNP-#19	3	3q13.13-3qtel	110.114	198.496	88.382	Region contains homozygous *BCL6* deletion
BCR-ABL-SNP-#9	3	3q21.3-3qtel	129.454	198.496	69.042	
Hyperdip>50-SNP-#8	4	All chromosome 4				
E2A-PBX1-SNP-#3	6	6p22.1-6ptel	0.1	27.775	27.675	
TEL-AML1-SNP-#10	6	6p21.1-6ptel	0.1	45.029	44.929	
Hyperdip47-50-SNP-#22	6	6q22.1-6qtel	114.8	170.608	55.808	
Hyperdip>50-SNP-#34	6	6p22.1-6ptel	0.1	27.948	27.848	Region includes 2 regions of homozygous deletions involving histone genes
Pseudodip-SNP-#8	6	6p21.31-6p12.3	35.226	49.853	14.627	
Hyperdip>50-SNP-#26	8	8q23.3-qtel	116.02	145.985	29.965	
Hyperdip>50-SNP-#3	8	All chromosome 8				
E2A-PBX1-SNP-#10	9	9p13.3-9ptel	0.031	34.495	34.464	Region includes focal homozygous 9p21.3 deletion (including *CDKN2A*)
TEL-AML1-SNP-#29	9	9p13.2-9ptel	0.031	36.489	36.458	Region includes large 9p21.2-3 deletion
TEL-AML1-SNP-#43	9	9p21.1-9ptel	0.031	28.875	28.844	Region includes focal homozygous deletion at *LRRN6C*
TEL-AML1-SNP-#7	9	9p13.3-9ptel	0.031	34.031	34	Region includes focal homozygous 9p21.3 deletion (including *CDKN2A*)
Hyperdip47-50-SNP-#22	9	9p24.3-9p13.3	0.215	34.404	34.189	Region includes large homozygous 9p21.3 deletion (including *CDKN2A*)
Hyperdip>50-SNP-#26	9	All chromosome 9				
Hyperdip>50-SNP-#37	9	All 9p	0.031	44.109	44.078	Region includes large homozygous 9p21.3 deletion

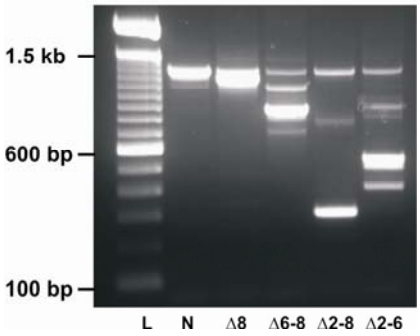
Case	Chr	Cytoband	Start	Stop	Size (Mb)	Comment
						(including *CDKN2A*)
Hyperdip>50-SNP-#4	9	All chromosome 9				Region includes focal homozygous 9p21.3 deletion (including *CDKN2A*)
Pseudodip-SNP-#12	9	9ptel-9p24.2	0.031	2.488	2.457	Adjacent to complex deletion
Pseudodip-SNP-#15	9	9ptel-9p13.2	0.031	37.322	37.291	Region extends centromeric of PAX5 to ZCCHC7, and has 9p21.3 deletion (including *CDKN2A*)
Pseudodip-SNP-#17	9	9ptel-9p13.2	0.031	36.948	36.917	Region ends at PAX5, and has focal homozygous 9p21.3 deletion (including *CDKN2A*). This case also has a homozygous *PAX5* IVS9+1 point mutation
T-ALL-SNP-#14	9	9ptel-9p21.1	0.031	32.527	32.496	Region includes focal homozygous 9p21.3 deletion (including *CDKN2A*)
T-ALL-SNP-#21	9	9p24.3-9p21.1	0.398	29.428	29.03	Region includes focal homozygous 9p21.3 deletion (including *CDKN2A*)
T-ALL-SNP-#23	9	9ptel-9p21.3	0.031	22.319	22.288	Region includes focal homozygous 9p21.3 deletion (including *CDKN2A*)
T-ALL-SNP-#24	9	9p24.3-9p13.3	0.508	35.219	34.711	Region includes focal homozygous 9p21.3 deletion (including *CDKN2A*)
T-ALL-SNP-#31	9	9ptel-9p13.3	0.031	34.933	34.902	Region includes homozygous 9p21.3-22.1 deletion (including *CDKN2A*)
T-ALL-SNP-#35	9	9ptel-9p13.2	0.031	38.479	38.448	
T-ALL-SNP-#45	9	9ptel-9p21.3	0.031	24.84	24.809	Region includes homozygous 9p21.3 deletion (including *CDKN2A*)
T-ALL-SNP-#7	9	All 9p	0.031	44.109	44.078	Region includes homozygous 9p21.3 deletion (including *CDKN2A*)
T-ALL-SNP-#9	9	9ptel-9p13.2	0.031	38.3	38.269	Complex deletion: heterozygous at 9p13.2 (telomeric of PAX5), then deletion from 9p21.1-9p21.3 that is homozygous at 9p21.1 and 9p21.3 (2 regions)
T-ALL-SNP-#23	10	10q11.21-10qtel	44.752	135.353	90.601	Region includes focal homozygous *PTEN* deletion at 10q23.31
T-ALL-SNP-#9	10	10p15.3-10p14	0.102	7.432	7.33	Region is adjacent to large deletion
TEL-AML1-SNP-#35	11	11p15.1-ptel	0.197	18.595	18.398	
TEL-AML1-SNP-#20	12	12p13.2-12ptel	0.037	11.925	11.888	Region is adjacent to large deletion from *ETV6* to 12p11.2; the junction of the regions of copy-neutral LOH and deletion is *ETV6*
T-ALL-SNP-#11	12	12p13.33-12p12.1	0.062	21.627	21.565	
Hyperdip>50-SNP-#3	13	13q21.33-qtel	69.186	113.531	44.345	Rest of chromosome shows deletion and LOH
T-ALL-SNP-#35	13	All chromosome 13				
Hyperdip>50-SNP-#21	15	15q12-qtel	23.666	99.887	76.221	Adjacent to gain
Hyperdip>50-SNP-#24	15	All chromosome 15				

Case	Chr	Cytoband	Start	Stop	Size (Mb)	Comment
Hyperdip47-50-SNP-#17	16	16q21-16qtel	59.176	88.691	29.515	
Hyperdip>50-SNP-#39	16	All chromosome 16				
Hyperdip>50-SNP-#17	17	All chromosome 17				Deletion at 17q23.3-q24.1 (57.360-59.959Mb)
Other-SNP-#15	17	17ptel-17p11.2	0.007	20.463	20.456	
T-ALL-SNP-#35	18	18q21.32-18qtel	55.752	76.116	20.364	
Hyperdip>50-SNP-#21	19	All chromosome 19				Deletion at 19p13.3 (0.229-2.89Mb)
Hyperdip>50-SNP-#26	19	19q13.32-qtel	52.139	63.611	11.472	
Hyperdip>50-SNP-#27	19	All chromosome 19				
Hyperdip>50-SNP-#5	19	19q13.32-qtel	51.712	63.611	11.899	
TEL-AML1-SNP-#42	20	20q11.21-qtel	29.31	62.192	32.882	Rest of chromosome gained
Hyperdip>50-SNP-#39	20	All chromosome 20				
Pseudodip-SNP-#11	20	20q11.2-20qtel	34.166	62.377	28.211	
T-ALL-SNP-#11	20	20q12-20qtel	39.747	62.377	22.63	

Supplementary Table 16. Internal deletions in PAX5 in B-progenitor ALL

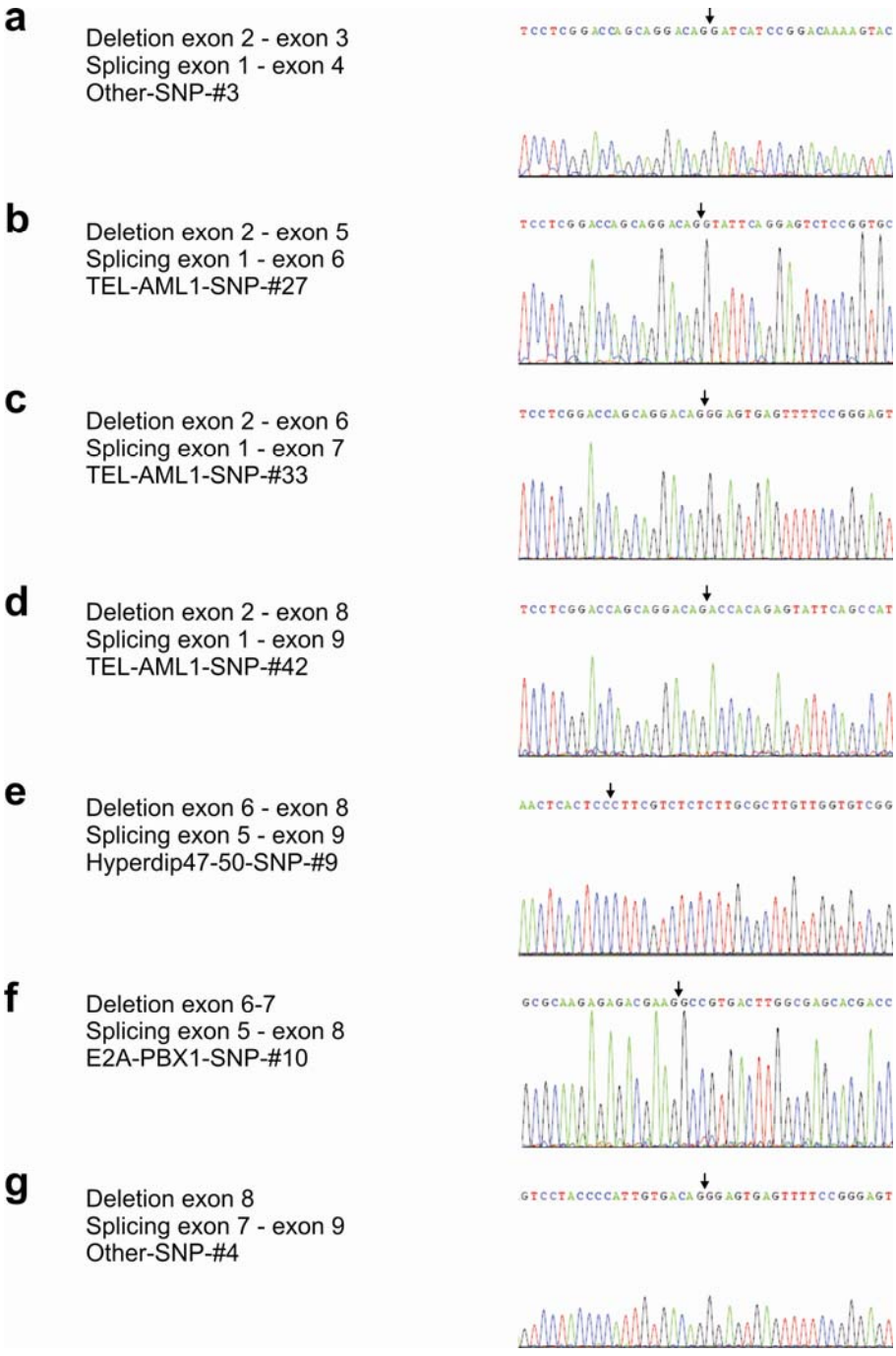
aa, amino acids.

Deletion	Number of cases	Effect on PAX5 coding sequence	Predicted protein size and domains
exons 2-3	2	Frameshift, stop in exon 4	36 aa, lacks all functional domains
exons 2-5	3	Frameshift, stop in exon 7	57 aa, lacks all functional domains
exons 2-6	9	Frameshift, stop in exon 7	42 aa, lacks all functional domains
exons 2-7	1	In frame	103 aa, lacks paired, octapeptide and homeodomain-like domains
exons 2-8	4	In frame	69 aa, lacks paired, octapeptide, homeodomain-like and activating part of transactivation domain
exons 6-8	2	In frame	255 aa, lacks homeodomain like and activating part of transactivation domain
exon 6-7	1	In frame	289 amino acids, lacks transactivating
exon 8	1	In frame	357 aa, lacks activating part of transactivation domain



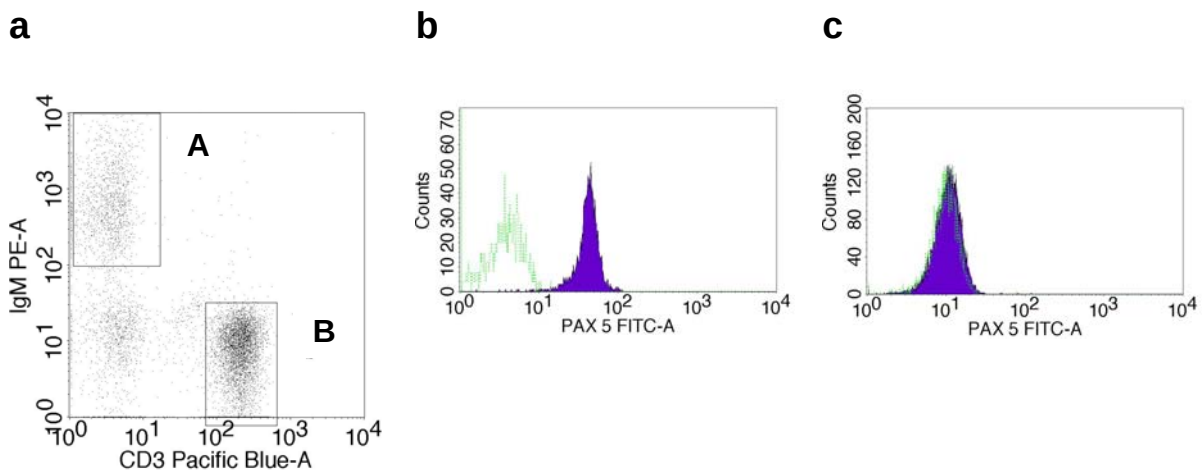
Supplementary Figure 17. RT-PCR demonstrates internally deleted transcripts in cases with focal *PAX5* deletions.

