



## Genomewide analysis of cytogenetic aberrations in ETV6/RUNX1 positive childhood acute lymphoblastic leukaemia

**Borst, Louise; Wesolowska, Agata; Joshi, Tejal; Borup, Rehannah; Nielsen, Finn; Andersen, Mette K.; Jonsson, Olafur G.; Wehner, Peder Skov; Wesenberg, Finn; Frost, BrittMarie**

*Total number of authors:*  
12

*Published in:*  
British Journal of Haematology

*Link to article, DOI:*  
[10.1111/j.1365-2141.2012.09083.x](https://doi.org/10.1111/j.1365-2141.2012.09083.x)

*Publication date:*  
2012

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Borst, L., Wesolowska, A., Joshi, T., Borup, R., Nielsen, F., Andersen, M. K., Jonsson, O. G., Wehner, P. S., Wesenberg, F., Frost, B.M., Gupta, R., & Schmiegelow, K. (2012). Genomewide analysis of cytogenetic aberrations in ETV6/RUNX1 positive childhood acute lymphoblastic leukaemia. *British Journal of Haematology*, 157(4), 476-482. <https://doi.org/10.1111/j.1365-2141.2012.09083.x>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Genome-wide analysis of cytogenetic aberrations in *ETV6/RUNX1*-positive childhood acute lymphoblastic leukaemia

Louise Borst,<sup>\*,1</sup> Agata Wesolowska,<sup>\*,2</sup> Tejal Joshi,<sup>2</sup> Rehannah Borup,<sup>3</sup> Finn C. Nielsen,<sup>3</sup> Mette K. Andersen,<sup>4</sup> Olafur G. Jonsson,<sup>5</sup> Peder S. Wehner,<sup>6</sup> Finn Weisenberg,<sup>7</sup> Britt-Marie Frost,<sup>8</sup> Ramneek Gupta<sup>2</sup> and Kjeld Schmiegelow<sup>1,9</sup>

<sup>1</sup>Clinic for Paediatric and Adolescent Medicine, The Juliane Marie Centre, The University Hospital Rigshospitalet, Copenhagen, <sup>2</sup>Centre for Biological Sequence Analysis, The Technical University of Denmark, Kgs. Lyngby, <sup>3</sup>Centre for Genomic Medicine, The University Hospital Rigshospitalet, Copenhagen, <sup>4</sup>Department of Clinical Genetics, The Juliane Marie Centre, The University Hospital, Rigshospitalet, Denmark, <sup>5</sup>Department of Paediatrics, University Hospital, Reykjavik, Iceland, <sup>6</sup>Department of Paediatric Haematology and Oncology, H. C. Andersen Children's Hospital, Odense University Hospital, Odense, Denmark, <sup>7</sup>Department of Paediatrics, National Hospital, Oslo, Norway, <sup>8</sup>Institute of Paediatric, Department of Women's and Children's Health, Uppsala, Sweden and <sup>9</sup>The Institute of Gynaecology, Obstetrics and Paediatrics, The Faculty of Health Sciences, The University of Copenhagen, Copenhagen, Denmark

Received 20 December 2011; accepted for publication 08 February 2012

Correspondence: Kjeld Schmiegelow, Clinic for Paediatric and Youth Medicine, The Juliane Marie Centre, The University Hospital Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark.  
E-mail: kschmiegelow@rh.dk

\*Joint first authorship.

Acute lymphoblastic leukaemia (ALL) is the most common cancer in children (Hjalgrim *et al*, 2003). In B-lineage childhood ALL the most frequent structural cytogenetic abnormality, present in 25% of the cases, is the chromosomal translocation t(12;21)(p13;q22) resulting in the *ETV6/RUNX1* fusion gene (Shurtleff *et al*, 1995; Forestier *et al*,

## Summary

The chromosomal translocation t(12;21) resulting in the *ETV6/RUNX1* fusion gene is the most frequent structural cytogenetic abnormality among patients with childhood acute lymphoblastic leukaemia (ALL). We investigated 62 *ETV6/RUNX1*-positive childhood ALL patients by single nucleotide polymorphism array to explore acquired copy number alterations (CNAs) at diagnosis. The mean number of CNAs was 2.82 (range 0–14). Concordance with available G-band karyotyping and comparative genomic hybridization was 93%. Based on three major protein-protein complexes disrupted by these CNAs, patients could be categorized into four distinct subgroups, defined by different underlying biological mechanisms relevant to the aetiology of childhood ALL. When recurrent CNAs were evaluated by an oncogenetic tree analysis classifying their sequential order, the most common genetic aberrations (deletions of 6q, 9p, 13q and X, and gains of 10 and 21) seemed independent of each other. Finally, we identified the most common regions with recurrent gains and losses, which comprise microRNA clusters with known oncogenic or tumour-suppressive roles. The present study sheds further light on the genetic diversity of *ETV6/RUNX1*-positive childhood ALL, which may be important for understanding poor responses among this otherwise highly curable subset of ALL and lead to novel targeted treatment strategies.

