Supplementary appendix

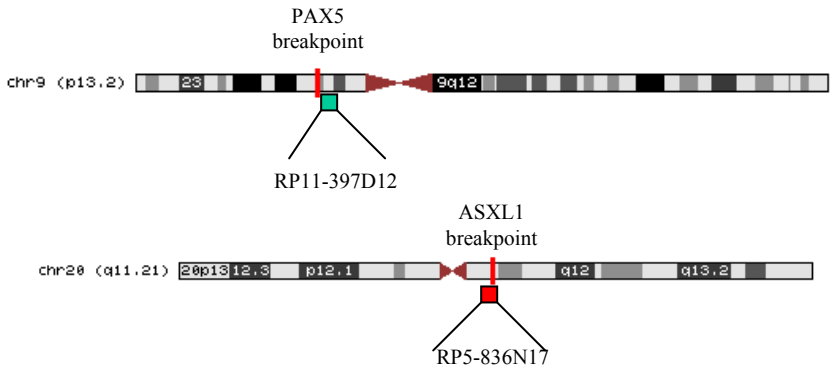
This appendix formed part of the original submission and has been peer reviewed. It is presented as supplied by the authors.

Supplement to: Den Boer ML, van Slegtenhorst M, De Menezes RX, et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncology* 2009; published online Jan 9. DOI:10.1016/S1470-2045(08)70339-5.

Figure 1: FISH analysis of a dic(9;20)(p13;q11) chromosome

A) The genomic position of BAC clones used for FISH analysis is depicted. (B) FISH analysis using RP11-397D12 (green) and RP5-836N17 (red) confirms a dic(9;20) in patient #2 (see also supplementary table 8).

A)



B)

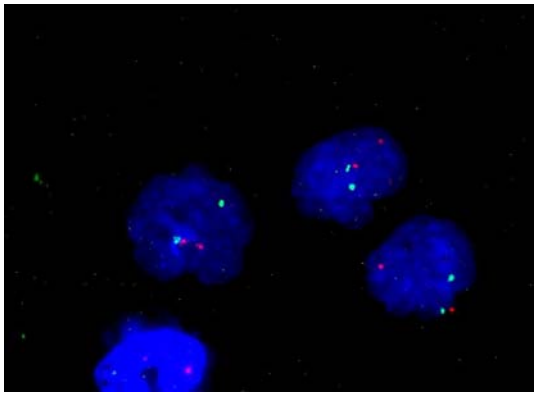


Table 1: Representation of ALL subtypes in training/discovery (COALL) and validation (DCOG) cohort

	COALL cohort		**DCOG cohort**		**Literature****
	training/discovery set		validation set		
	N	%	N	%	%
ALL subtype					
T-lineage ALL	36	*19*	15	*14*	*15*
Precursor B-ALL:					
ETV6-RUNX1 (*TEL-AML1*)-positive	44	*23*	24	*22*	*20-25*
Hyperdiploid*	44	*23*	28	*26*	*25*
TCF3 (*E2A*)-rearranged	13	7	2	*2*	*5*
BCR-ABL1-positive	5	*3*	1	*1*	*<5*
MLL-rearranged	4	*2*	4	*4*	*<5*
B-other	44	*23*	33	*31*	*25*
Total	190	*100*	107	*100*	*100*

* Hyperdiploid: >50 chromosomes and/or DNA index ≥1.16
** Pui CH et al., *New Engl J Med* 2004;**350**:1535-48; Pieters R and Carroll WL, *Pediatr Clin North Am* 2008;**55**:1-20.

Table 2: Number of genes discriminative for 6 major subtypes of pediatric ALL at selected P-value cut-offs[1]

ALL subtype	$P<10^{-8}$ #genes	$P<10^{-6}$ #genes	$P<10^{-4}$ #genes	P<0.001 #genes	P<0.005 #genes	P<0.01 #genes	P<0.05 #genes
T-lineage	988	1722	2871	3831	4777	5283	7005
ETV6-RUNX1-positive	152	293	665	1077	1581	1905	3116
Hyperdiploid	87	242	668	1152	1670	2011	3098
TCF3 (E2A)-rearranged	0	0	23	131	332	473	1052
BCR-ABL1-positive[2]	0	0	0	0	0	0	0
MLL-rearranged[2]	0	0	0	0	0	0	0

[1] Wilcoxon P-value corrected for multiple testing (=false discovery rate)

[2] No significant probe sets were found for *BCR-ABL1*-positive and *MLL*-rearranged cases in the COALL training cohort. Instead of newly selected probe sets, we have used the top-50 of the previously compiled decision tree list reported by Ross ME *et al*. (*Blood* 2003;**102**:2951-9) in the double-loop cross validation procedure to construct our classifier.