PAX5 RT-PCR. PCR primers C282 and C302 (Supplementary Table 7) amplify the entire coding region of *PAX5* (exon 1a isoform). Size of the predominant transcript corresponds to the extent of deletion (Δ) on SNP array analysis. L, ladder; N, *PAX5* wild type.



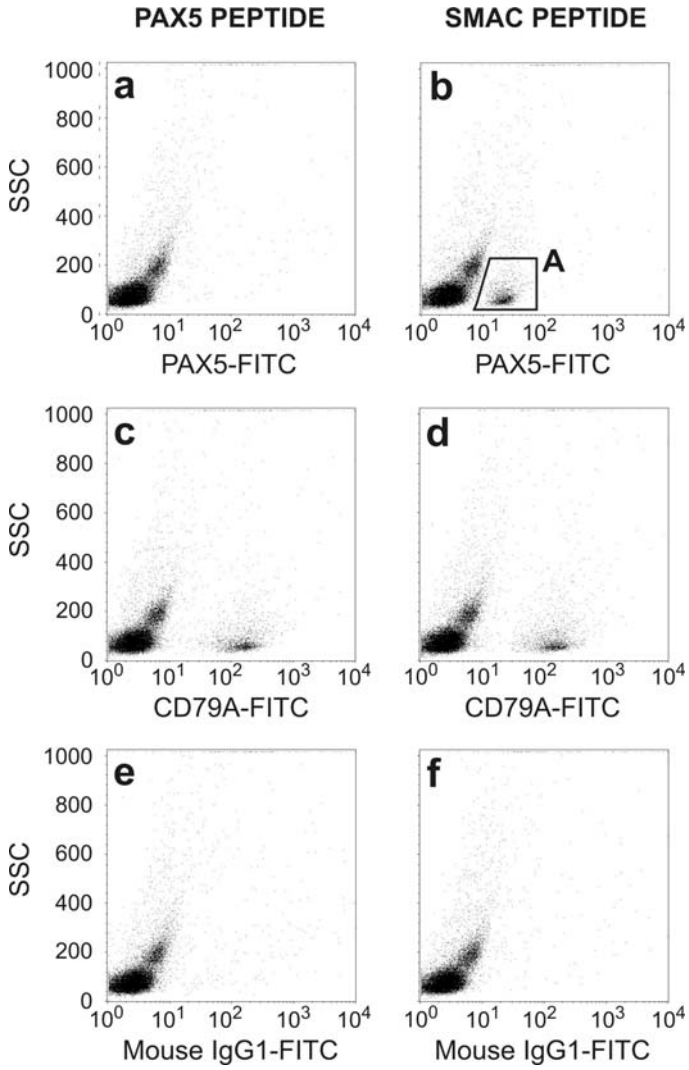
Supplementary Figure 18. Sequencing pherograms confirming aberrant *PAX5* splicing in cases with internal *PAX5* deletions

Representative pherograms of PAX5 exon 1a variant alleles amplified by RT-PCR (primers C282 and C302) and sequenced, showing splicing across the deleted region.



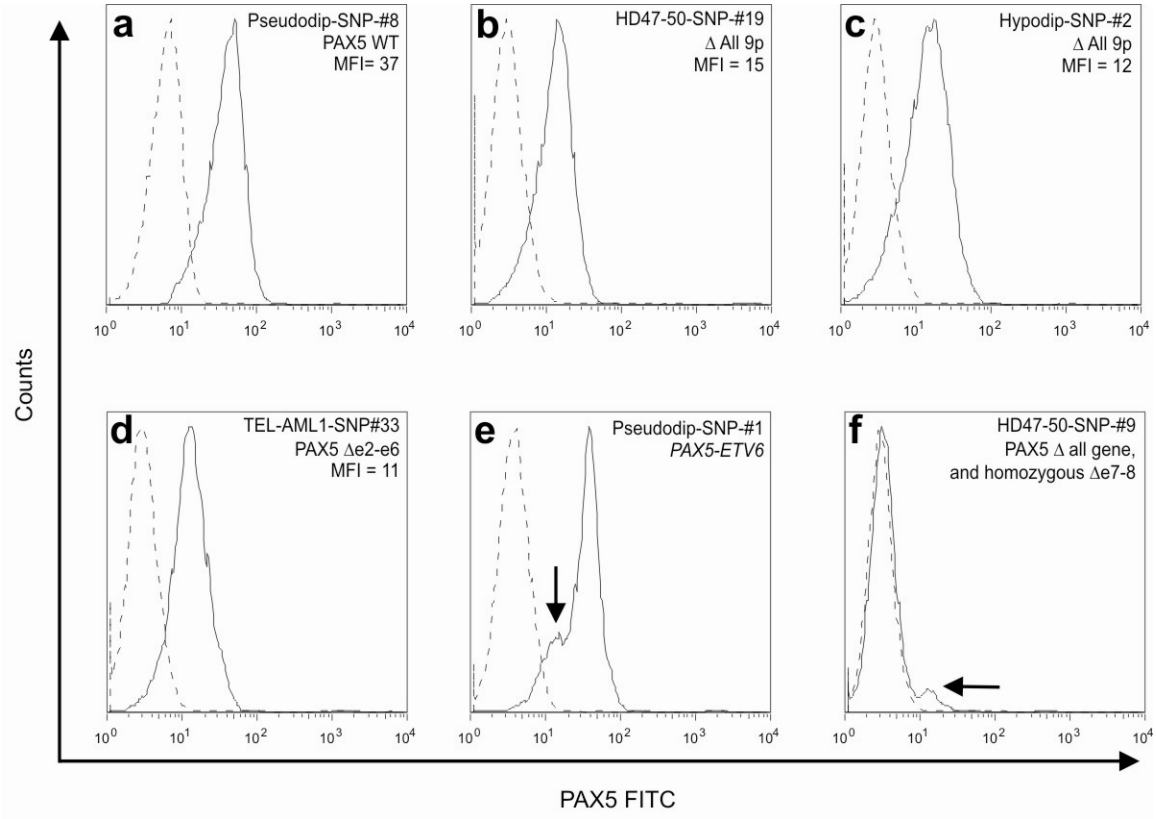
Supplementary Figure 19. Specific staining of normal blood B-lymphocytes by anti-PAX5

Peripheral blood mononuclear cells from a normal donor were stained for surface CD3 and IgM, followed by permeabilization and staining for PAX5. **a** shows distinct B cell (CD3-IgM+, region A) and T cell (CD3+IgM-, region B) populations. The B cell population is PAX5 positive **(b**) and the T cell population PAX5 negative (**c**)



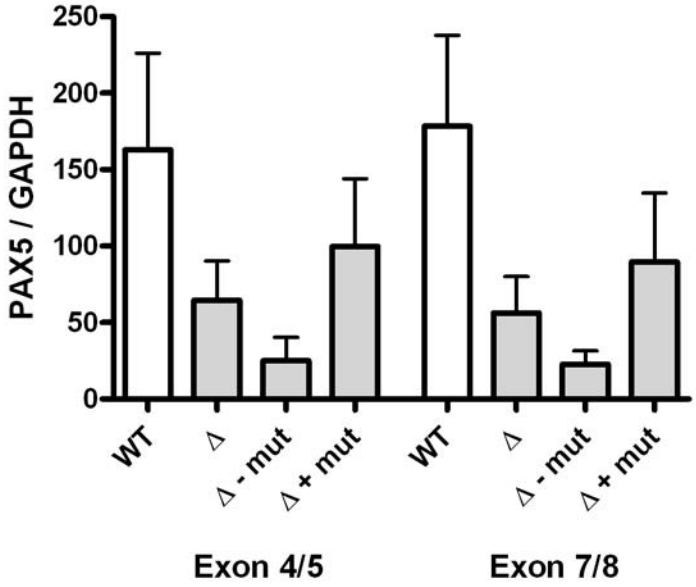
Supplementary Figure 20. Blocking studies demonstrate PAX5 specificity of the anti-PAX5 antibody used for immunophenotyping of leukaemic blasts

Peripheral blood from a normal donor was stained with PAX5, CD79A or isotype control antibodies preincubated with 15ng of pooled PAX5 peptide or non-specific peptide (SMAC). Staining of B-lymphocytes (gate **A**) is abolished by preincubation of PAX5 antibody with PAX5 peptide (**a**) but not control peptide (**b**). Specificity of inhibition by the PAX5 peptide pool is shown by lack of inhibition of staining with anti-CD79A antibody (**c-d**). SSC, side scatter.



Supplementary Figure 21. *PAX5*-deleted B-lineage ALLs show reduced PAX5 expression by flow cytometry

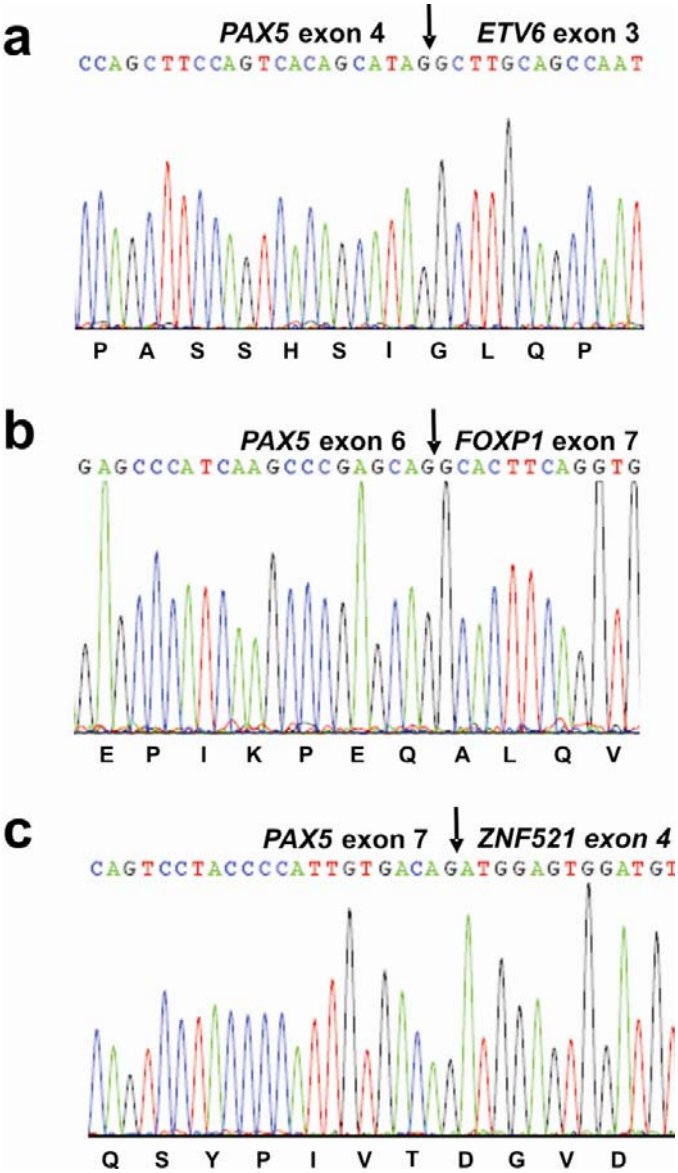
PAX5 flow cytometry was performed using a monoclonal antibody raised against the mid-portion (residues 151-306, encoded by exons 4-8) of PAX5. A histogram is shown for each sample. Samples were gated on CD19 positive lymphoblasts. Dotted lines indicate the isotype negative control. **a-c,** cases with hemizygous deletions; cases in B and C show reduced PAX5 expression compared to a case without *PAX5* deletion (a). **d,** Focal deletion also results in reduced expression of wild-type PAX5. **e,** A *PAX5-ETV6* case has two populations of cells with near normal and reduced PAX5 expression (↓), reflecting the wild-type and *PAX5-ETV6* fusion transcripts expressed in this case. **f,** A case with homozygous deletion shows near-negative PAX5 expression, with a population of cells showing weak PAX5 staining (←) due to partial deletion of the epitope recognized by the PAX5 antibody. MFI, mean fluorescence intensity, and is the difference of the MFI of PAX5 and isotype control. Of the 17 cases examined, nine cases had hemizygous *PAX5* deletion (but no mutation or translocation) were examined, which had a mean MFI of 13.56 (s.e.m. ± 1.63), compared with a mean MFI of 30.0 (± 7.0) for B-ALL cases without deletions (t test P=0.0048)



Supplementary Figure 22. Quantitation of *PAX5* gene expression by real-time PCR

Plots of PAX5 expression levels quantitated by real-time PCR using assays specific for PAX5 exons 4-5 and exons 7-8. Bars show mean + s.e.m. There was a broad range of PAX5 mRNA levels, particularly in *PAX5* wild-type cases. *PAX5* exon 4-5 mean (± SEM) expression relative to GAPDH was 163.0 (±63.1) for PAX5 wild-type cases, compared to 64.7 (±25.6) for all *PAX5*-deleted cases (t test with Welch's correction P=0.12), 25.1 (±15.2) for *PAX5*-deleted cases without concomitant point mutations (P=0.04), and 99.9 (±44.3) for *PAX5*-deleted cases with concomitant point mutations (P=NS). *PAX5* exon 7-8 mean (± SEM) expression relative to GAPDH was 178.4 (±59.4) for *PAX5* wild-type cases, compared to 56.3 (±23.8) for all *PAX5*-deleted cases (P=0.06), 22.9 (±8.8) for *PAX5*-deleted cases without concomitant point mutations (P=0.015), and 89.7 (±45.0) for *PAX5*-deleted cases with concomitant point mutations (P=NS).