**Keywords:** childhood acute lymphoblastic leukaemia, *ETV6/RUNX1* translocation, genome-wide screening.

2008). The event-free survival (EFS) of childhood ALL patients after first line therapy is approximately 80% (Schmiegelow *et al*, 2010), however presence of the *ETV6/RUNX1* fusion gene is generally associated with a more favourable prognosis, with an EFS of approximately 90% (Forestier *et al*, 2008). The *ETV6/RUNX1* fusion gene is believed to

represent an initiating event in the development of this childhood ALL subset. This initiating event, seen as early as *in utero* (Wiemels *et al*, 1999; Hjalgrim *et al*, 2002), most likely requires additional genetic changes in order to lead to leukaemia e.g. *ETV6/RUNX1* transcripts have been demonstrated in normal cord blood samples at low levels (Mori *et al*, 2002; Lausten-Thomsen *et al*, 2011) and neither *ETV6/RUNX1*-positive knock-in mouse models nor carriers of *ETV6/RUNX1*-positive pre-leukaemic cells will all develop leukaemia (Fischer *et al*, 2005; Hong *et al*, 2008). Thus, the development of *ETV6/RUNX1*-positive leukaemia requires one or more secondary genetic changes and previous studies have indeed shown additional genetic changes in the majority of *ETV6/RUNX1*-positive childhood ALL patients (Forestier *et al*, 2007; Mullighan *et al*, 2007; Kawamata *et al*, 2008; Liljebjorn *et al*, 2007, 2010; Parker *et al*, 2008; Mullighan *et al*, 2008; van Delft *et al*, 2011).

In the present study, we applied high-resolution Affymetrix GeneChip® 500K profiling to 62 *ETV6/RUNX1*-positive childhood ALL patients to explore in detail copy number alterations (CNAs) throughout the genome. Furthermore, we investigated the inter-dependencies of the recurrent CNAs and involvement in protein-protein complexes among them, and correlated the CNAs with biologically relevant microRNAs (miRNAs) affected by the aberrations.

## Methods and materials

### Patients

The *ETV6/RUNX1*-positive childhood ALL patients (aged between 1 and 15 years at the time of diagnosis) from Denmark, Norway, Sweden and Iceland were diagnosed and enrolled in the Nordic Society for Paediatric Haematology and Oncology (NOPHO) treatment protocols (NOPHO ALL-92 or NOPHO ALL-2000) (Schmiegelow *et al*, 2010). A total of 133 patients were found to be positive for *ETV6/RUNX1* by fluorescent *in situ* hybridization (FISH) and/or reverse transcription polymerase chain reaction (RT-PCR) by routine investigation from 1996–2007, and diagnostic tumour DNA samples were available for 99 of these patients. Ten samples were of poor quality and/or had insufficient amount of DNA to be analysed and of the remaining samples, 62 fulfilled the quality criteria (QC) for inclusion in the bioinformatics analyses (Table 1). The study was approved by The Capital Region of Denmark Committee on Biomedical Research Ethics and The Danish Data Protection Agency.

### G-band karyotyping and comparative genomic hybridization

G-band karyotyping was performed as part of the routine investigation on short-term cultured leukaemic cells by standard techniques. As part of a separate study, high-resolution comparative genomic hybridization (CGH) was performed

**Table 1.** Patient characteristics and copy number alterations.

Characteristics	Patients
Gender	
Male	37 (59.7%)
Female	25 (40.3%)
Age, years, median (range)	4.17 (1.30–15)
WBC, $\times 10^9/l$ , median (range)	11.5 (0.80–110)
Risk group	
SR	28 (45.2%)
IR	23 (37.1%)
HR	11 (17.7%)
Treatment protocol	
NOPHO ALL-92	33 (53.2%)
NOPHO ALL-2000	29 (46.8%)
Events	
Induction failure	0 (0%)
Resistant disease	0 (0%)
Relapse	4 (6.5%)
Death in remission	0 (0%)
Secondary malignancy	0 (0%)
Total events	4 (6.5%)
EFS	93.5%
CNA (mean)	
Gains > 1 Mb	0.73 (0–8)
Gains < 1 Mb	0.02 (0–1)
Losses > 1 Mb	1.24 (0–8)
Losses < 1 Mb	0.85 (0–6)
Total > 1 Mb	1.97 (0–9)
Total < 1 Mb	0.85 (0–6)
Total	2.82 (0–14)

WBC, white blood cell count; SR, standard risk; IR, intermediate risk; HR, high risk; NOPHO, Nordic Society for Paediatric Haematology and Oncology; EFS, event-free survival; CNA, copy number alteration; Mb, megabase.

for a subset of the patients (Kristensen *et al*, 2003). G-band karyotyping and/or CGH data was available in 54 (87.1%) of the included cases (Table S1).

