Contents lists available at ScienceDirect

Genomics Data

Data in Brief

Identification of TEL-AML1 (ETV6-RUNX1) associated DNA and its impact on mRNA and protein output using ChIP, mRNA expression arrays and SILAC

Yvonne Linka [a,1], Sebastian Ginzel [a,b,1], Arndt Borkhardt [a,*], Pablo Landgraf [a]

[a] Heinrich-Heine-University Düsseldorf, Medical Faculty, Clinic for Pediatric Oncology, Hematology and Clinical Immunology, Duesseldorf, Germany
[b] Department of Computer Science, Bonn-Rhine-Sieg University of Applied Sciences, St. Augustin, Germany

ARTICLE INFO

ABSTRACT

The contribution of the most common reciprocal translocation in childhood B-cell precursor leukemia t(12;21)(p13;q22) to leukemia development is still under debate. Direct as well as secondary indirect effects of the TEL-AML1 fusion protein are commonly recorded by using cell lines and patient samples, often bearing the TEL-AML1 fusion protein for decades. To identify direct targets of the fusion protein a short-term induction of TEL-AML1 is needed. We here describe in detail the experimental procedure, quality controls and contents of the ChIP, mRNA expression and SILAC datasets associated with the study published by Linka and colleagues in the Blood Cancer Journal [1] utilizing a short term induction of TEL-AML1 in an inducible precursor B-cell line model.

Specifications

Organism/cell line/tissue	*Mus musculus*, BA/F3 cell line; *Homo sapiens*, bone marrow, peripheral blood, NALM6
Array type	mRNA expression arrays: GeneChip Mouse Genome 430 2.0 Array; Agilent Whole Human Genome Oligo Microarrays 8 × 60K. ChIP promoter array: NimbleGen 385 K RefSeq mouse promoter arrays; NimbleGen Human ChIP-chip 3 × 720K RefSeq Promoter Arrays
Data format	Raw data: mRNA: CEL files, ChIP promoter arrays: TXT Normalized data: mRNA: TXT; ChIP promoter arrays: TXT; SILAC: TXT
Experimental factors	Mouse and human cell lines: TEL-AML1 (ETV6-RUNX1) positive vs. negative Human patients: TEL-AML1 + vs. CD19+ cells
Experimental features	Induction of stable TEL-AML1 BA/F3 cell lines, expression profiling by array, genome binding/occupancy profiling by ChIP and genome tiling array, protein profiling by SILAC
Consent	All patients gave their written informed consent before study entry

Direct link to deposited data

Deposited data can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50736.