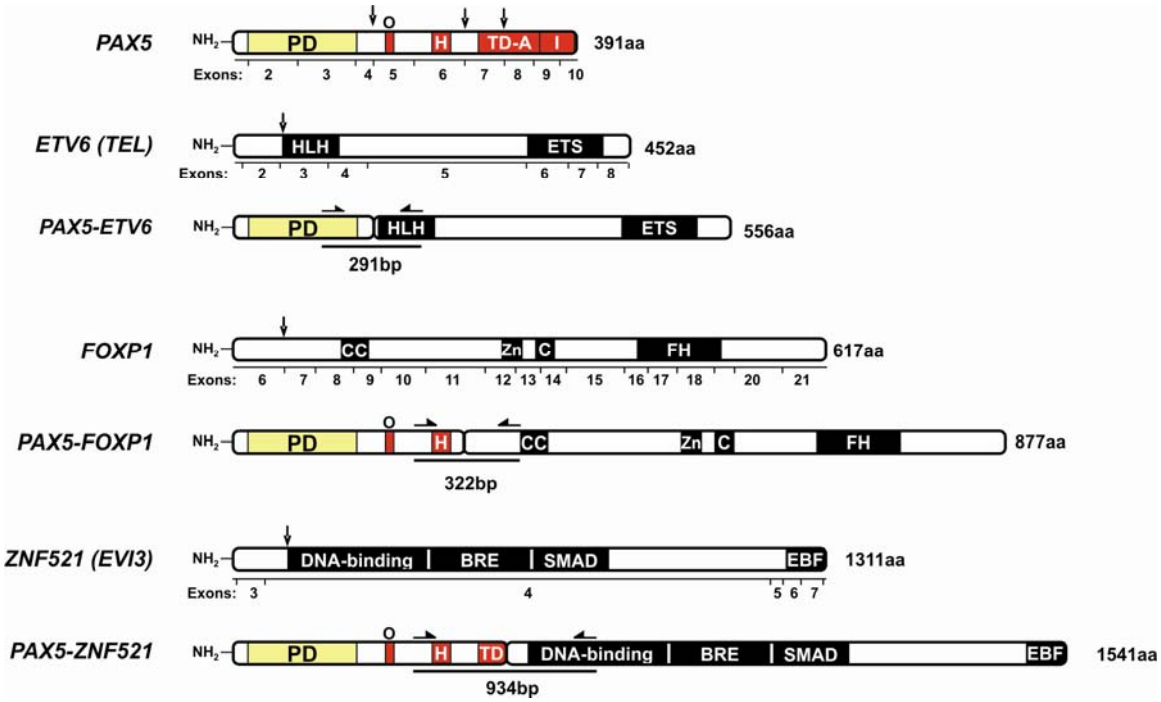
CRYPTIC TRANSLOCATIONS INVOLVING *PAX5* IN B-ALL



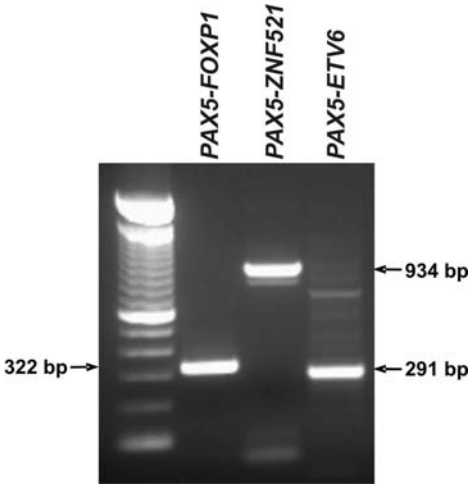
Supplementary Figure 23. Sequencing chromatograms of *PAX5* translocations

Direct sequencing of RT-PCR products demonstrating in-frame fusions of *PAX5* to a, *ETV6*, b, *FOXP1*, and c, *ZNF521* (*EVI3*). Breakpoints are shown by an arrow (↓). Protein coding sequence is shown below each pherogram.

a



b



Supplementary Figure 24. Fusion-specific RT-PCR confirms *PAX5* translocations in B-progenitor ALL

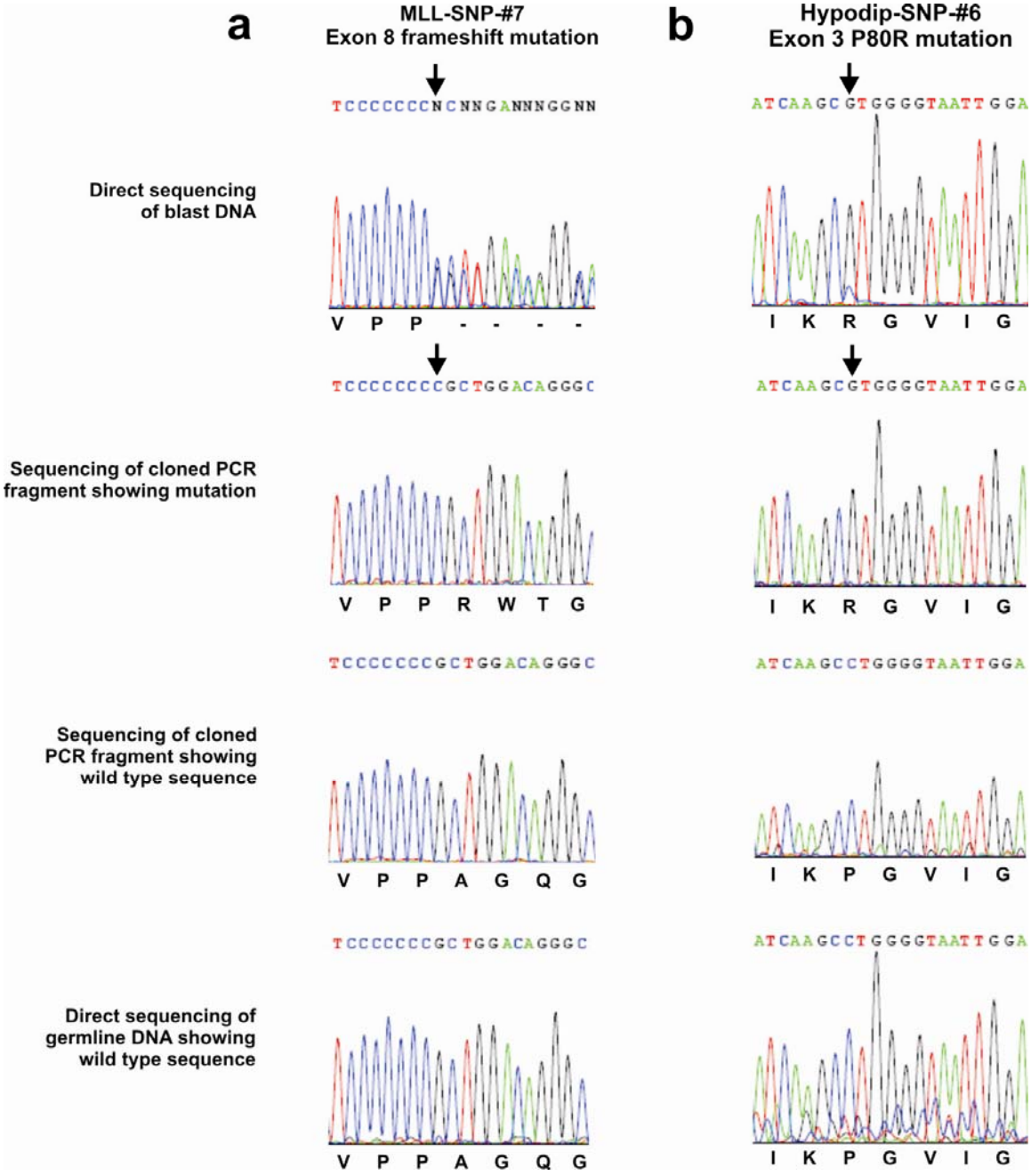
a, Schematic of *PAX5*, fusion partners and the fusion proteins. Breakpoints are indicated by (↓), and location of RT-PCR primers by (→/←); bp, base pairs; BRE, BMP2 response element domain; EBF, EBF-interaction domain; ETS, Ets domain; C and CC, coiled-coil domain; FH, forkhead domain; H, homeodomain-like; HLH, helix-loop-helix; O, octapeptide domain; PD, paired domain; SMAD, SMAD-interacting domain; Zn, zinc finger domain. b, fusion specific RT-PCR confirms the presence of the translocations.

POINT MUTATIONS OF *PAX5* IN B-ALL

Supplementary Table 17. Location of mutations and predicted effects on PAX5 amino acid sequence

Mutated residues are shown in red. Nucleotide numbering is according to position in mRNA (reference sequence NM_016734) if not otherwise indicated.
[1]Deletion of ATACACTGTAAGCACGACCCGTTTGCATCCATGCATA and insertion of CCCCCCCCAACG at position 1131 from the transcription start site (position +7 from the translation start). Sequence with Genbank accession AF386789 is used as the reference. This mutation alters amino acid sequence after residue E2, and results in a premature stop after a 10 amino acid peptide.
[2]Deletion of ATGACACCGTGC and insertion of GGG, resulting in deletion of N126 to P130, and insertion of inserts R126A127, with normal amino acid sequence from S128.
[3]Deletion of G, and insertion of TGTCACTAC, altering the amino acid sequence after N210, resulting in a premature stop in exon 7.
[4]Insertion of C, resulting in altered amino acid sequence after P321, and premature stop after E339.
[5]Insertion of CCCGGGGG; this alters the amino acid sequence after S339 with elongation of the PAX5 protein by 10 amino acids.
[6]This splice site mutation abolishes the splice donor site at the 3' end of exon 9. Sequencing of cloned *PAX5* cDNA from this case demonstrated that this mutation results in deletion of exon 9 and in frame fusion of exon 8 to exon 10.