### Single nucleotide polymorphism (SNP) array analysis

DNA from bone marrow or blood samples acquired at diagnosis was extracted and purified by sodium chloride and ethanol precipitation. For SNP array analysis the DNA was processed and hybridized to Affymetrix GeneChip® Mapping 500K array set according to the manufacturer's instructions (Affymetrix, Santa Clara, CA, USA). The Affymetrix GeneChip® Mapping 500K array set comprises two arrays and is processed by two assay kits differing only in the restriction enzyme used (Affymetrix GeneChip® Mapping 250K Nsp Assay Kit and Affymetrix GeneChip® Mapping 250K Sty Assay Kit). Briefly, 250 ng DNA was digested by either Nsp I or Sty I, followed by ligation of adaptors, allowing PCR amplification of fragments in sizes from 200 to 1100 bp. PCR products were fragmented and end-labelled with biotin. Samples were subsequently hybridized to the arrays. The

arrays were washed and stained with phycoerythrin-conjugated streptavidin (SAPE) using the Affymetrix Fluidics Station<sup>®</sup> 450 and were scanned in the Affymetrix GeneChip<sup>®</sup> 2500 scanner to generate fluorescent images, as described in the Affymetrix GeneChip<sup>®</sup> protocol. Cell intensity files (CEL files) were generated in the GeneChip<sup>®</sup> operating software (GCOS) version 5.0.

### Data processing

The signal intensity data based on the raw CEL files was generated using the Affymetrix Power Tools (APT) Software Package. First, for each array a QC call rate was generated using the Dynamic Model algorithm (Di *et al*, 2005). According to the manufacturer's recommendations, only samples achieving call rates  $\geq 93\%$  were used for further analysis. For the 62 samples and 200 controls fulfilling QC, genotype calls were generated using the BRLMM algorithm (<http://array.mc.vanderbilt.edu/microarray/dna/brlmm.pdf>). Allele-specific signals were extracted from median polish summarized and quantile normalized Perfect Match (PM) probe-intensity values. The R-GADA package (Pique-Regi *et al*, 2010) was employed to detect the recurrent CNAs. The package implements a segmentation algorithm to call CNAs based on genome alternation detection analysis (GADA). During the segmentation procedure the parameters controlling the trade-off between the sensitivity and false discovery rate (FDR) were set to achieve high specificity at the cost of sensitivity. The minimum segment length required for classification as a CNA was 8 probes, in order to exclude detection of false alterations due to extreme outliers. The 200 controls were used to exclude alterations representing the normal CNAs. Final estimation of copy number states was verified by visual examination.

### Data visualization and biological correlations

The display of all the CNAs in circos format was done according to Krzywinski *et al* (2009). Recurrent CNAs were defined as gains or deletions of material occurring in the same region in at least two patients and plotted as a heatmap (Fig S1). The recurrent CNAs were then analysed further with a CRAN R package Oncotree 0.31 to determine a tree model for oncogenesis (Desper *et al*, 1999; Szabo & Boucher, 2002).

To explore functional modules that are potentially affected by the CNAs, enrichment analysis on protein-protein interaction complexes was performed using custom written Perl scripts. Only patients with losses of genetic material were included in this analysis, as these are likely to result in a direct loss of function and hence are most disruptive to the complexes. Using a curated collection of protein-protein complexes (Lage *et al*, 2007), we defined a set of complexes that categorized the patients into distinct subgroups defined by different underlying biological mechanisms. All combinations of 2–5 protein-protein complexes were tested to find a

set of complexes representing the largest possible group of patients with the smallest possible overlap between different complexes. From the final collection of complex sets that best fulfilled the criteria, only those that comprised of genes highly expressed in B lymphoblasts were selected and prioritized according to leukaemia-related annotations among Bio-Alma terms assigned to the complexes. The miRNA annotations were based on miRBase version 16 (Griffiths-Jones, 2010) while their experimentally validated target information was collected from miRecords and miRWalk databases (Xiao *et al*, 2009; Dweep *et al*, 2011).