Table 3: Definition of diagnostic test values

		True subtype	
		positive	**negative**
Predicted subtype	**positive**	a	b
	negative	c	d

sensitivity	a/(a+c)
specificity	d/(b+d)
positive predictive value (PPV)	a/(a+b)
negative predictive value (NPV)	d/(c+d)
prediction accuracy	(a+d)/(a+b+c+d)

Table 4: Overview of overlapping probe sets identified by 3 gene expression classification studies in pediatric ALL

Probe Set ID	Group	Ratio [1]	Yeoh [2]	Ross [3]	Gene Symbol
200600_at	Hyperdiploid	2.21		*	MSN
200953_s_at	TCF3(E2A)-rearranged	0.08	*	*	CCND2
201132_at	Hyperdiploid	2.12	*	*	HNRPH2
201136_at	Hyperdiploid	3.04		*	PLP2
201153_s_at	MLL-rearranged	4.05	*	*	MBNL1
201443_s_at	Hyperdiploid	2.75	*		ATP6AP2
201579_at	TCF3(E2A)-rearranged	21.50	*	*	FAT
201811_x_at	Hyperdiploid	9.65	*		SH3BP5
202086_at	BCR-ABL1	0.35	*		MX1
202123_s_at	BCR-ABL1	2.35	*	*	ABL1
202478_at	TCF3(E2A)-rearranged	20.84	*		TRIB2
202517_at	ETV6-RUNX1	5.34	*		CRMP1
202838_at	ETV6-RUNX1	2.89	*		FUCA1
203372_s_at	TCF3(E2A)-rearranged	0.04		*	SOCS2
203373_at	TCF3(E2A)-rearranged	0.03	*	*	SOCS2
203434_s_at	MLL-rearranged	0.05		*	MME
203611_at	T-ALL	0.15	*		TERF2
203865_s_at	TCF3(E2A)-rearranged	27.45	*		ADARB1
204069_at	MLL-rearranged	79.80	*	*	MEIS1
204446_s_at	MLL-rearranged	0.02		*	ALOX5
204674_at	TCF3(E2A)-rearranged	6.88	*	*	LRMP
204849_at	ETV6-RUNX1	11.09	*	*	TCFL5
204890_s_at	T-ALL	13.07		*	LCK
204891_s_at	T-ALL	24.00	*	*	LCK
205067_at	Hyperdiploid	3.90	*		IL1B
205821_at	MLL-rearranged	3.37	*		KLRK1
206001_at	T-ALL	0.03	*		NPY
206231_at	ETV6-RUNX1	11.92	*	*	KCNN1
208894_at	MLL-rearranged	0.54		*	HLA-DRA
209695_at	ETV6-RUNX1	3.70	*		PTP4A3
209905_at	MLL-rearranged	194.60	*		HOXA9
210982_s_at	MLL-rearranged	0.53		*	HLA-DRA
211796_s_at	T-ALL	48.81		*	TRBV /// TRBC1
213317_at	ETV6-RUNX1	50.34	*	*	CLIC5
213423_x_at	ETV6-RUNX1	12.90		*	TUSC3
213539_at	T-ALL	90.03	*	*	CD3D
218499_at	Hyperdiploid	1.92		*	MASK
219686_at	T-ALL & MLL-rearranged	0.03 / 0.07		*	STK32B
220451_s_at	ETV6-RUNX1	4.88		*	BIRC7
221747_at	ETV6-RUNX1	4.11		*	TNS1
221748_s_at	ETV6-RUNX1	7.23		*	TNS1
35974_at	TCF3(E2A)-rearranged	7.66		*	LRMP
37986_at	ETV6-RUNX1	7.38		*	EPOR
39318_at	T-ALL	0.03		*	TCL1A

[1] Ratio indicates the ratio in expression level of the indicated gene probe set between the subtype for which the probe set was found discriminative (listed in column with heading 'Group') and the remaining subtypes. T-ALL cases are compared to precursor B-ALL cases and within precursor B-ALL each subtype is compared with the remaining subtypes (e.g. *ETV6-RUNX1*-positive *versus* *ETV6-RUNX1*-negative precursor B-ALL). *ETV6-RUNX1*, *ETV6-RUNX1 (TEL-AML1)*-positive; Hyperdiploid, >50 chromosomes or DNA-index ≥1.16; *BCR-ABL1*, *BCR-ABL1* positive.