Location	*Mutation*	*Nucleotide change*	*Patient ID*	*Protein sequence*
Exon 1B	WT			MEIHCKHDPFASMH
	Frame-shift E2	1131del37ins12[1]	Hyperdip47-50-SNP-#8	MEPPPTDMEE*
Exon 2	WT			GHGGVNQLGGVFVNGRPLPDVVRQRIVELAHQGVRPCDISRQLRVSHGCVSHKILG
	V26G	525T→G	Hyperdip47-50-SNP-#5 Hyperdip47-50-SNP-#19 Hypodip-SNP-#4	GHGGVNQLGGGFVNGRPLPDVVRQRIVELAHQGVRPCDISRQLRVSHGCVSHKILG
	P34Q	549C→A	Hypodip-SNP-#5	GHGGVNQLGGVFVNGRQLPDVVRQRIVELAHQGVRPCDISRQLRVSHGCVSHKILG
Exon 3	WT			RYYETGSIKPGVIGGSKPKVATPKVVEKIAEYKRQNPTMFAWEIRDRLLAERVCDNDTVPSVSSIN
	P80R	687C→G	Hypodip-SNP-#6 Hypodip-SNP-#10 Pseudodip-SNP-#2 other-SNP-#10	RYYETGSIKRGVIGGSKPKVATPKVVEKIAEYKRQNPTMFAWEIRDRLLAERVCDNDTVPSVSSIN
	IN/DEL NDTVP126RA	824del12ins3[2]	Hyperdip47-50-SNP-#8	RYYETGSIKPGVIGGSKPKVATPKVVEKIAEYKRQNPTMFAWEIRDRLLAERVCDRASVSSIN
Exon 6	WT			GIQESPVPNGHSLPGRDFLRKQMRGDLFTQQQLEVLDRVFERQHYSDIFTTTEPIPEQ
	Frame-shift N210	1079delGins9, 1237ins104[3]	other-SNP-#10	GIQESPVPNCHYATRFRAETSSGSRCGETCSGSSSWRCTACLRGSTTQTSSPPQSPSSPSRPQSIQPW PRWLVGWTT*
Exon 8	WT			GRDLASTTLPGYPPHVPPAGQGSYSAPTLTGMVP
	Frame-shift P321	1411insC[4]	E2A-PBX1-SNP-#3 MLL-SNP-#7	GRDLASTTLPGYPPHVPPRWTGQLLSTDADRDGAWE*
Exon 9	WT			GSEFSGSPYSPYSHPQYSSYNDSWRFPNPGLL
	G338W	1460G→T	Hypodip-SNP-#2	WSEFSGSPYSPYSHPQYSSYNDSWRFPNPGLL
	Frame-shift S339	1467ins8[5]	Pseudodip-SNP-#4	GSDPGGFPGVPTATLSIPRTTTPGGSPTRGCLAPPTIIALPPEEPPHLQPPLPMTVTDPWSQAGTKH*
Intron 9	IVS9+1	C→A[6]	Pseudodip-SNP-#17	GRDLASTTLPGYPPHVPPAGQGSYSAPTLTGMVPGSPYYYSAAARGAAPPAAATAYDRH*



Supplementary Figure 25. Sequencing chromatograms of *PAX5* point mutations

Chromatograms of genomic PCR products either directly or after cloning from cases with *PAX5* mutations, illustrating a, *PAX5* exon 8 frameshift mutation and b, *PAX5* exon 3 point mutation

Supplementary Table 18. Location of mutations, corresponding *PAX5* deletion status, blast and germline *PAX5* mutation status, and estimation of the ratio of wild type to mutated *PAX5* transcripts in cases with *PAX5* mutations

Proportions of wild-type and mutated *PAX5* transcripts were calculated by sequencing multiple colonies of cloned genomic PCR products.

Patient ID	*Mutation*	*Exon/ Intron*	*Germ-line Status*	*Deletion (hemizygous)*	*Blast %*	*Ratio WT/mutation in cloned PCR product*	*Predicted % mutants*	*Actual % mutants*
Hypodip-SNP-#4	V26G	Exon 2	WT	all 9p	98	10/6	98	37.5
Hyperdip47-50-SNP-#5	V26G	Exon 2	WT	all 9p	97	2/10	97	83.3
Hyperdip47-50-SNP-#19	V26G	Exon 2	WT	all 9p	97	2/27	97	93.1
Hypodip-SNP-#5	P34Q	Exon 2	WT	e1-e10	93	4/23	93	85.2
Hypodip-SNP-#6	P80R	Exon 3	WT	all 9p	94	2/16	94	88.9
Hypodip-SNP-#10	P80R	Exon 3	WT	all 9p	98	2/26	98	92.9
Pseudodip-SNP-#2	P80R	Exon 3	WT	all 9p	100	0/30	100	100
Other-SNP-#10	P80R Frame-shift N210	Exon 3 Exon 6	NA WT		93	NA 16/12	46.5 46.5	 42.9
Hyperdip47-50-SNP-#8	Frame-shift E2 IN/DEL NDTVP126RA	Exon 1B Exon 3	NA WT	e2-e10	95	7/34 13/16	95 95	82.9 55.2
E2A-PBX1-SNP-#3	Frame-shift P321	Exon 8	WT		94	13/7	47	35.0
MLL-SNP-#7	Frame-shift P321	Exon 8	WT		91	11/12	45.5	52.2
Hypodip-SNP-#2	G338W	Exon 9	WT	all 9p	96	6/11	96	64.7
Pseudodip-SNP-#4	Frame-shift S339	Exon 9	WT	e2-e6	90	16/8	45	33.3
Pseudodip-SNP-#17	IVS9+1	Intron9	WT		97	0/15	48.5	100

Supplementary Table 19. Patterns of mutations and deletions in cases with *PAX5* point mutations

*Two cases are heterozygous for mutations, one case is homozygous (no wild type genomic or RT-PCR transcripts identified on cloning and sequencing). **This case (Hyperdip47-50-SNP-#8) has a complex pattern of *PAX5* abnormalities with deletion of one allele from exon 1B to distal to exon 10 (thus resulting in no functional transcript from this allele), and two mutations: a truncating frameshift mutation of exon 1B (truncating in exon 2) and an in-frame insertion/deletion of exon 3 (Supplementary Table 17). Cloning and sequencing of PAX5 exon 1B transcripts by RT-PCR demonstrated that all (n=35) clones carry the exon 1B mutation; 13 (37.1%) have splicing of exon 2 to exon 4, 7 (20.0%) have exon 1B mutated but no exon 3 mutation, and 13 (37.1%) have both the exon1B mutation and exon 3 mutation in *cis*. Single clones had splicing of mutated exon 1B to either exon 3 or 4. This suggests that the exon 1B mutation was acquired prior to the exon 3 insertion/deletion, and the latter mutation is present in a subclone. This is compatible with the results of cloning and sequencing of exon 3 genomic PCR products in this case, in which the mutation was detected in 16 of 29 (55%) of clones (Supplementary Table 18). However, the exon 3 mutation is of no functional relevance in this transcript, as the exon 1B mutation truncates PAX5 in exon 2.

Pattern of deletion and/or mutation	Number of cases
Mutation, no deletion*	3
Two mutations in trans	1
Focal deletion and hemizygous mutation in trans	1
Deletion of all of PAX5, hemizygous point mutation	8
Focal deletion and two mutations in trans to deletion**	1

STRUCTURAL MODELLING OF *PAX5* POINT MUTATIONS

1. Domain structure

The DNA-binding paired domain of PAX5 contains two structurally independent globular subdomains (N- and C-terminal) [52]. The structure and DNA docking mechanism of the N-terminal domain (consisting of two antiparallel beta-sheets and three alpha helices) are highly conserved in the PAX family. The C-terminal domain of PAX5 and PAX6 includes three alpha-helices and contributes to the DNA-binding of PAX5 and PAX6 but is less conserved among the PAX family. Described developmental missense mutations in mice and humans map to the N-terminal domain [52, 67] but affect other residues than those mutated in ALL cases in this study.

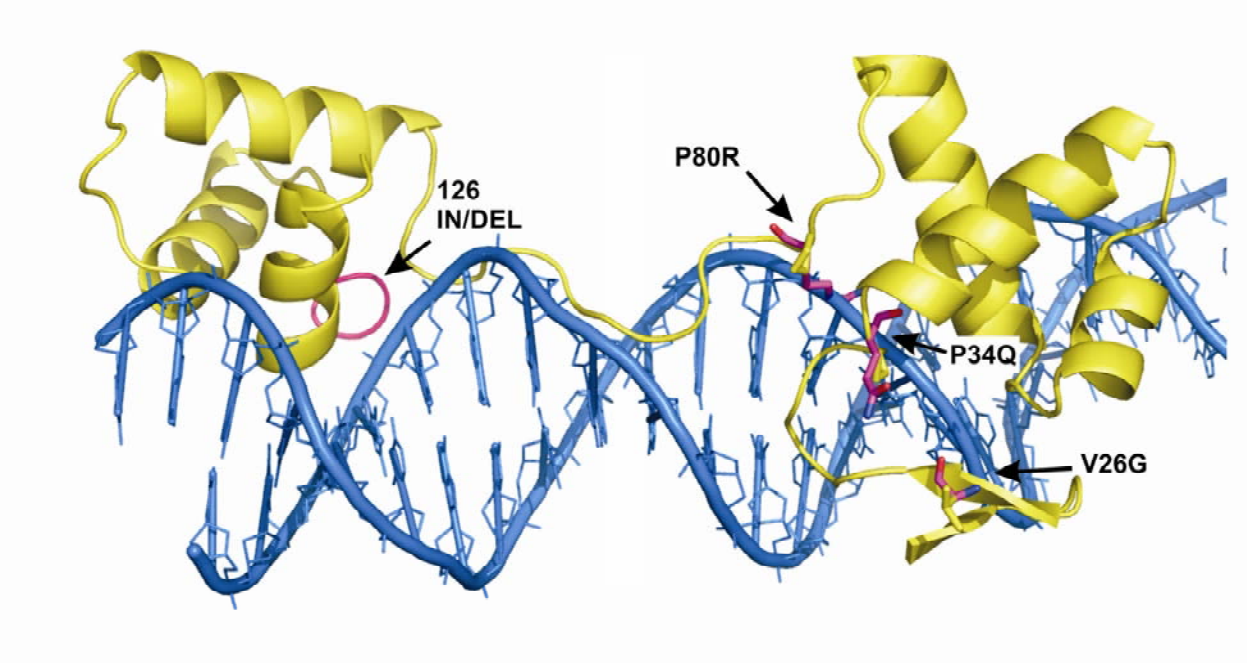
2. Structural consequences of *PAX5* point mutations

Replacement of prolines has a significant impact on the protein structure as proline has stronger stereochemical constraints than other amino acids (the rotational freedom around the central carbon is restricted by inclusion of the amid-nitrogen in the side-chain ring). It usually forms turns in the protein chain. The proline residues mutated in B-progenitor ALL cases (P34, P80) are invariant among the PAX family members. The proline in position 80 of the protein contacts the phosphate backbone of the DNA and is replaced by an arginine in the P80R mutant (4 cases). The arginine will clash directly into the minor groove of the DNA. The alterations in the protein backbone caused by replacement of proline by glutamine in P34Q (1 case) may affect positioning of the preceding DNA-contacting residues, in particular L33 (Supplementary Figure 26).

In contrast to the rigid proline, introduction of glycine significantly increases the flexibility of the protein backbone. The replacement of valine by glycine in V26G (3 patients) could result in structural alterations. It is part of the second beta-sheet containing the DNA-contacting residue F27. Other PAX family members contain aliphatic residues in that position (alanine in human PAX8, leucine in mouse PAX4) but not glycine.

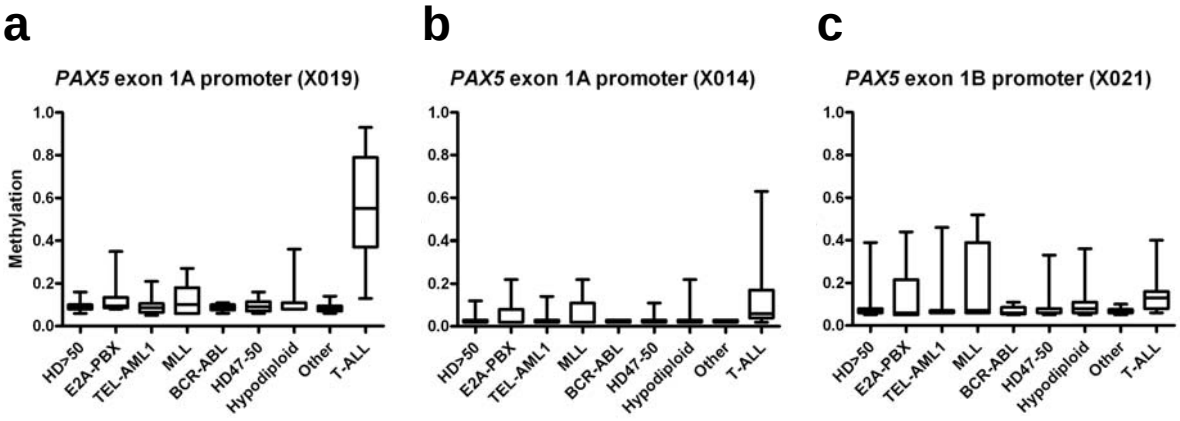
The deletion of NDTVP and in-frame insertion of RA (1 case) occurs in the loop between alpha-helix 5 and 6 of the C-terminal part of the paired domain. The backbone oxygen of the valine that resides in the deleted stretch of amino acids directly contacts the DNA. The backbone nitrogen of the subsequent serine also contacts the DNA and may be displaced by deletion of the preceding five residues. Proline and serine are conserved in almost all paired domains. Studies of other helix-turn-helix structures indicated that large insertions can be tolerated between helix 5 and 6 [68] but the observed insertion/deletion mutation shortens the ten amino acid turn by 3 residues and also deletes a DNA-contacting residue.