## Results

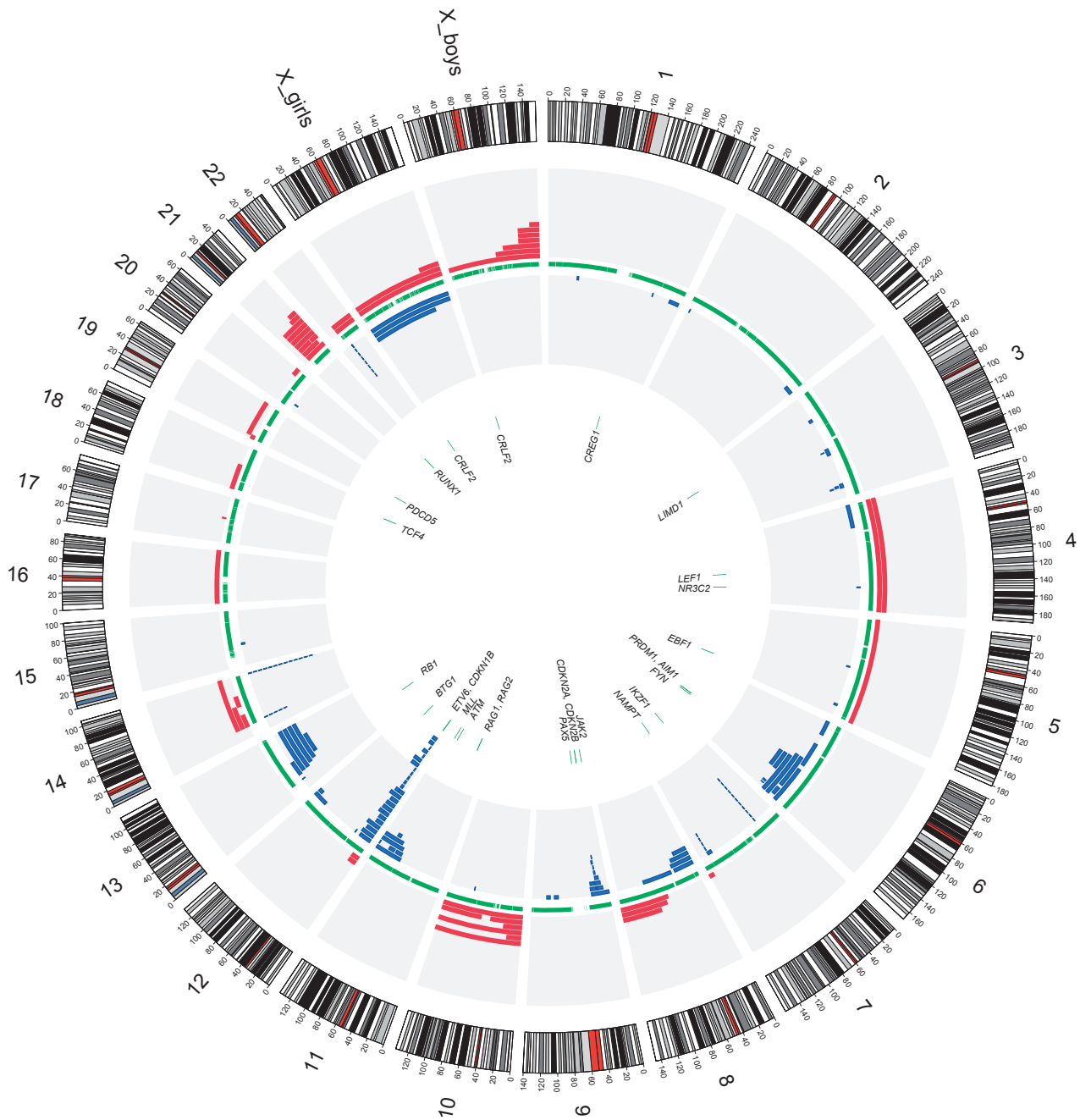
### Genomic profiling of ETV6/RUNX1-positive childhood ALL patients

The SNP array analysis revealed a total of 174 CNAs (30 were recurrent) among the 62 successfully analysed cases (whole chromosome deletions and amplifications each counted as one cytogenetic change). Of these, 129 were deletions, 77 above 1 Mb (including two whole X chromosome deletions), and 52 focal deletions below 1 Mb. A total of 45 gains were detected, of which 44 were above 1 Mb (including 18 whole chromosome gains), and only one focal amplification below 1 Mb (Fig 1 and Table S1). Since the 500K arrays do not have any probes covering the p arms of chromosomes 13, 14, 15, 21 and 22, some large amplifications of these chromosomes may actually be trisomies. The mean number of CNAs per patient was 2.82 (range 0–14) and 13 cases (21%) did not display any CNAs (Table 1 and Table S1).

Concordance between G-band karyotyping and/or CGH and the SNP array analysis data was 93%. Thus, in 51 of the 54 cases with available data, the SNP array results either confirmed and/or added to G-band karyotyping and/or CGH (Table S1). As summarized in Fig 1, all but chromosome 17 demonstrated CNAs above 1 Mb. If focal lesions are included, all chromosomes demonstrated alterations. The most common CNAs above 1 Mb involved deletions of 12p (39%), 6q (13%), 9p (10%), 11q (10%), 13q (10%), 8p (8%) and amplifications of Xq (11%) (one female and six males), 21 (10%), 10 (5%) and Xp (5%) (two females and one male). Of the focal changes, the most common included deletions in 14q32.33 (21%), 7p14.1 (18%), 22q11.22 (13%), 14q11.2 (8%) and 7q34 (6%).

