CD371 cell surface expression: a unique feature of *DUX4*-rearranged acute lymphoblastic leukemia

Dagmar Schinnerl,[1,] Ester Mejstrikova,[2,3,4,*] Angela Schumich,[1,*] Marketa Zaliova,[2,3,4,*] Klaus Fortschegger,[1] Karin Nebral,[1] Andishe Attarbaschi,[5] Karel Fiser,[2,3] Maximilian O. Kauer,[1] Niko Popitsch,[1] Sabrina Haslinger,[1] Andrea Inthal,[1] Barbara Buldini,[6] Giuseppe Basso,[6] Jean-Pierre Bourquin,[7] Giuseppe Gaipa,[8] Monika Brüggemann,[9] Tamar Feuerstein,[10] Margarita Maurer-Granofszky,[1] Renate Panzer-Grümayer,[1] Jan Trka,[2,3,4] Georg Mann,[5] Oskar A. Haas,[1] Ondrej Hrusak,[2,3,4] Michael N. Dworzak[1,5,#] and Sabine Strehl[1,#]*

**DS, EM, AS and MZ contributed equally as first authors. #MND and SS contributed equally as senior authors.*

[1]CCRI, Children's Cancer Research Institute, St. Anna Kinderkrebsforschung, Vienna, Austria; [2]CLIP – Childhood Leukemia Investigation Prague, Prague, Czech Republic; [3]Department of Pediatric Hematology and Oncology, Second Faculty of Medicine, Charles University, Prague, Czech Republic; [4]University Hospital Motol, Prague, Czech Republic; [5]Department of Pediatric Hematology and Oncology, St. Anna Children's Hospital; Medical University of Vienna, Vienna, Austria; [6]Department of Woman's and Child's Health, Laboratory of Pediatric Oncohematology, University of Padova, Padova, Italy; [7]Department of Oncology and Children's Research Center, University Children's Hospital Zürich, Zürich, Switzerland; [8]Department of Pediatrics, University of Milano-Bicocca, Fondazione Tettamanti – Centro Ricerca M. Tettamanti, Monza, Italy; [9]Department of Hematology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany and [10]Pediatric Hematology Oncology, Schneider Children's Medical Center, Petah-Tikva, Israel

Correspondence: SABINE STREHL/MICHAEL N. DWORZAK.
sabine.strehl@ccri.at/michael.dworzak@stanna.at
doi:10.3324/haematol.2018.214353

SUPPLEMENTARY INFORMATION

Schinnerl *et. al*.
CD371 cell surface expression:
A unique feature of DUX4-rearranged acute lymphoblastic leukemia

Supplementary Methods

Patients

All patient analyzed in this study were enrolled in the ALL-BFM 2000 (NCT00430118), 2009 (NCT01117441), Interfant-06 (NCT00550992) or EsPhALL 2009 (NCT00287105) clinical trials after obtaining informed consent from the patients, their parents, or legal guardians. The use of human material was approved by the institutional review boards in accordance with the Declaration of Helsinki.

RNA-seq

All samples subjected to RNA-seq contained ≥70% blast cells by morphology or blasts were enriched by a Ficoll-Paque gradient centrifugation and/or flow sorting.

RNA-seq cohort 1

Library preparation and RNA-seq were conducted at the Vienna BioCenter Core Facilities (VBCF; www.viennabiocenter.org/facilities) Next Generation Sequencing (NGS) Unit. Stranded RNA-seq libraries were prepared from 200-500 ng total RNA (RIN value >7) on a Hamilton STAR robot with NGS configuration using the NEB Next Ultra II Kit including poly(A)+ enrichment using oligo(dT)-beads. Paired-end (2x125 bp) RNA-seq was performed on an Illumina HiSeq2500 always multiplexing 7 libraries per lane. Data quality was checked by FastQC.

RNA-seq cohort 2

Sequencing libraries were prepared from total RNA using Agilent SureSelect mRNA Strand kits according to the manufacturer's instructions (Agilent, Technologies, USA). High throughput sequencing was performed on HiSeq2500 (2x50 or 2x100 bp) or NextSeq500 (2x75 bp) using TruSeq Rapid SBS and PE Cluster kits and High Output Kit, respectively, according to the manufacturer's instructions (Illumina, USA). Data quality was checked by FastQC and Qualimap.

Bioinformatics analysis

Alignment and differential expression analysis

Reads were aligned to human genome GRCh38 without alt loci (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.g

z) using the genecode annotation v28 (ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_28/gencode.v28.annotation.gtf.gz) and STAR-2.6.0c[1] run with the following parameter settings: --outSAMtype BAM Unsorted --outSAMunmapped Within --outFilterMultimapNmax 1 --outFilterMismatchNmax 3 --outFilterMismatchNoverLmax 0.3 --alignIntronMax 500000 --alignMatesGapMax 500000 --chimSegmentMin 10 --chimJunctionOverhangMin 10 --chimScoreMin 1 --chimScoreDropMax 30 --chimScoreJunctionNonGTAG 0 --chimScoreSeparation 1 --alignSJstitchMismatchNmax 5 -1 5 5 --chimSegmentReadGapMax 3 --chimMainSegmentMultNmax 10 --chimOutType SeparateSAMold.