In addition, the paired-domain mutations affect amino acids that are in close proximity to residues critical for recruitment of the ETS1, a co-activator of the PAX5 target gene *CD79A* [69].



Supplementary Figure 26. Modelling of *PAX5* paired domain mutations

The PAX5 paired domain is yellow, DNA in blue, and mutated residues pink. The mutations are predicted to impair DNA-binding: V26G increases protein backbone flexibility and affects the adjacent DNA contact residue F27; P34Q affects positioning of the preceding L33 DNA-contacting residue; P80R disrupts contact with the phosphate backbone of DNA by placing arginine directly into the minor groove of DNA; and NDTVP126RA shortens a 10 amino acid turn and deletes a DNA-contacting residue.

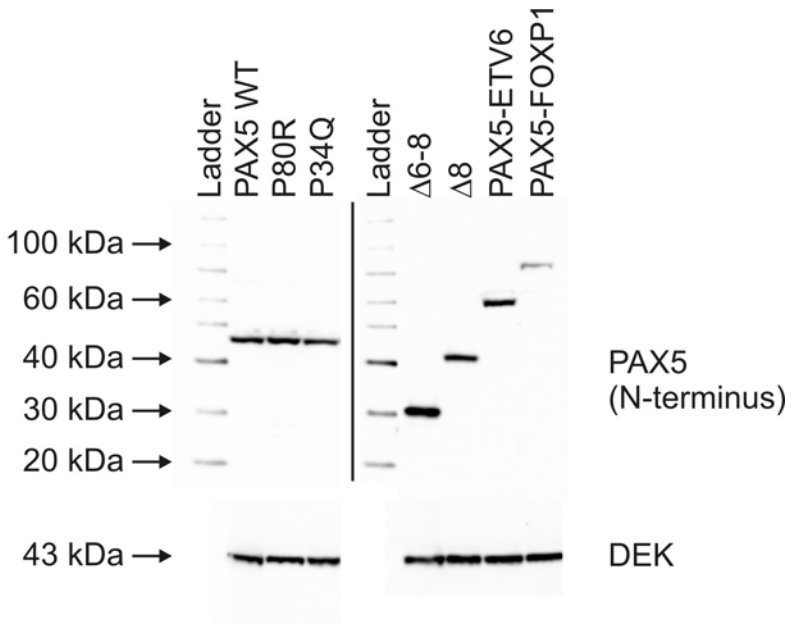


Supplementary Figure 27. Hypermethylation of *PAX5* in T-ALL

a-c, Box-and-whisker plots showing methylation levels for each of the *PAX5* amplicons in ALL. The mean methylation level of all CpG-containing fragments within an amplicon for each case is used to generate the plots. Methylation levels were significantly greater in T-ALL than each B-ALL subtype for the *PAX5* X019 amplicon, located in the distal upstream promoter of *PAX5* exon 1a.

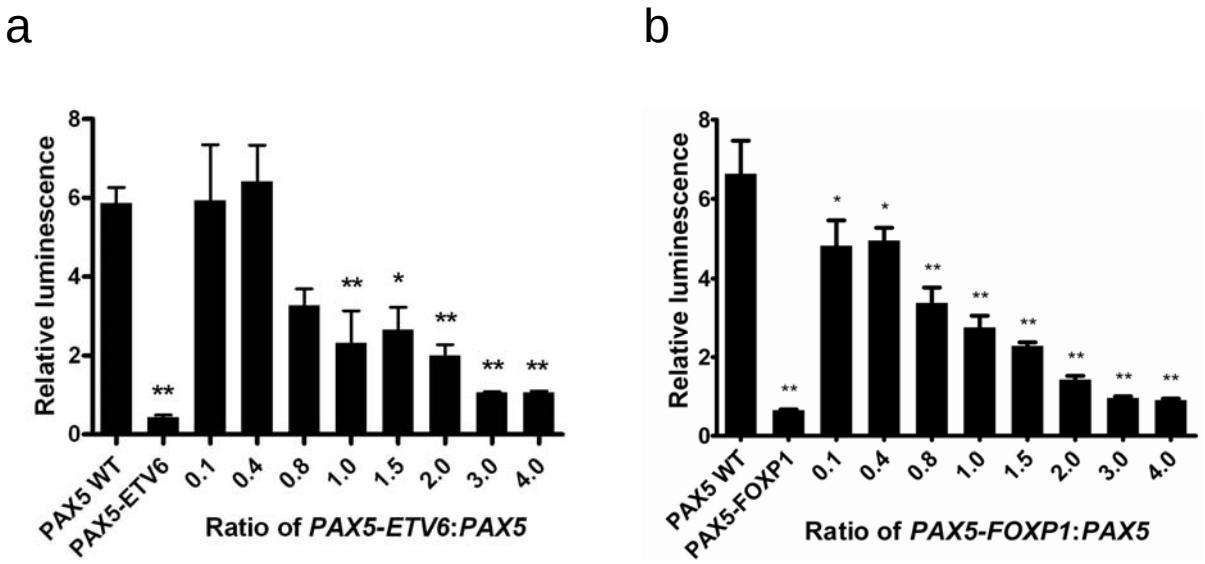
PAX5 MUTATIONS COMPROMISE DNA-BINDING AND TRANSCRIPTIONAL ACTIVATION

Western blot analysis of nuclear extracts confirmed comparable expression in the assays for wild-type and each of the mutant PAX5 variants examined shown in Figure 4a-b.



Supplementary Figure 28. PAX5 western blots of nuclear extracts of transfected 293T cells used for *luc*-CD19 reporter assays

Western blots using an N-terminus PAX5 specific antibody (Chemicon) were performed on nuclear extracts of transiently transfected 293T cells. Neither this antibody, nor any of the alternative available PAX5 specific antibodies detect the PAX5 D2-6 or D2-8 internal truncation mutants.



Supplementary Figure 29. *PAX5-ETV6* and *PAX5-FOXP1* competitively inhibit the transcriptional activity of wild-type PAX5

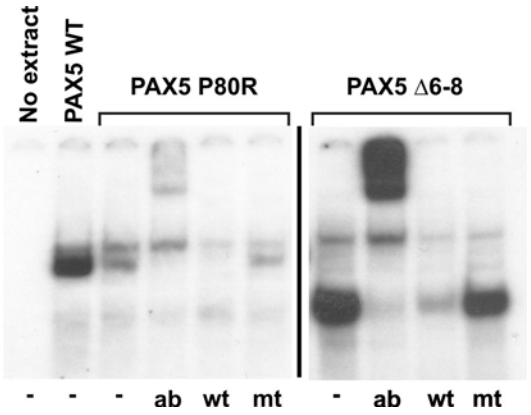
Relative luminescence of 293T cells transfected with *luc*-CD19, fixed amount of PAX5 WT and increasing amounts of *PAX5-ETV6* (**a**) or *PAX5-FOXP1* (**b**) plasmids. Empty MSCV-IRES-mRFP (MIR) expression vector was used to maintain a constant mass of DNA in each transfection experiment. Luminescence is normalized to *Renilla* luciferase activity and is shown relative to 293T cells infected with empty MIR vector. Bars show means ± SEM of triplicate experiments. *ANOVA with Dunnett's test $P<0.05$; **$P<0.01$. Importantly, expression of the various PAX5 mutants did not inhibit the transcription of the pRL-TK *Renilla* luciferase plasmid used as an internal control for transfection efficiency (Supplementary Table 20).

Supplementary Table 20. Raw firefly and *Renilla* luciferase data for *luc*-CD19 reporter assays

The table lists the raw firefly (representing activity of each tested vector) and Renilla luciferase activities. Each construct was tested in triplicate. Different experiments are separated by blank lines, and each incorporated empty vector (MIR) and PAX5 wild-type (PAX5 WT) controls. indel, insertion-deletion (mutation).

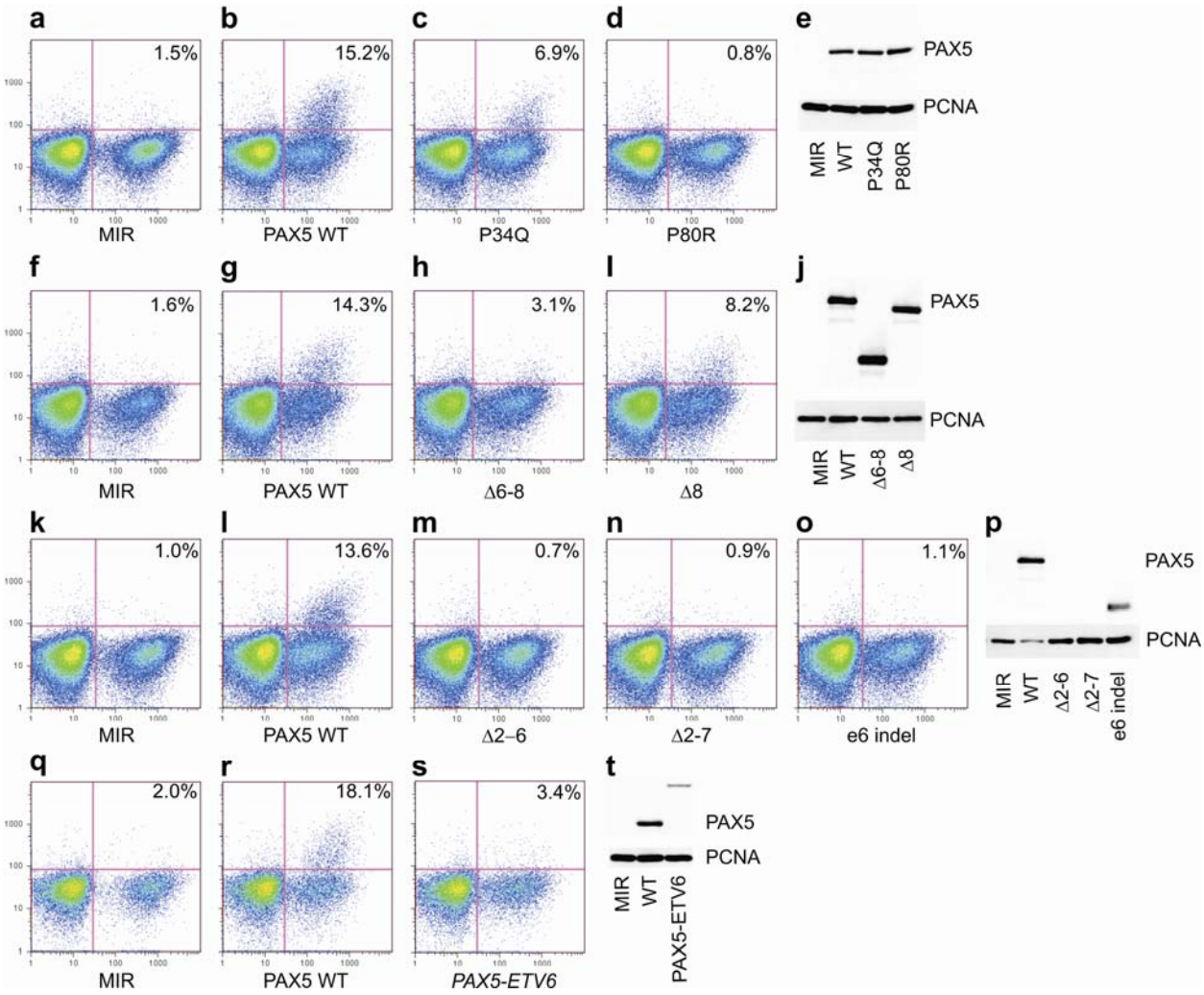
Construct	Firefly	*Renilla*	Firefly/*Renilla*
MIR 1	84707	136964	61.8
MIR 2	116256	167273	69.5
MIR 3	127609	181232	70.4
PAX5 WT 1	790084	240045	329.1
PAX5 WT 2	627370	194924	321.9
PAX5 WT 3	703101	216221	325.2
e6 indel 1	92701	146778	63.2
e6 indel 2	67709	121331	55.8
e6 indel 3	126277	174269	72.5
MIR 1	83865	2808373	3.0
MIR 2	57990	2173186	2.7
MIR 3	63302	2135190	3.0
PAX5 WT 1	391490	1780924	22.0
PAX5 WT 2	220611	1071109	20.6
PAX5 WT 3	451802	2115741	21.4
P80R 1	110983	1084975	10.2
P80R 2	199076	2003527	9.9
P80R 3	175299	1459224	12.0
Δ2-6 1	61392	1890958	3.2
Δ2-6 2	61700	1647098	3.7
Δ2-6 3	72603	2009871	3.6
Δ2-7 1	106263	2968500	3.6
Δ2-7 2	37203	1018654	3.7
Δ2-7 3	27060	778764	3.5
Δ6-8 1	154214	3209041	4.8
Δ6-8 2	128451	2748889	4.7
Δ6-8 3	119306	2396928	5.0
Δ8 1	349881	2598802	13.5
Δ8 2	369893	2225925	16.6
Δ8 3	424839	2736297	15.5
MIR 1	98333	301351	32.6
MIR 2	190841	533933	35.7
MIR 3	216003	568882	38.0
PAX5 WT 1	818075	487476	167.8
PAX5 WT 2	1050106	602576	174.3
PAX5 WT 3	920393	543643	169.3