### Biological correlations

Overall, neither individual aberrations nor heatmap clustering of the recurrent aberrations (Fig S1) correlated significantly with age or white blood cell count (WBC). In an attempt to classify the patients more distinctively, a systems biology approach was applied investigating the likely effects of losses of genetic material at the protein complex level. Grouping of patients in this way reflects more closely the clinical param-



**Fig 1.** The recurrent cytogenetic alterations in *ETV6/RUNX1* patients plotted in Circos format (Krzywinski *et al*, 2009). The chromosomes are presented clockwise with the p arm starting from 0, followed by the centromere marked in red and by the q arm. The gains are shown in red, and the deletions in blue. The location of the probes from the 500K arrays is shown in green. The location of genes found to be important in other microarray based ALL studies is marked by green lines in the middle of the plot. Note: the alterations are stacked together; so one circle does not necessarily represent one patient.

ters and suggests different potential biological mechanisms underlying the leukaemogenesis (Data S1, Fig S2).

Next, the inter-dependency of the recurrent CNAs was investigated by an oncogenetic tree analysis to explore the sequential order of CNAs as previously done by Lilljebjorn *et al* (2010). By this approach, deletions of 12p, 14q32 and gain of 21q were classified as early events, i.e. being close to

the root and acting as new roots of their own subtree (Fig 2). Importantly, even though gain of 12p (the most common additional aberration) was close to the root, it did not seem to be a required precursor for other changes. Furthermore, the most common aberrations in *ETV6/RUNX1*-positive patients (deletions of 6q, 9p, 12p, 13q, and X, and chromosome 10 and 21 amplifications) (Forestier

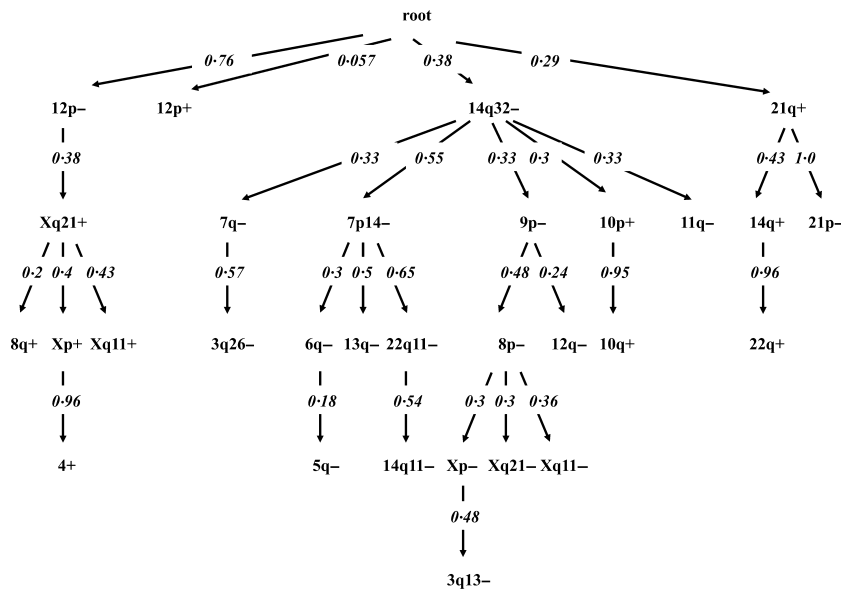


Fig 2. Oncogenetic tree. A probabilistic model of the sequential order of acquiring the copy number alterations. The root of the tree represents the initial event - the t(12;21) chromosomal translocation. The numbers at each edge indicate the probabilities of transition along the given edge by the time of observation.

*et al*, 2007), seemed independent of each other, i.e. the occurrence of one did not change the likelihood of occurrence of the others. Still, 9p deletions seemed to precede 8p deletions and chromosome X deletions were rooted in 8p deletion according to the oncogenetic branching model (Fig 2). The reported estimated false-positive error rate was 0.0007 and the false-negative error rate was 0.30.

Finally, regions undergoing recurrent aberrations in the patients were also investigated for transcribed miRNAs. miRNAs can regulate protein-coding gene targets, thus they may play an oncogenic or tumour suppressive role, depending on their transcription levels, DNA methylation and CpG content of the surrounding regions, mutations and the genes they regulate (Data S2). The most frequently deleted regions, containing important miRNA clusters, were on 11q and 13q, while the most frequently amplified regions were on chromosomes 21 and X, comprising other miRNA clusters (Data S2 and Table S2).

## Discussion

The present study sheds further light on the genetic diversity of *ETV6/RUNX1*-positive childhood ALL. Overall, the mean number of alterations per patient (2.82, range 0–14) is in agreement with previous studies of *ETV6/RUNX1*-positive childhood ALL patients (Forestier *et al*, 2007; Mullighan *et al*, 2007; Parker *et al*, 2008; Lilljebjorn *et al*, 2010; van Delft *et al*, 2011). Higher specificity was prioritized in attempt to reduce the rate of false positive findings, at the cost of reduced sensitivity, especially with regard to detection of focal aberrations. Furthermore, due to the size of the study, some recurrent changes may have been missed if they