[2] The asterix indicates overlapping probe sets out of the list of 110 probe sets found in the current study and those reported by Yeoh EJ et al. (*Cancer Cell* 2002;**1**:133-43) and Ross ME et al. (*Blood* 2003;**102**:2951-9).

Table 5: The predictive value of the gene expression signature to classify 107 DCOG pediatric ALL patients

		True subtype of DCOG cases						
		T-ALL	***ETV6-RUNX1*-positive**	**Hyperdiploid**	***TCF3 (E2A)*-rearranged**	***BCR-ABL1*-positive**	***MLL*-rearranged**	**B-other**
Predicted subtype	**T-ALL**	**15**	0	0	0	0	0	0
	***ETV6-RUNX1*-positive**	0	**24**	0	0	0	0	2
	Hyperdiploid	0	0	**28**	0	0	0	5
	***TCF3 (E2A)*-rearranged**	0	0	0	**2**	0	0	1
	***BCR-ABL1*-positive**	0	0	0	0	**0**	0	0
	***MLL*-rearranged**	0	0	0	0	0	**0**	0
	B-other	0	0	0	0	1	4	**25**

Correctly predicted cases in blue.

Table 6: Overview of misclassified B-other patients in DCOG validation cohort

Patient	Predicted	True karyotype*	Follow up
#1	*ETV6-RUNX1*	46,XY,**del(12)(p12)**	CCR (10.7 yr)
#2	*ETV6-RUNX1*	47,XX,der(7)t(7;12)(p1?2;q1?3),+10,**add(12)(p1?1)**,add(12)(q1?1),der(17)t(12;17)(?;q22)[19]/46,XX[12]	CCR (9.2 yr)
#3	Hyperdiploid	unknown; retrospective FISH revealed -X(14% of cells) and **+18** (48% of cells)**	REL (0.4yr)
#4	Hyperdiploid	46,X,-X,der(1)(q12),der(7)(p11),+mar [2]	REL (2.3 yr)
#5	Hyperdiploid	unknown, DNA index **1.14****	CCR (9.6 yr)
#6	Hyperdiploid	46,XX,r(7),del(8)(p21p23)[17]/45,sl,-7,-r(7)[6]/45,sl,-19[5]/45,sl,add(4)(q23),-7,-r(7)[3]/45,sl,-7,-r(7),-19[2] /46,XX[1]	CCR (9.8 yr)
#7	Hyperdiploid	48,XY**,+X**,t(8;14)(q12;q32),**+21c**[13]/47,XY,**+21c**[13]	CCR (13.1 yr)
#8	*TCF3(E2A)*-rearranged	47,XX,+?i(X)(q10),der(11)t(1;11)(q12;q14),der(16)?add(16)(p?)?del(16)(q?)[2]/47,sl,+der(?)t(?;X)(?;q13)[12] /47,sdl1,-X,?del(X)(q?)[2]/46,XX[2] retrospective FISH revealed **del(19)(p13)** in 21% of cells**	REL (4.0 yr)

***in bold:** genetic abnormality that may explain the predicted subtype
**additional information obtained after gene expression profiling

The majority of misclassified B-other cases could be 'justified' by the presence of additional genetic abnormalities. E.g. in the DCOG validation cohort 8 B-other cases were misclassified: one hyperdiploid-predicted case turned out to be a Down syndrome patient constitutionally having an additional copy of chromosome 21, another hyperdiploid-predicted case had a DNA-index of 1.14 indicating that this patient is near-hyperdiploid and a third predicted hyperdiploid case turned out to have an extra copy of chromosome 18 in half of the cells (more retrospectively obtained information is lacking due to limited material). Two *ETV6-RUNX1* (*TEL-AML1*)-predicted cases turned out to have a (partial) deletion or duplication of chromosome 12p affecting a.o. the *ETV6 (TEL)*-locus. Additional (post-profiling) FISH analysis of the predicted *TCF3 (E2A)*-rearranged case indicated that this case had a 19p13 deletion (affecting the *TCF3 (E2A)*-locus) in a subset of the leukemic cells. This patient was officially diagnosed as B-other based on available karyotypic data whereas this patient would have been assigned to the *TCF3* (*E2A*)-rearranged subtype if cytogenetic analysis by FISH had been performed in the diagnostic procedure (see also Figure 2B of Den Boer ML et al., Lancet Oncology 2009). The two remaining hyperdiploid-predicted cases had a complex karyotype including both numerical and structural abnormalities or karyotype was not assessed.