Construct	Firefly	*Renilla*	Firefly/*Renilla*
P34Q 1	367710	268528	136.9
P34Q 2	326414	276698	118.0
P34Q 3	660728	426421	154.9
MIR 1	47390	2583848	1.8
MIR 2	37698	2119538	1.8
MIR 3	34191	1948844	1.8
PAX5 WT 1	334413	3167117	10.6
PAX5 WT 2	483449	4135683	11.7
PAX5 WT 3	219414	2370826	9.3
PAX5-ETV6 1	89014	9241342	1.0
PAX5-ETV6 2	56733	7684280	0.7
PAX5-ETV6 3	39733	6473496	0.6
MIR 1	67082	3529265	1.9
MIR 2	140590	5274656	2.7
MIR 3	110557	4461232	2.5
PAX5 WT 1	381000	3217954	11.8
PAX5 WT 2	786202	4778264	16.5
PAX5 WT 3	811377	4398572	18.4
PAX5-FOXP1 1	95878	6405809	1.5
PAX5-FOXP1 2	94912	6321917	1.5
PAX5-FOXP1 3	103563	6442528	1.6



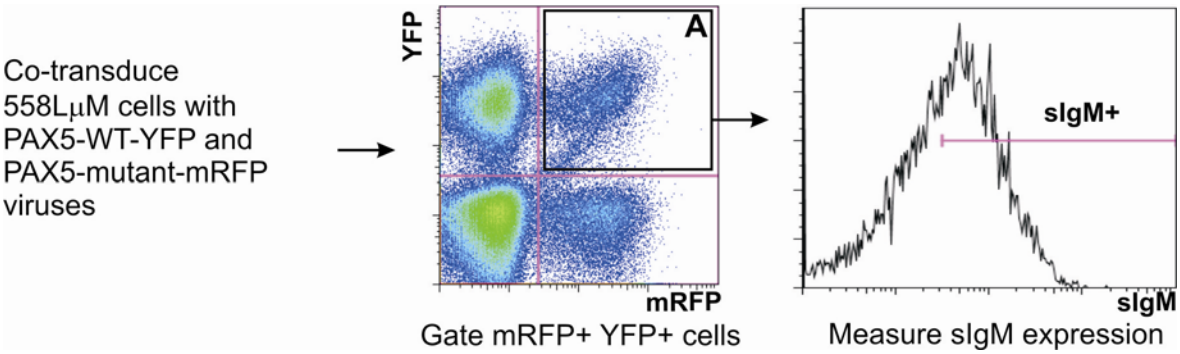
Supplementary Figure 30. DNA-binding of PAX5 mutant alleles

Gel-shift assay using showing reduced (P80R) and normal (Δ6-8) binding of PAX5 variants to a CD19 promoter binding site. ab, supershift with PAX5 antibody; wt, wild-type and mt, mutant competitor oligonucleotides; -, no antibody or competitor oligonucleotides.



Supplementary Figure 31. *PAX5* mutations impair Cd79a transactivation and sIgM expression in the 558LμM cell line

558LμM cells were transduced with MSCV-PAX5-IRES-mRFP retroviruses expressing wild type or mutant PAX5. sIgM expression of the mRFP-positive population was measured. Four experiments examining different PAX5 variants were performed (**a-e**, **f-j**, **k-p** and **q-t**). Empty vector (MIR) and wild-type PAX5 (PAX5 WT) were included as controls in each experiment. All transductions were performed in triplicate, and representative flow plots are shown. Percentages refer to the proportion of RFP+ cells expressing sIgM. Two million RFP+ cells were sorted from a transduction experiment for each PAX5 variant, and western blotting for PAX5 and the control antigen PCNA performed on nuclear extracts of the RFP+ population. Sort purities were >90%. Western blot analysis confirmed comparable protein expression of the PAX5 variants analysed, with the exception of PAX5-ETV6, which is larger than the other variants (approx. 65kDa *c.f.* 45 kDa for WT PAX5), and is consistently expressed at lower protein levels than the other variants. The PAX5 antibody used (BD Transduction Labs) was raised against an internal PAX5 epitope and does not recognize the Δ2-6 or Δ2-7 *PAX5* internal deletion mutants shown in **p**.



Supplementary Figure 32. Design of 558LμM PAX5 WT and mutant co-transduction experiments

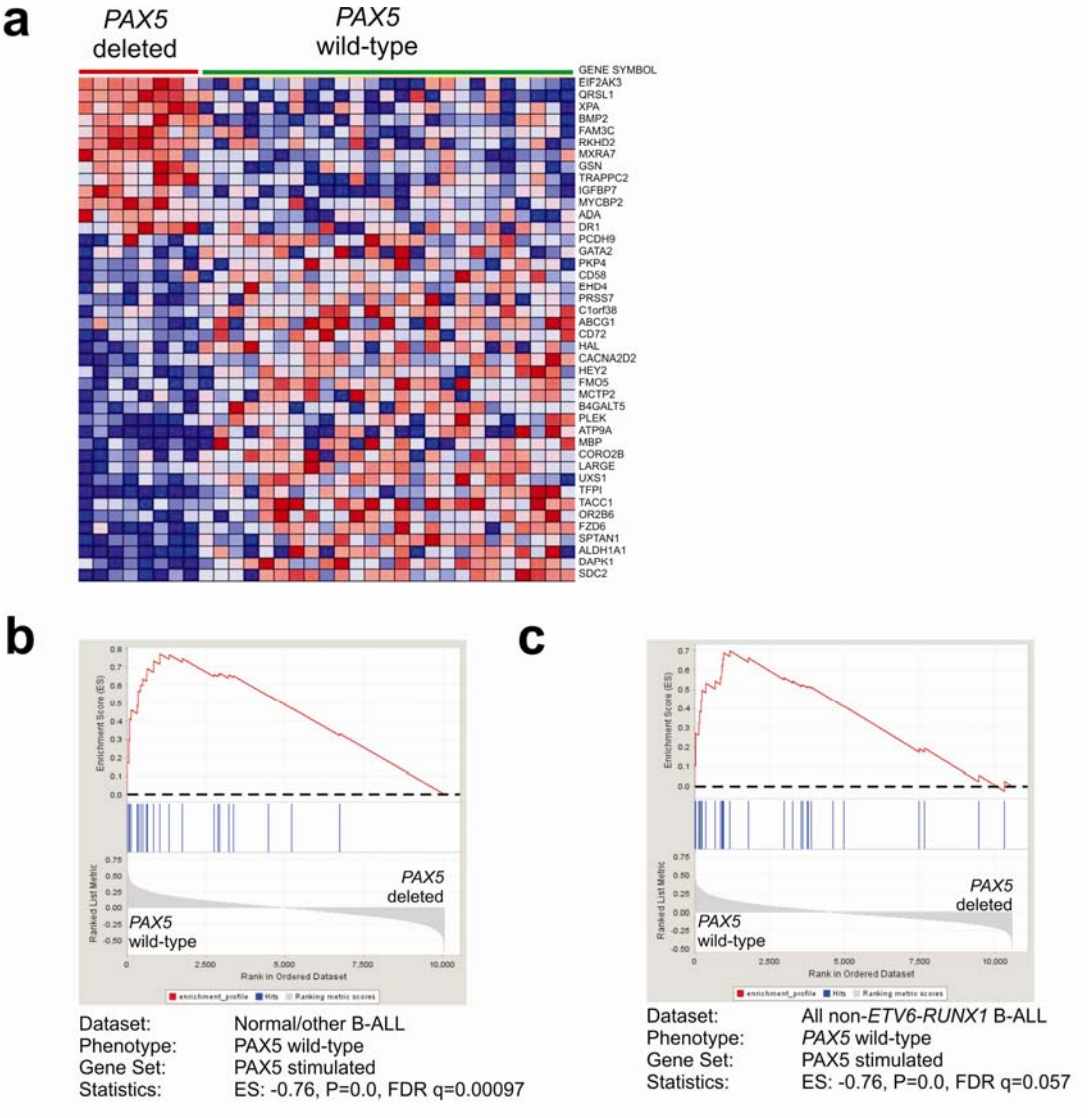
558LμM cells were co-transduced with bicistronic MSCV retroviruses co-expressing wild type PAX5 with YFP, and mutant PAX5 alleles with mRFP. Surface IgM expression was then quantitated on the dual positive population

CROSS-SUBTYPE GENE SET ENRICHMENT ANALYSIS OF PAX5 TARGETS IN B-PROGENITOR ALL

Supplementary Table 21. Affymetrix HG-U133A probe sets showing differential expression between *PAX5*-deleted and *PAX5* wild type *ETV6-RUNX1* B-progenitor ALL at a FDR of <0.3

HG U133A probe set	Gene symbol	Gene name	Expression level *PAX5* mutated	Expression level *PAX5* wild type	Log_2 Fold change	T	P	FDR
203139_at	DAPK1	death-associated protein kinase 1	560.8	1546.8	-1.46	-6.19	5.03E-07	0.007
218696_at	EIF2AK3	eukaryotic translation initiation factor 2-alpha kinase 3	3361.0	1619.6	1.05	5.31	6.94E-06	0.023
207638_at	PRSS7	protease, serine, 7 (enterokinase)	50.1	284.6	-2.51	-5.18	1.01E-05	0.023
212154_at	SDC2	syndecan 2	436.4	2002.1	-2.20	-5.30	7.18E-06	0.023
213258_at	TFPI	tissue factor pathway inhibitor	69.2	390.3	-2.50	-5.19	9.93E-06	0.023
210664_s_at	TFPI	tissue factor pathway inhibitor	117.0	377.3	-1.69	-5.45	4.56E-06	0.023
212224_at	ALDH1A1	aldehyde dehydrogenase 1 family, member A1	21.3	206.9	-3.28	-4.63	5.18E-05	0.079
203987_at	FZD6	frizzled homolog 6	173.2	350.3	-1.02	-4.68	4.49E-05	0.079
218949_s_at	QRSL1	glutaminyl-tRNA synthase (glutamine-hydrolyzing)-like 1	2625.8	1106.3	1.25	4.63	5.18E-05	0.079
205290_s_at	BMP2	bone morphogenetic protein 2	1560.3	596.7	1.39	4.53	6.89E-05	0.086
212157_at	SDC2	syndecan 2	441.9	1062.2	-1.27	-4.55	6.52E-05	0.086
200911_s_at	TACC1	transforming, acidic coiled-coil containing protein 1	1682.2	2482.4	-0.56	-4.44	9.05E-05	0.104
209789_at	CORO2B	Coronin, actin binding protein, 2B	138.8	327.7	-1.24	-4.31	0.0001	0.134
215543_s_at	LARGE	like-glycosyltransferase	613.1	1100.5	-0.84	-4.30	0.0001	0.134
208611_s_at	SPTAN1	spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	2058.1	3451.2	-0.75	-4.20	0.0002	0.166
204811_s_at	CACNA2D2	calcium channel, voltage-dependent, alpha 2/delta subunit 2	119.2	390.2	-1.71	-4.17	0.0002	0.172
200696_s_at	GSN	gelsolin (amyloidosis, Finnish type)	4308.7	1800.2	1.26	4.12	0.0002	0.184
212509_s_at	MXRA7	matrix-remodelling associated 7	1894.5	957.6	0.98	4.07	0.0003	0.186
216522_at	OR2B6	olfactory receptor, family 2, subfamily B, member 6	247.4	580.4	-1.23	-4.06	0.0003	0.186
215235_at	SPTAN1	spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	1371.3	2387.5	-0.80	-4.08	0.0003	0.186
201889_at	FAM3C	family with sequence similarity 3, member C	1134.5	685.5	0.73	4.04	0.0003	0.189
201162_at	IGFBP7	insulin-like growth factor binding protein 7	432.8	68.7	2.65	4.01	0.0003	0.197
209187_at	DR1	down-regulator of transcription 1, TBP-binding	1081.3	694.5	0.64	3.94	0.0004	0.223
201960_s_at	MYCBP2	MYC binding protein 2	2585.1	1873.6	0.46	3.95	0.0004	0.223