did not occur in at least two patients. On the other hand, the strict criteria increased the reliability of the oncogenetic tree and systems biology analyses, and is furthermore supported by the high concordance with previous results of G-band karyotyping and CGH analyses. The larger alterations: 12p, 6q, 9p, 11q, 13q and gains of chromosome 21 and Xq occurred in at least 10% of the patients, supporting previous reports (Forestier *et al*, 2007; Mullighan *et al*, 2007; Kawamata *et al*, 2008; Lilljebjorn *et al*, 2007, 2010; Parker *et al*, 2008; Mullighan *et al*, 2008; van Delft *et al*, 2011). Moreover, gain of Xq is far more common in males than in females, which supports the previous findings of Lilljebjorn *et al* (2007); (Lilljebjorn *et al*, 2010) and could imply loss of X chromosome silencing of that specific region in females leading to similar gene dosage effects in both sexes. The affected gene(s) remains to be determined.

Importantly, these non-random aberrations should be interpreted by their putative biological consequence. Here we proposed an alternative grouping of patients based on integrative analysis of recurrent CNAs and involved protein-protein interaction complexes, which may suggest different underlying biological mechanisms. This classification reflects that deletions of different genes, often from different chromosomes, may lead to disruption of the same complex resulting in a similar phenotype. In most cases the loss of a gene, evidenced through CNAs, leads to perturbation of interaction with another gene with high connectivity in the complex, and therefore is likely to have a functional effect.

To speculate on the relationship and order by which these CNAs occur, we used a branching oncogenetic tree model. The model identified deletions on 12p, 14q32 and gain of

21q as important early leukaemogenic events secondary to the *ETV6/RUNX1* translocation. Later, independent events include the most common changes in *ETV6/RUNX1*-positive childhood ALL patients, such as deletions of 6q, 9p, 13q and X and amplification of chromosome 10, suggesting these changes occur later in the development of the leukaemic clone. These implications need to be confirmed, if possible, in studies directly observing the sequential order of occurrence of CNAs. A recent study also demonstrated a similar model with data from 164 *ETV6/RUNX1*-positive childhood ALL patients (Liljebjorn *et al*, 2010). The authors demonstrated independence between deletions on 6q, 9p and 12p, however with 6q and 9p closer to the root. Also amplification of chromosome 21 was far down a subtree rooting from 6q deletion. However, the two models cannot be compared directly because the resolution of analysed aberrations differed between the two studies. Furthermore, Anderson *et al* (2011) has also recently performed an oncogenetic tree mapping, however the study only explored pre-selected CNAs at single cell level in individual patients. Thus, the results and those from the present study are not directly comparable, but the nonlinear branching nature of the sequence of the acquired CNAs is alike.

Finally, we associated recurrent CNAs with miRNAs transcribed in those regions in order to elucidate the consequences on gene regulation of the investigated miRNAs targets. Here we observed deletions and amplifications in several chromosomal regions containing miRNAs known to be associated with tumorigenicity, leukaemia and regulation of apoptosis among others (Data S2). Thus, miRNAs may play a crucial role in childhood ALL by targeting a number of proteins and affecting various functional pathways.

Even though many clinical trials classify *ETV6/RUNX1*-positive childhood ALL as one group, their cytogenetic diversity calls for refinement in the classification of these patients, which ultimately may lead to a better understanding of their natural history (including determination of pre- and postnatal hits) (Greaves, 2006), the clinical presentation and their specific treatment requirements. Based on CNAs in protein-protein complexes, we described four groups of patients. This analysis could be further refined in the future with a larger number of samples and higher density arrays.

## Acknowledgements

First of all, we are very grateful to the patients who participated in the study and their referring physicians. We also thank Hanne Maage and Charlotte Scherling for very helpful technical assistance. Acknowledgements are also made to The RH Microarray Centre/KB4111 – Department of Clinical

Biochemistry – The University Hospital Rigshospitalet, Copenhagen for providing technology consultation and laboratory resources. This study has received financial support from Ministry of Health (Grant no. 2006-12103-250), The Novo Nordisk Foundation, The Danish Research Council for Health and Disease (Grant no 271-06-0278 and 271-08-0684), The Danish Childhood Cancer Foundation. Kjeld Schmiegelow holds The Danish Childhood Cancer Foundation Professorship in Paediatric Oncology.

## Author contributions

LB, AW and KS designed the study, interpreted data and drafted the manuscript. LB performed the SNP array analyses in collaboration with RB and FCN, and collected patient samples and clinical data. AW performed the data processing, visualization and biological correlations. TJ performed the miRNA analyses, interpreted miRNA data and provided critical input to the manuscript. MKA was responsible for the G-band karyotyping and CGH data, while OGJ, PSW, FW and BMF provided patient samples. RG and KS contributed to the data analyses and interpretation, together with critical input to the writing of the manuscript. All authors approved the submitted and final versions of the manuscript.

## Conflict of interest

The authors have no competing interests.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Data S1.** Protein-protein complexes analyses.