Further analysis was performed in R (version 3.4.3) statistical environment using Bioconductor packages.[2] Count statistics for Refseq genes were obtained by the "featureCounts" function (package Rsubread 1.28.1)[3] and differential expression analysis was performed by DESeq2 (1.18.1).[4] Normalized counts (outlier replacement=true) were log2 transformed for data visualizations with ggplot2 (2.2.1). For clustering analysis with pheatmap[5] counts were transformed using the wrapper for the varianceStabilizingTransformation (vst) from the DESeq2 package. Data were batch corrected for different read length using the removeBatchEffect function from the limma package.[6] DUX4-specific gene sets used for clustering were taken from Yeoh *et al*.[7] (ChiSquare_NOVEL_above), Harvey *et al*.[8] (ROSE cluster 6) and Zhang *et al*.[9] (top up- and downregulated genes, supplementary table S3a_FPKM_SAM); gene names were updated to the current annotation version.

To investigate the expression level of *DUX4*, we mapped reads to the *DUX4* cDNA (NM_001293798.2) using STAR-2.6.0c (default parameters). The number of uniquely mapping reads per million assigned reads determined by featureCounts was calculated.

Detection of *DUX4* fusion transcripts

Fusion genes were detected with FusionCatcher (version 1.00),[10] Arriba (https://github.com/suhrig/arriba/) and STAR-Fusion (version 1.4.0)[11] using default settings and human genome GRCh38. Detected fusion genes were verified manually in IGV.[12,13] Cases with *DUX4* expression but no detectable fusion gene employing the three tools, were further inspected manually for the presence of fusion gene supporting reads with IGV.

Detection of copy number alterations

ERG and *IKZF1* copy number alterations were assessed using CytoScan HD arrays and the Chromosome Analysis Suite version 3.1 with NetAffx Annotation Version 33 (ThermoFisher Scientific) and/or genomic multiplex PCR[14] and/or multiplex ligation-dependent probe amplification (MLPA) using the SALSA MLPA P335 ALL-IKZF1 kit (MRC Holland, Amsterdam, The Netherlands).[15]

Immunophenotyping

Immunophenotypic studies were performed on erythrocyte-lysed diagnostic whole bone marrow samples. In a multi-tube 7-color panel approach we used lymphoid lineage markers together with APC-H7 or Pacific Orange conjugated mouse anti-human CD45 (clone 2D1, BD Pharmingen or clone HI30, Exbio) as backbone together with an extensive list of other markers of interest. CD371 expression was determined using the mouse-anti-human CD371 clone 50C1 in PE- or APC- conjugated format (BD Biosciences or BioLegend). The live-cell-permeant nucleated-cell dye (Syto41) was used to clear non-nucleated events (erythrocytes, debris) from analysis. A minimum of 30.000 nucleated cells were acquired per tube and doublets with increased FSC-Width were excluded. Blasts were identified by CD45/SSC gating plus lineage-marker gating as appropriate. Expression levels of CD371 on blasts were rated as being negative, weak or strong positive according to the AIEOP-BFM consensus guidelines 2016.[17] Acquisition was done on a Becton Dickinson™ LSR II (3 lasers) cytometer and data were analyzed using FACSDiva™ software (BD Biosciences).

CD371 antigen expression analysis

CD371 antigen expression was analyzed prospectively in n=258 consecutive B-ALL patients (Austrian BFM cohort; supplemental Table 3). Immunophenotyping of *DUX4*+ cases that were not included in this cohort and of selected cases from earlier periods because of their switching behavior to the monocytic lineage (swALL)[16,17] was retrospectively performed on biobanked material as available.

Definition of antigen expression of CD371 was essentially done according to the AIEOP-BFM consensus guidelines.[17] A threshold of ≥10% of gated cells expressing the respective antigen is used to consider an antigen as “positive” (“negative”: <10%). Expression is referred to as “weak”, if ≥10% but <50% of a cell population expresses the respective antigen (i.e. the majority of the cell population overlaps with the negative control). This applies irrespective of whether clearly two cell populations are present (subclonal pattern with partial expression), with only a minor cell population being positive (<50% of blasts), or whether there is a shift of the whole cell population with an extensive overlap of the distribution curve with the negative population. Expression is referred to as “strong”, if the positive cell population is ≥50%.

Supplementary Tables (Excel files)

Supplementary Table S1.

Cohort 1 subjected to RNA-seq.