HG U133A probe set	Gene symbol	Gene name	Expression level *PAX5* mutated	Expression level *PAX5* wild type	Log_2 Fold change	T	P	FDR
219675_s_at	UXS1	UDP-glucuronate decarboxylase 1	1330.9	1881.5	-0.50	-3.92	0.0004	0.226
211113_s_at	ABCG1	ATP-binding cassette, sub-family G (WHITE), member 1	262.4	471.5	-0.85	-3.90	0.0004	0.227
218247_s_at	RKHD2	ring finger and KH domain containing 2	803.2	485.9	0.73	3.88	0.0005	0.234
212062_at	ATP9A	ATPase, Class II, type 9A	23.3	176.0	-2.91	-3.82	0.0005	0.251
221484_at	B4GALT5	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 5	529.7	985.8	-0.90	-3.83	0.0005	0.251
209710_at	GATA2	GATA binding protein 2	164.2	389.2	-1.25	-3.83	0.0005	0.251
219738_s_at	PCDH9	protocadherin 9	221.7	376.4	-0.76	-3.77	0.0006	0.272
219351_at	TRAPPC2	trafficking protein particle complex 2	1759.9	1149.2	0.61	3.77	0.0006	0.272
206643_at	HAL	histidine ammonia-lyase	61.2	180.3	-1.56	-3.74	0.0007	0.278
219743_at	HEY2	hairy/enhancer-of-split related with YRPW motif 2	124.1	331.1	-1.42	-3.73	0.0007	0.278
210136_at	MBP	myelin basic protein	317.6	709.3	-1.16	-3.73	0.0007	0.278
207571_x_at	C1orf38	chromosome 1 open reading frame 38	1277.3	1833.4	-0.52	-3.70	0.0008	0.278
209536_s_at	EHD4	EH-domain containing 4	165.1	308.5	-0.90	-3.72	0.0007	0.278
220603_s_at	MCTP2	multiple C2 domains, transmembrane 2	295.2	504.2	-0.77	-3.68	0.0008	0.278
201929_s_at	PKP4	plakophilin 4	146.4	342.6	-1.23	-3.69	0.0008	0.278
203471_s_at	PLEK	pleckstrin	415.3	752.4	-0.86	-3.69	0.0008	0.278
212158_at	SDC2	syndecan 2	300.0	1211.7	-2.01	-3.67	0.0008	0.279
215925_s_at	CD72	CD72 molecule	453.5	844.4	-0.90	-3.64	0.0009	0.295
204639_at	ADA	adenosine deaminase	2353.9	1350.4	0.80	3.61	0.001	0.298
211744_s_at	CD58	CD58 molecule	885.4	1517.4	-0.78	-3.62	0.0009	0.298
205672_at	XPA	xeroderma pigmentosum, complementation group A	449.5	174.8	1.36	3.61	0.001	0.298
205776_at	FMO5	flavin containing monooxygenase 5	141.2	230.4	-0.71	-3.60	0.001	0.298



Supplementary Figure 33. Cross-subtype gene set enrichment analysis (GSEA) of PAX5-regulated genes in B-progenitor ALL

a, heatmap of 42 differentially expressed genes (at FDR <0.3) in *ETV6-RUNX1 PAX5*-deleted *v. ETV6-RUNX1 PAX5* wild-type ALL. **b**, GSEA showing significant enrichment of the PAX5-stimulated genes shown in B-ALL samples lacking recurring cytogenetic abnormalities. The ranked gene list of *PAX5*-wild-type *v. PAX5*-deleted cases is shown in the lower, grey plot. Vertical blue lines indicated where probe sets in the PAX5-stimulated gene set fall in the ranked gene list. The top, red line indicates the running enrichment score (*ES*) that becomes more positive as probe sets are encountered at the top of the list. **c**, significant enrichment for PAX5-stimulated genes in all non-*ETV6-RUNX1* B-progenitor ALL. Enrichment for the PAX5-repressed gene set in PAX5-mutated cases was also observed in both non-*ETV6-RUNX1* cohorts, but was not significant (i.e. FDR>0.25) after correction for multiple hypothesis testing.

MONO-ALLELIC DELETIONS OF OTHER B-CELL DEVELOPMENT GENES IN PAEDIATRIC ALL

Supplementary Table 22. Frequency of B cell development gene mutations in B-precursor ALL

For each gene listed, cases with deletions limited to the gene of interest were identified. Interestingly, 10 of 17 *IKZF1* deletions and the single *LEF1* deletion occurred in cases that also had mono-allelic deletions of *PAX5* (two cases also had *PAX5* point mutations) (Supplementary Table 24). Deletions were also identified in a component of the pre-B cell receptor complex, *VPREB1* (55 cases). Deletion of *VPREB1*, however, is a consequence of rearrangement of the immunoglobulin lambda light chain (*IGL*) locus at chromosome 22q11. [1]Includes pseudodiploid cases and cases with normal cytogenetics. The exact likelihood-ratio chi-square test was calculated to examine differences in lesion frequency between B-progenitor ALL subtypes: [2] P=0.0693; [3] P<0.0001; [4] P=0.0001; [5] P=0.0167; [6] P=NS. To explore whether significant p-values are driven by the hypodiploid or hyperdiploid groups, we performed analyses excluding those two subtypes using the exact chi-square test. The remaining 6 subgroups have differing frequencies of the PAX5 point mutation (p = 0.0895) and Ikaros deletions (p = 0.0028). However, there was not strong evidence suggesting that the remaining 6 subgroups have differing frequencies of mutations in *EBF* (p = 0.2748), the PAX5 deletion/internal amplification (p = 0.4822), *Aiolos* (p = 0.6713), or *LEF1* (p = 0.3981).

		PAX5					
Genetic subtype (N)	**EBF [2]**	**Deletion / Internal amplification [3]**	**Point mutation [4]**	**IKZF1 (Ikaros) [4]**	**IKZF3 (Aiolos) [5]**	**LEF1 [6]**	**Total cases with mutation (N,%) [3]**
Hypodiploid (10)	1	10	5	5	2	1	10 (100%)
ETV6-RUNX1 (47)	5	13	0	0	0	0	16 (34%)
Other[1] (36)	1	12	4	5	1	1	19 (53%)
HD47-50 (23)	0	7	3	1	0	0	8 (35%)
TCF3-PBX1 (17)	0	7	1	0	0	1	8 (47%)
BCR-ABL (9)	1	4	0	3	0	0	6 (66%)
MLL (11)	0	1	1	1	0	0	3 (27%)
HD>50 (39)	0	3	0	2	0	0	5 (13%)

.

Supplementary Table 23. Full listing of genomic lesions in the B cell differentiation pathway for the entire ALL cohort

*With focal homozygous deletion at *IKZF1* (Ikaros)

Case	*EBF1* deletion	*PAX5* deletion	*PAX5* mutation	Aiolos deletion	*BLNK* deletion	*IGLL1* deletion	Ikaros deletion	*LEF1* deletion	*TCF3* deletion	*VPREB1* deletion
Hyperdip>50-SNP-#1										
Hyperdip>50-SNP-#2										Yes
Hyperdip>50-SNP-#3										
Hyperdip>50-SNP-#4										
Hyperdip>50-SNP-#5										
Hyperdip>50-SNP-#6						Yes				
Hyperdip>50-SNP-#7										Yes
Hyperdip>50-SNP-#8										
Hyperdip>50-SNP-#9										
Hyperdip>50-SNP-#10										
Hyperdip>50-SNP-#11										
Hyperdip>50-SNP-#12										
Hyperdip>50-SNP-#13										Yes
Hyperdip>50-SNP-#14										
Hyperdip>50-SNP-#15										
Hyperdip>50-SNP-#16										
Hyperdip>50-SNP-#17		Broad								
Hyperdip>50-SNP-#18										
Hyperdip>50-SNP-#19										
Hyperdip>50-SNP-#20										
Hyperdip>50-SNP-#21									Yes	
Hyperdip>50-SNP-#22										
Hyperdip>50-SNP-#23										
Hyperdip>50-SNP-#24		Broad								
Hyperdip>50-SNP-#25										
Hyperdip>50-SNP-#26										
Hyperdip>50-SNP-#27						Yes				
Hyperdip>50-SNP-#28		Focal								Yes
Hyperdip>50-SNP-#29										Yes
Hyperdip>50-SNP-#30										Yes
Hyperdip>50-SNP-#31							Focal			
Hyperdip>50-SNP-#32										Yes
Hyperdip>50-SNP-#33										
Hyperdip>50-SNP-#34							All 7p			
Hyperdip>50-SNP-#35										
Hyperdip>50-SNP-#36										
Hyperdip>50-SNP-#37										
Hyperdip>50-SNP-#38										
Hyperdip>50-SNP-#39										
E2A-PBX1-SNP-#1										
E2A-PBX1-SNP-#2										

Case	*EBF1* deletion	*PAX5* deletion	*PAX5* mutation	Aiolos deletion	*BLNK* deletion	*IGLL1* deletion	Ikaros deletion	*LEF1* deletion	*TCF3* deletion	*VPREB1* deletion
E2A-PBX1-SNP-#3			exon 8 FS P321							
E2A-PBX1-SNP-#4										
E2A-PBX1-SNP-#5		All 9p								
E2A-PBX1-SNP-#6		All 9p						Yes		
E2A-PBX1-SNP-#7										
E2A-PBX1-SNP-#8		Focal								
E2A-PBX1-SNP-#9										Yes
E2A-PBX1-SNP-#10		Focal								
E2A-PBX1-SNP-#11		All 9p								
E2A-PBX1-SNP-#12										
E2A-PBX1-SNP-#13		All 9p								
E2A-PBX1-SNP-#14										
E2A-PBX1-SNP-#15										
E2A-PBX1-SNP-#16		All 9p								
E2A-PBX1-SNP-#17										
TEL-AML1-SNP-#1										Yes
TEL-AML1-SNP-#2										Yes
TEL-AML1-SNP-#3										
TEL-AML1-SNP-#4										Yes
TEL-AML1-SNP-#5	Yes									Yes
TEL-AML1-SNP-#6										Yes
TEL-AML1-SNP-#7										
TEL-AML1-SNP-#8										Yes
TEL-AML1-SNP-#9		Focal								Yes
TEL-AML1-SNP-#10										Yes
TEL-AML1-SNP-#11		Focal								Yes
TEL-AML1-SNP-#12	Yes									Yes
TEL-AML1-SNP-#13										Yes
TEL-AML1-SNP-#14										
TEL-AML1-SNP-#15		Focal								
TEL-AML1-SNP-#16										Yes
TEL-AML1-SNP-#17										
TEL-AML1-SNP-#18		Focal								
TEL-AML1-SNP-#19		Focal , with homozygous 5' deletion								
TEL-AML1-SNP-#20										Yes
TEL-AML1-SNP-#21										Yes
TEL-AML1-SNP-#22										Yes
TEL-AML1-SNP-#23		Focal								Yes
TEL-AML1-SNP-#24										Yes
TEL-AML1-SNP-#25										Yes
TEL-AML1-SNP-#26	Yes									
TEL-AML1-SNP-#27		Focal								Yes
TEL-AML1-SNP-#28		Focal								Yes
TEL-AML1-SNP-#29										Yes
TEL-AML1-SNP-#30										Yes

Case	*EBF1* deletion	*PAX5* deletion	*PAX5* mutation	Aiolos deletion	*BLNK* deletion	*IGLL1* deletion	Ikaros deletion	*LEF1* deletion	*TCF3* deletion	*VPREB1* deletion
TEL-AML1-SNP-#31										Yes
TEL-AML1-SNP-#32		Focal								Yes
TEL-AML1-SNP-#33	Yes	Focal								Yes
TEL-AML1-SNP-#34										Yes
TEL-AML1-SNP-#35										Yes
TEL-AML1-SNP-#36										
TEL-AML1-SNP-#37										
TEL-AML1-SNP-#38		Focal								Yes
TEL-AML1-SNP-#39										Yes
TEL-AML1-SNP-#40										Yes
TEL-AML1-SNP-#41										Yes
TEL-AML1-SNP-#42	Yes	Focal								
TEL-AML1-SNP-#43										
TEL-AML1-SNP-#44					Yes					Yes
TEL-AML1-SNP-#45										
TEL-AML1-SNP-#46		Focal								
TEL-AML1-SNP-#47										
MLL-SNP-#1										
MLL-SNP-#2		All 9p								
MLL-SNP-#3										
MLL-SNP-#4										
MLL-SNP-#5										
MLL-SNP-#6						Yes	Focal			
MLL-SNP-#7			exon 8 FS P321							
MLL-SNP-#8										
MLL-SNP-#9										
MLL-SNP-#10										
MLL-SNP-#11										
BCR-ABL-SNP-#1		Focal								
BCR-ABL-SNP-#2										
BCR-ABL-SNP-#3							Focal			Yes
BCR-ABL-SNP-#4		Broad					Focal			
BCR-ABL-SNP-#5	Yes									Yes
BCR-ABL-SNP-#6										
BCR-ABL-SNP-#7		Focal								
BCR-ABL-SNP-#8										
BCR-ABL-SNP-#9		All 9p					All Chr7*			
Hyperdip47-50-SNP-#1										
Hyperdip47-50-SNP-#2										
Hyperdip47-50-SNP-#3										
Hyperdip47-50-SNP-#4										
Hyperdip47-50-SNP-#5		All 9p	exon 2 V26G							Yes
Hyperdip47-50-SNP-#6		All 9p								
Hyperdip47-50-SNP-#7										Yes
Hyperdip47-50-SNP-#8		Focal	exon 1B FS E2 exon 3 NDTVP126RA							