**Data S2.** MicroRNAs.

**Table S1.** Comparison of chromosomal alterations analysed by G-band karyotyping, CGH and SNP array analysis.

**Table S2.** MicroRNA clusters located in recurrent amplified and deleted regions.

**Figure S1.** Heatmap clustering of patients and their recurrent aberrations summarized in cytoband, chromosome arm or whole chromosome regions.

**Figure S2.** Protein-protein complexes.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## References

- Anderson, K., Lutz, C., van Delft, F.W., Bateman, C.M., Guo, Y., Colman, S.M., Kempfski, H., Moorman, A.V., Tittle, I., Swansbury, J., Kearney, L., Enver, T. & Greaves, M. (2011) Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*, **469**, 356–361.
- van Delft, F.W., Horsley, S., Colman, S., Anderson, K., Bateman, C., Kempfski, H., Zuna, J., Eckert, C., Saha, V., Kearney, L., Ford, A. & Greaves, M. (2011) Clonal origins of relapse in *ETV6*-

- RUNX1 acute lymphoblastic leukemia. *Blood*, **117**, 6247–6254.
- Desper, R., Jiang, F., Kallioniemi, O.P., Moch, H., Papadimitriou, C.H. & Schaffer, A.A. (1999) Inferring tree models for oncogenesis from comparative genome hybridization data. *Journal of Computational Biology*, **6**, 37–51.
- Di, X., Matsuzaki, H., Webster, T.A., Hubbell, E., Liu, G., Dong, S., Bartell, D., Huang, J., Chiles, R., Yang, G., Shen, M.M., Kulp, D., Kennedy, G. C., Mei, R., Jones, K.W. & Cawley, S. (2005) Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. *Bioinformatics*, **21**, 1958–1963.
- Dweep, H., Sticht, C., Pandey, P. & Gretz, N. (2011) miRWalk-database: prediction of possible miRNA binding sites by “walking” the genes of three genomes. *Journal of Biomedical Informatics*, **44**, 839–847.
- Fischer, M., Schwieger, M., Horn, S., Niebuhr, B., Ford, A., Roscher, S., Bergholz, U., Greaves, M., Lohler, J. & Stocking, C. (2005) Defining the oncogenic function of the TEL/AML1 (ETV6/RUNX1) fusion protein in a mouse model. *Oncogene*, **24**, 7579–7591.
- Forestier, E., Andersen, M.K., Autio, K., Blennow, E., Borgstrom, G., Golovleva, I., Heim, S., Heinonen, K., Hovland, R., Johannsson, J.H., Kerndrup, G., Nordgren, A., Rosenquist, R., Swolin, B. & Johannsson, B. (2007) Cytogenetic patterns in ETV6/RUNX1-positive pediatric B-cell precursor acute lymphoblastic leukemia: a Nordic series of 245 cases and review of the literature. *Genes, Chromosomes and Cancer*, **46**, 440–450.
- Forestier, E., Heyman, M., Andersen, M.K., Autio, K., Blennow, E., Borgstrom, G., Golovleva, I., Heim, S., Heinonen, K., Hovland, R., Johannsson, J.H., Kerndrup, G., Nordgren, A., Rosenquist, R., Swolin, B. & Johannsson, B. (2008) Outcome of ETV6/RUNX1-positive childhood acute lymphoblastic leukaemia in the NOPHO-ALL-1992 protocol: frequent late relapses but good overall survival. *British Journal of Haematology*, **140**, 665–672.
- Greaves, M. (2006) Infection, immune responses and the aetiology of childhood leukaemia. *Nature Reviews Cancer*, **6**, 193–203.
- Griffiths-Jones, S. (2010) miRBase: microRNA sequences and annotation. *Current Protocols in Bioinformatics*, **29**, 12.9.1–12.9.10.
- Hjalgrim, L.L., Madsen, H.O., Melbye, M., Jorgensen, P., Christiansen, M., Andersen, M.T., Pallisgaard, N., Hokland, P., Clausen, N., Ryder, L. P., Schmiegelow, K. & Hjalgrim, H. (2002) Presence of clone-specific markers at birth in children with acute lymphoblastic leukaemia. *British Journal of Cancer*, **87**, 994–999.
- Hjalgrim, L.L., Rostgaard, K., Schmiegelow, K., Soderhall, S., Kolmannskog, S., Vetenranta, K., Kristinsson, J., Clausen, N., Melbye, M., Hjalgrim, H. & Gustafsson, G. (2003) Age- and sex-specific incidence of childhood leukemia by immunophenotype in the Nordic countries. *Journal of the National Cancer Institute*, **95**, 1539–1544.
- Hong, D., Gupta, R., Ancliff, P., Atzberger, A., Brown, J., Soneji, S., Green, J., Colman, S., Piacibello, W., Buckle, V., Tsuzuki, S., Greaves, M. & Enver, T. (2008) Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. *Science*, **319**, 336–339.
- Kawamata, N., Ogawa, S., Zimmermann, M., Kato, M., Sanada, M., Hemminki, K., Yamamoto, G., Nannya, Y., Koehler, R., Flohr, T., Miller, C.W., Harbott, J., Ludwig, W.D., Stanulla, M., Schrappe, M., Bartram, C.R. & Koefler, H.P. (2008) Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood*, **111**, 776–784.