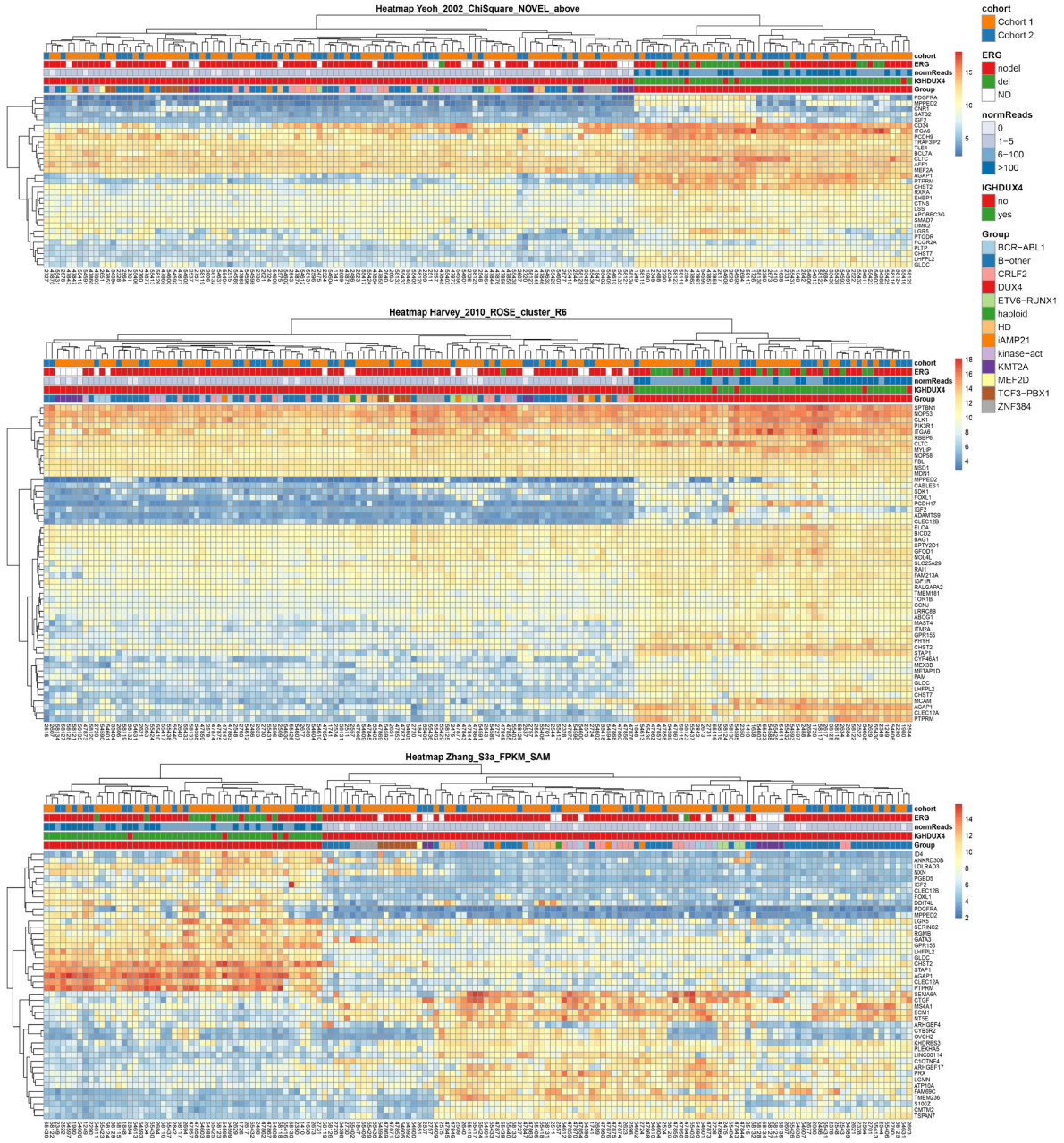
Supplementary Table S2.

Cohort 2 subjected to RNA-seq.