Case	*EBF1* deletion	*PAX5* deletion	*PAX5* mutation	Aiolos deletion	*BLNK* deletion	*IGLL1* deletion	Ikaros deletion	*LEF1* deletion	*TCF3* deletion	*VPREB1* deletion
Hyperdip47-50-SNP-#9		9p up to PAX5, with focal homozygous deletion exon 7-8								
Hyperdip47-50-SNP-#10		9p up to PAX5; PAX5-ETV6 fusion								
Hyperdip47-50-SNP-#11							Broad			Yes
Hyperdip47-50-SNP-#12										
Hyperdip47-50-SNP-#13										
Hyperdip47-50-SNP-#14										
Hyperdip47-50-SNP-#15										
Hyperdip47-50-SNP-#16										
Hyperdip47-50-SNP-#17										Yes
Hyperdip47-50-SNP-#18										
Hyperdip47-50-SNP-#19		All 9p	exon 2 V26G							
Hyperdip47-50-SNP-#20		Focal								
Hyperdip47-50-SNP-#21										
Hyperdip47-50-SNP-#22										
Hyperdip47-50-SNP-#23										
Hypodip-SNP-#1		9p up to PAX5					All 7p			
Hypodip-SNP-#2		All 9p	exon 9 G338W							Yes
Hypodip-SNP-#3		9p up to PAX5; PAX5-ZNF521 fusion								
Hypodip-SNP-#4		All 9p	exon 2 V26G				Focal			Yes
Hypodip-SNP-#5	Yes	Broad	exon 2 P34Q				Broad			
Hypodip-SNP-#6		All 9p	exon 3 P80R							
Hypodip-SNP-#7		Amplified		Yes		Yes	All Chr 7			
Hypodip-SNP-#8		All 9p		Yes			All Chr 7	Yes		
Hypodip-SNP-#9		Broad								Yes
Hypodip-SNP-#10		All 9p	exon 3 P80R							
Other-SNP-#1										
Other-SNP-#2							All Chr 7			
Other-SNP-#3	Yes	Focal					Focal			
Other-SNP-#4		Focal, with focal homozygous deletion of exon 8								Yes
Other-SNP-#5		All 9p								
Other-SNP-#6										
Other-SNP-#7		All 9p								
Other-SNP-#8										
Other-SNP-#9							Focal			Yes
Other-SNP-#10			exon 3 P80R							

Case	*EBF1* deletion	*PAX5* deletion	*PAX5* mutation	Aiolos deletion	*BLNK* deletion	*IGLL1* deletion	Ikaros deletion	*LEF1* deletion	*TCF3* deletion	*VPREB1* deletion
			and exon 6 FS N310							
Other-SNP-#11										Yes
Other-SNP-#12										
Other-SNP-#13										
Other-SNP-#14		9p up to PAX5; PAX5-FOXP1 fusion								
Other-SNP-#15										
Other-SNP-#16										
Pseudodip-SNP-#1		9p up to PAX5; PAX5-ETV6 fusion								
Pseudodip-SNP-#2		All 9p	exon 3 P80R							
Pseudodip-SNP-#3										
Pseudodip-SNP-#4		Focal	exon 9 FS S339							Yes
Pseudodip-SNP-#5										
Pseudodip-SNP-#6		All 9p					Broad			
Pseudodip-SNP-#7										
Pseudodip-SNP-#8										
Pseudodip-SNP-#9		Broad								
Pseudodip-SNP-#10										Yes
Pseudodip-SNP-#11		Focal								
Pseudodip-SNP-#12								Yes		
Pseudodip-SNP-#13		Focal				Yes				
Pseudodip-SNP-#14				Yes						
Pseudodip-SNP-#15										
Pseudodip-SNP-#16										Yes
Pseudodip-SNP-#17			IVS9+1							
Pseudodip-SNP-#18										
Pseudodip-SNP-#19					Yes					
Pseudodip-SNP-#20							Broad			
T-ALL-SNP-#1										
T-ALL-SNP-#2										
T-ALL-SNP-#3		Broad					Broad			
T-ALL-SNP-#4										
T-ALL-SNP-#5										
T-ALL-SNP-#6		All 9p								
T-ALL-SNP-#7										
T-ALL-SNP-#8										Yes
T-ALL-SNP-#9										
T-ALL-SNP-#10										
T-ALL-SNP-#11										
T-ALL-SNP-#12										
T-ALL-SNP-#13										
T-ALL-SNP-#14										
T-ALL-SNP-#15										

Case	*EBF1* deletion	*PAX5* deletion	*PAX5* mutation	Aiolos deletion	*BLNK* deletion	*IGLL1* deletion	Ikaros deletion	*LEF1* deletion	*TCF3* deletion	*VPREB1* deletion
T-ALL-SNP-#16										
T-ALL-SNP-#17										
T-ALL-SNP-#18										
T-ALL-SNP-#19		Broad								
T-ALL-SNP-#20		Broad						Yes		
T-ALL-SNP-#21										
T-ALL-SNP-#22										
T-ALL-SNP-#23										
T-ALL-SNP-#24										
T-ALL-SNP-#25										
T-ALL-SNP-#26										
T-ALL-SNP-#27										
T-ALL-SNP-#28										
T-ALL-SNP-#29										
T-ALL-SNP-#30										
T-ALL-SNP-#31										Yes
T-ALL-SNP-#32										
T-ALL-SNP-#33										
T-ALL-SNP-#34										
T-ALL-SNP-#35										
T-ALL-SNP-#36										
T-ALL-SNP-#37										
T-ALL-SNP-#38										
T-ALL-SNP-#39										
T-ALL-SNP-#40										
T-ALL-SNP-#41										
T-ALL-SNP-#42										
T-ALL-SNP-#43										
T-ALL-SNP-#44										
T-ALL-SNP-#45										
T-ALL-SNP-#46										
T-ALL-SNP-#47										
T-ALL-SNP-#48										
T-ALL-SNP-#49		All 9p								
T-ALL-SNP-#50										

Supplementary Table 24. Cases with multiple genomic lesions in the B cell differentiation pathway

FS, frameshift

Case	*EBF1* deletion	*PAX5* deletion	*PAX5* mutation	Aiolos deletion	Ikaros deletion	*LEF1* deletion	*TCF3* deletion
E2A-PBX1-SNP-#6		All 9p				Yes	
TEL-AML1-SNP-#33	Yes	Focal					
TEL-AML1-SNP-#42	Yes	Focal					
MLL-SNP-#6					Yes		
BCR-ABL-SNP-#4		Broad			Yes		
BCR-ABL-SNP-#9		All 9p			Yes		
Hyperdip47-50-SNP-#5		All 9p	exon 2 V26G				
Hyperdip47-50-SNP-#8		Focal	exon 3 NDTVP126RA				
Hyperdip47-50-SNP-#19		All 9p	exon 2 V26G				
Hypodip-SNP-#1		9p up to PAX5			Yes		
Hypodip-SNP-#2		All 9p	exon 9 G338W				
Hypodip-SNP-#4		All 9p	exon 2 V26G		Yes		
Hypodip-SNP-#5	Yes	Broad	exon 2 P34Q		Yes		
Hypodip-SNP-#6		All 9p	exon 3 P80R				
Hypodip-SNP-#7		Amplified		Yes	Yes		
Hypodip-SNP-#8		All 9p			Yes	Yes	
Hypodip-SNP-#10		All 9p	exon 3 P80R				
Other-SNP-#3	Yes	Focal			Yes		
Pseudodip-SNP-#2		All 9p	exon 3 P80R				
Pseudodip-SNP-#4		Focal	exon 9 FS S339				
Pseudodip-SNP-#6		All 9p			Yes		
Pseudodip-SNP-#13		Focal					
T-ALL-SNP-#3		Broad			Yes		
T-ALL-SNP-#20		Broad				Yes	

SUPPLEMENTARY NOTES

39. Ross, M. E. *et al.* Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* **102**, 2951-9 (2003).
40. Yeoh, E. J. *et al.* Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* **1**, 133-43 (2002).
41. Lin, M. *et al.* dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**, 1233-40 (2004).
42. Zhao, X. *et al.* An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* **64**, 3060-71 (2004).
43. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557-72 (2004).
44. Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* **2**, research0032.1-0032.11 (2001).
45. Willenbrock, H. & Fridlyand, J. A comparison study: applying segmentation to array CGH data for downstream analyses. *Bioinformatics* **21**, 4084-91 (2005).
46. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* **Vienna**, Austria (2006).
47. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).
48. Rozen, S. & Skaletsky, H. J. in Bioinformatics Methods and Protocols: Methods in Molecular Biology (eds. Krawetz, S. & Misener, S.) 365-386 (Humana Press, Totowa, NJ, 2000).
49. Molnar, A. *et al.* The Ikaros gene encodes a family of lymphocyte-restricted zinc finger DNA binding proteins, highly conserved in human and mouse. *J Immunol* **156**, 585-92 (1996).
50. Sun, L., Liu, A. & Georgopoulos, K. Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. *Embo J* **15**, 5358-69 (1996).
51. Garvie, C. W., Hagman, J. & Wolberger, C. Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* **8**, 1267-76 (2001).
52. Xu, H. E. *et al.* Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev* **13**, 1263-75 (1999).
53. Strehl, S., Konig, M., Dworzak, M. N., Kalwak, K. & Haas, O. A. PAX5/ETV6 fusion defines cytogenetic entity dic(9;12)(p13;p13). *Leukemia* **17**, 1121-3 (2003).
54. Persons, D. A. *et al.* Retroviral-mediated transfer of the green fluorescent protein gene into murine hematopoietic cells facilitates scoring and selection of transduced progenitors in vitro and identification of genetically modified cells in vivo. *Blood* **90**, 1777-86 (1997).
55. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **99**, 7877-82 (2002).

56. Czerny, T. & Busslinger, M. DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). *Mol Cell Biol* **15**, 2858-71 (1995).
57. Andrews, N. C. & Faller, D. V. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res* **19**, 2499 (1991).
58. Kozmik, Z., Wang, S., Dorfler, P., Adams, B. & Busslinger, M. The promoter of the CD19 gene is a target for the B-cell-specific transcription factor BSAP. *Mol Cell Biol* **12**, 2662-72 (1992).
59. Maier, H., Colbert, J., Fitzsimmons, D., Clark, D. R. & Hagman, J. Activation of the early B-cell-specific mb-1 (Ig-alpha) gene by Pax-5 is dependent on an unmethylated Ets binding site. *Mol Cell Biol* **23**, 1946-60 (2003).
60. Ehrich, M. *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* **102**, 15785-90 (2005).
61. Stanssens, P. *et al.* High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Res* **14**, 126-33 (2004).
62. Rahman, M. *et al.* A repressor element in the 5'-untranslated region of human Pax5 exon 1A. *Gene* **263**, 59-66 (2001).
63. Mahmoud, M. S. & Kawano, M. M. Cloning and analysis of the human Pax-5 gene promoter. *Biochem Biophys Res Commun* **228**, 159-64 (1996).
64. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
65. 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).
66. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
67. Xu, W., Rould, M. A., Jun, S., Desplan, C. & Pabo, C. O. Crystal structure of a paired domain-DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations. *Cell* **80**, 639-50 (1995).
68. Finney, M. The homeodomain of the transcription factor LF-B1 has a 21 amino acid loop between helix 2 and helix 3. *Cell* **60**, 5-6 (1990).
69. Maier, H. *et al.* Requirements for selective recruitment of Ets proteins and activation of mb-1/Ig-alpha gene transcription by Pax-5 (BSAP). *Nucleic Acids Res* **31**, 5483-9 (2003).

FUNDING AND GRANT SUPPORT

This work was supported in part by National Cancer Institute grants PO1 CA71907-07 (J.R.D.), R37 CA-36401 (W.E.E.), CA-21765 (Cancer Center CORE grant to St Jude Children's Research Hospital), the NIH/NIGMS Pharmacogenetics Research Network and Database grants U01 GM61393, U01GM61374 (http://pharmgkb.org/), and by the American Lebanese and Syrian Associated Charities (ALSAC) of St Jude Children's Research Hospital. C.G.M. is supported by a St Jude Children's Research Hospital Physician Scientist Fellowship, a National Health and Medical Research Council (Australia) CJ Martin Postdoctoral Travelling Fellowship, a Royal Australasian College of Physicians/CSL Travelling Fellowship, and a Haematology Society of Australia and New Zealand/AMGEN Fellowship.