- Kristensen, T.D., Wesenberg, F., Jonsson, O.G., Carlsen, N.T., Forestier, E., Kirchoff, M., Lundsteen, C. & Schmiegelow, K. (2003) High-resolution comparative genomic hybridisation yields a high detection rate of chromosomal aberrations in childhood acute lymphoblastic leukaemia. *European Journal of Haematology*, **70**, 363–372.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. & Marra, M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.
- Lage, K., Karlberg, E.O., Stirling, Z.M., Olason, P. I., Pedersen, A.G., Rigina, O., Hinsby, A.M., Tumer, Z., Pociot, F., Tommerup, N., Moreau, Y. & Brunak, S. (2007) A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnology*, **25**, 309–316.
- Lausten-Thomsen, U., Madsen, H.O., Vestergaard, T.R., Hjalgrim, H., Nersting, J. & Schmiegelow, K. (2011) Prevalence of t(12;21)[ETV6-RUNX1]-positive cells in healthy neonates. *Blood*, **117**, 186–189.
- Lilljebjorn, H., Heidenblad, M., Nilsson, B., Lassen, C., Horvat, A., Heldrup, J., Behrendtz, M., Johannsson, B., Andersson, A. & Fioretos, T. (2007) Combined high-resolution array-based comparative genomic hybridization and expression profiling of ETV6/RUNX1-positive acute lymphoblastic leukemias reveal a high incidence of cryptic Xq duplications and identify several putative target genes within the commonly gained region. *Leukemia*, **21**, 2137–2144.
- Lilljebjorn, H., Sonesson, C., Andersson, A., Heldrup, J., Behrendtz, M., Kawamata, N., Ogawa, S., Koefler, H.P., Mitelman, F., Johannsson, B., Fontes, M. & Fioretos, T. (2010) The correlation pattern of acquired copy number changes in 164 ETV6/RUNX1-positive childhood acute lymphoblastic leukemias. *Human Molecular Genetics*, **19**, 3150–3158.
- Mori, H., Colman, S.M., Xiao, Z., Ford, A.M., Healy, L.E., Donaldson, C., Hows, J.M., Navarrete, C. & Greaves, M. (2002) Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proceedings of the National Academy of Sciences USA*, **99**, 8242–8247.
- Mullighan, C.G., Goorha, S., Radtke, I., Miller, C. B., Coustan-Smith, E., Dalton, J.D., Girtman, K., Mathew, S., Ma, J., Pounds, S.B., Su, X., Pui, C.H., Relling, M.V., Evans, W.E., Shurtleff, S.A. & Downing, J.R. (2007) Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*, **446**, 758–764.
- Mullighan, C.G., Miller, C.B., Radtke, I., Phillips, L.A., Dalton, J., Ma, J., White, D., Hughes, T.P., Le Beau, M.M., Pui, C.H., Relling, M.V., Shurtleff, S.A. & Downing, J.R. (2008) BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature*, **453**, 110–114.
- Parker, H., An, Q., Barber, K., Case, M., Davies, T., Konn, Z., Stewart, A., Wright, S., Griffiths, M., Ross, F.M., Moorman, A.V., Hall, A.G., Irving, J.A., Harrison, C.J. & Strefford, J.C. (2008) The complex genomic profile of ETV6-RUNX1 positive acute lymphoblastic leukemia highlights a recurrent deletion of TBL1XR1. *Genes, Chromosomes and Cancer*, **47**, 1118–1125.
- Pique-Regi, R., Caceres, A. & Gonzalez, J.R. (2010) R-Gada: a fast and flexible pipeline for copy number analysis in association studies. *BMC Bioinformatics*, **11**, 380.
- Schmiegelow, K., Forestier, E., Hellebostad, M., Heyman, M., Kristinsson, J., Soderhall, S. & Taskinen, M. (2010) Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. *Leukemia*, **24**, 345–354.
- Shurtleff, S.A., Buijs, A., Behm, F.G., Rubnitz, J.E., Raimondi, S.C., Hancock, M.L., Chan, G.C., Pui, C.H., Grosveld, G. & Downing, J.R. (1995) TEL/AML1 fusion resulting from a cryptic t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis. *Leukemia*, **9**, 1985–1989.
- Szabo, A. & Boucher, K. (2002) Estimating an oncogenetic tree when false negatives and positives are present. *Mathematical Biosciences*, **176**, 219–236.
- Wiemels, J.L., Cazzaniga, G., Daniotti, M., Eden, O.B., Addison, G.M., Masera, G., Saha, V., Biondi, A. & Greaves, M.F. (1999) Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet*, **354**, 1499–1503.
- Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X. & Li, T. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research*, **37**, D105–D110.