Table 7: Overview of genomic lesions in *BCR-ABL1*-like cases

Patient#	IKZF1 [1] (IKAROS)	TCF3[1] (E2A)	EBF1[1]	PAX5[1]	VPREB1[1]	Other (molecular) cytogenetic data[2]
1						46,XY[20]
2	focal		focal			
3				deleted		46,XY,del(9)(p21p23)[29]
4	deleted			deleted		45,XY,t(7;9)(p15;p22),t(3;12;10;6;9)(q21;q24;q22;q15;q21)
5				deleted		45,XY,dic(9;20)(p13;q11)
6	deleted					46,XX,r(7),del(8)(p21p23)[17]/45,sl,-7,-r(7)[6]/45,sl,-19[5]/45,sl,add(4)(q23),-7,-r(7)[3]/45,sl,-7,-r(7),-19[2] /46,XX[1]
7	deleted			deleted		46,XX,dic(9;20)(p11;q11),+21
8				focal	deleted	
9			deleted			
10	focal					
11						46,XX,t(10;13)(q24;q14)[12]
12			focal			
13	focal		focal	focal	deleted	
14	focal				deleted	
15	deleted					arr cgh 21q22(AML1)x3 (chr21:30.8-40.4 Mb): iAMP21
16						arr cgh 21q22(AML1)x3 (chr21:23.2-44.0 Mb): iAMP21
17			focal		focal	
18	focal				focal	
19						arr cgh 21q22(AML1)x3 (chr21:24.1-43.1 Mb): iAMP21
20	focal		focal		focal	
21	deleted			deleted		dic(9;20)(p13;q11)
22				deleted		
23				deleted		
24				deleted		46,XY,dup(4)(p?p?),t(8;20)(p2;?),del(9)(p11),inv(12)(q21q24),del(20)(p1?2) or del(20)(q1?2)
25	focal			focal		
26				deleted		
27	focal			deleted	focal	dic(9;20)(p13;q11)
28					focal	
29	focal			deleted		dic(9;20)(p13;q11)
30						
31				focal		46,XX,der(8)(?p),t(9;12)(p12;q14),der(10)(?p),del(11)(q22q25)[12]
32					deleted	
33					deleted	
34					deleted	
35					deleted	
36						
37						
38		deleted		focal	deleted	
39		deleted				
40		deleted				
41	deleted					
42	focal				focal	
43						
44	focal				deleted	

[1] Deleted: deleted area >0.5 Mbp; Focal: deleted area <0.5 Mbp. Blanks: no genetic lesion observed.

[2] Other molecular genetic data comprise information about karyotype and molecular cytogenetics. In addition cases with an intrachromosomal amplification of chromosome 21 (iAMP21) are indicated (data obtained by array-CGH). All cases were negative for *ETV6-RUNX1* (*TEL-AML1)* fusion, hyperdiploidy, *TCF3* (*E2A)*-rearrangement, *MLL*-rearrangement and *BCR-ABL1* fusion.

Table 8: Breakpoints in chromosome 9(p) and 20(q) are conserved in ALL

patient	Break chr 9(p)[1]	Break chr 20(q)[2]	dic(9;20)
1	*PAX5*	*SPAG4L*	No
2	*PAX5*	*ASXL1*	Yes
3	*PAX5*	*ASXL1*	Yes
4	*PAX5*	*DNMT3B*	Yes
5	*ZCCHC7*	*COMMD7*	Not tested
6	*ZCCHC7*	*COMMD7*	Yes
7	*ZCCHC7*	centromere	Yes
8	76685	centromere	Not tested

[1] *PAX5* and *ZCCHC7* are two adjacent genes on chromosome 9(p13), location: 36,828,531-37,348,145
[2] *ASXL1, COMMD7, DNMT3B, SPAG4L* are adjacent genes on chromosome 20(q11), location: 30,420,479-31,055,900