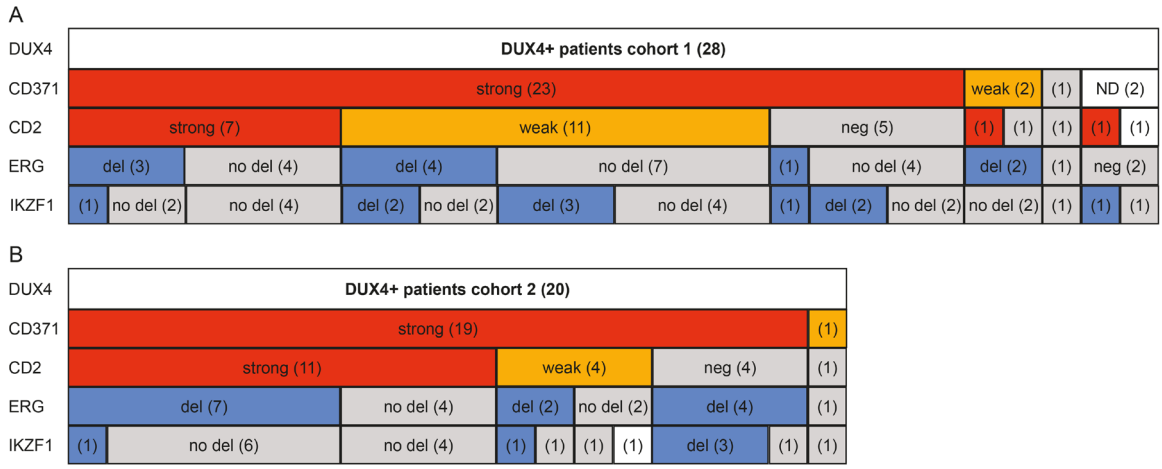
Supplementary Table S3.

CD371 expression and genetic subtypes of n=258 consecutive patients enrolled in Austrian ALL-BFM clinical trials.

Supplementary Figures



Supplementary Figure S1. Heatmaps DUX4-specific gene sets. Published gene sets[7-9] were used for clustering of the variance stabilization transformed counts by pheatmap; *ERG* nodel, no *ERG* deletion, *ERG* del, *ERG* deletion (detected by SNP-array and/or RT-PCR); ND, not determined; normReads, number of reads that mapped to *DUX4* cDNA (NM_001293798.2) per million total mapped reads assigned to RefSeq entries (RPM). Abbreviations: B-other, cases lacking any sentinel alteration; CRLF2, cases with *P2YR8-CRLF2* or *IGH-CRLF2* rearrangement; *DUX4*, cases with expression of *DUX4*; haploid, masked near haploid; HD, high hyperdiploid; iAMP21, intrachromosomal amplification of chromosome 21; kinase-act, cases harboring a kinase activation fusion gene; KMT2A, *KMT2A* fusion gene; MEF2D, *MEF2D* fusion gene; ZNF384, *ZNF384* fusion gene.



Supplementary Figure S2. Mosaic plot of *DUX4*+ samples showing the relationship with CD371 and CD2 antigen expression as well as *ERG* and *IKZF1* deletions. (**A**) **Cohort 1**. Deletions of *ERG* and *IKZF1* were determined by SNP array analysis and/or MLPA, respectively. **(B**) **Cohort 2**. Deletions of *ERG* and *IKZF1* were determined by genomic PCR and SNP array analysis, respectively. Dark orange, strong antigen expression; orange, weak antigen expression; grey, negative (no antigen expression or no deletion); blue, deletion; white, not determined; the number of patients in the respective category is indicated in parentheses; del, deletion; no del, no deletion; neg, negative; ND, not determined.

Supplementary References

1. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

2. Huber W, Carey VJ, Gentleman R, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. 2015;12(2):115-121.

3. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013;41(10):e108.

4. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

5. Kolde R. pheatmap: Pretty Heatmaps. R package version 1.0.8. https://CRAN.R-project.org/package=pheatmap 2015.

6. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

7. Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell. 2002;1(2):133-143.

8. Harvey RC, Mullighan CG, Wang X, et al. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. Blood. 2010;116(23):4874-4884.

9. Zhang J, McCastlain K, Yoshihara H, et al. Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. Nat Genet. 2016;48(12):1481-1489.

10. Nicorici D, Satalan M, Edgren H, et al. FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data. bioRxiv. 2014:doi: 10.1101/011650.

11. Haas B, Dobin A, Stransky N, et al. STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. bioRxiv. 2017; doi: 10.1101/120295.

12. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-26.

13. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-192.

14. Zaliova M, Zimmermannova O, Dorge P, et al. ERG deletion is associated with CD2 and attenuates the negative impact of IKZF1 deletion in childhood acute lymphoblastic leukemia. Leukemia. 2014;28(1):182-185.

15. Schwab CJ, Jones LR, Morrison H, et al. Evaluation of multiplex ligation-dependent probe amplification as a method for the detection of copy number abnormalities in B-cell precursor acute lymphoblastic leukemia. Genes Chromosomes Cancer. 2010;49(12):1104-1113.

16. Slamova L, Starkova J, Fronkova E, et al. CD2-positive B-cell precursor acute lymphoblastic leukemia with an early switch to the monocytic lineage. Leukemia. 2014;28(3):609-620.

17. Dworzak MN, Buldini B, Gaipa G, et al. AIEOP-BFM consensus guidelines 2016 for flow cytometric immunophenotyping of pediatric acute lymphoblastic leukemia. Cytometry B Clin Cytom. 2018;94(1):82-93.