Experimental design, materials and methods

Establishment of TEL-AML1 inducible cell line system

The inducible TEL-AML1 cell line system has been previously described [2] and has been a kind gift of Anthony Ford. This system is based on the GeneSwitch™ System (Invitrogen), where the TEL-AML1 encoding cDNA is transcribed after induction with mifepristone (BA/F3^{TA+} cells). The control cells (BA/F3^{TA-}) contained only the pSwitch vector. Induction conditions were optimized for this study since it was crucial to obtain a sufficient number of viable cells just after one cell doubling. Limiting the cell doublings was necessary to minimize the occurrence of secondary effects of TEL-AML1 overexpression regularly seen in cells and patient samples constitutively expressing the TEL-AML1 fusion protein. BA/F3^{TA+} cells bearing the inducible TEL-AML1 fusion gene were induced with 30 to 40 pM concentration of mifepristone for 72 h. The expression of the fusion transcript was tested by FACS analysis using the V5-tag FITC-conjugated antibody (Abcam, Cambridge, UK) while cell viability was assessed by trypan blue dye exclusion test of cell viability [3] (Fig. 1A). 91.7% of viable cells were positive for the fusion protein upon induction with 32.5 pM mifepristone, with
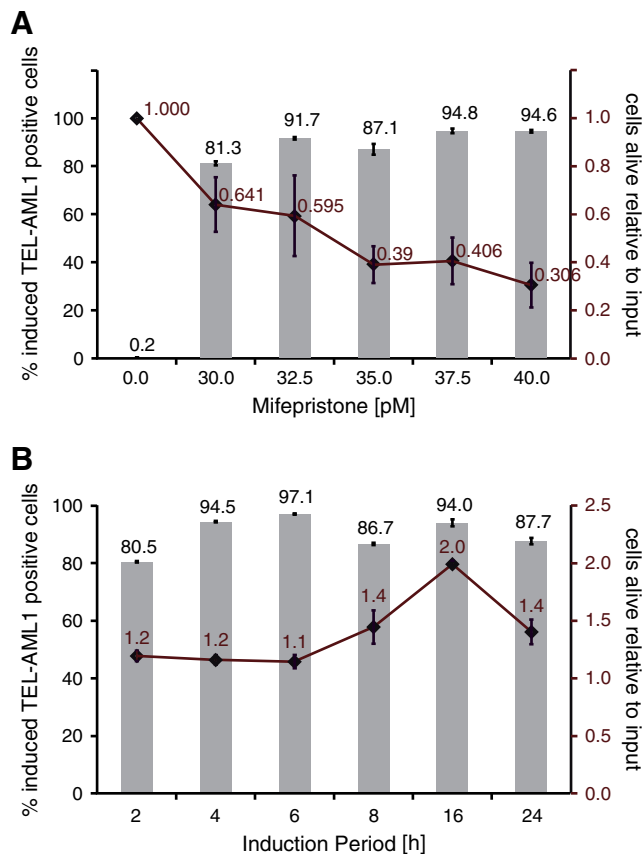
Fig. 1. (A) Evaluation of optimal mifepristone concentration for induction of TEL-AML1 expression was performed by treatment of 1.25 million cells with increasing amounts of mifepristone (0, 30.0, 32.5, 35.0, 37.5 and 40.0 pM) for 72 h (n = 3; SEM: standard error of the mean). (B) Optimization of mifepristone induction time was carried out by induction of 0.75 million cells each with 32.5 pM mifepristone. Cells were harvested after 2, 4, 6, 8, 16 and 24 h (n = 3; SEM: standard error of the mean). TEL-AML1 positive cells were analyzed by FACS using an anti-V5 FITC-conjugated antibody (bars; left y-axis). Living cells were counted in a Neubauer chamber by simultaneously staining dead cells with trypan blue and are shown in relation to input cell quantity (diamonds; right y-axis).

59.5% of cells surviving compared to non-induced cells. Within the first 16 h one cell doubling was reached while maintaining the rate of TEL-AML1 induced cells of 94% (± 1.2%, ± 1 SEM) (Fig. 1B). Therefore BA/F3 cells with the inducible TEL-AML1 fusion gene were treated with 32.5 pM mifepristone for 16 h and compared to empty vector control cells treated the same.

Cells were grown in RPMI 1640 (Gibco®, Camarillo, CA) supplemented with 2 mM L-glutamine (Gibco®), 10% fetal bovine serum (PAA, Pasching, Austria), 50 μg/ml Gentamicin (Gibco®) and 10 ng/ml IL3 (Gibco®). Moreover, BA/F3^{TA-} cells were cultured with 200 μg/ml Hygromycin B (Invitrogen). Growth medium for BA/F3^{TA+} cells additionally contained 50 μg/ml Zeocin™ (Invitrogen).

Two replicates of each, TEL-AML1 induced cells and treated empty vector control cells, were performed and analyzed by ChIP, mRNA expression microarrays and SILAC. Induction was always controlled by FACS analysis and an induction rate of at least 90% was required for all experiments.

Other cell line models used

The human precursor B-cell line NALM-6 (DSMZ ACC 128) was stably transfected with a TEL-AML1 fusion gene. TEL-AML1 cDNA was generated with overlap extension PCR from TEL (primers: 5′-GGCGCTCGCG AATGTCTGAGACTCCTGCTCAG, 5′-GGATTCATTCCAAGTATGCATTCTGC

TATTCTCCCAATGGGCATGG) and AML1 (primers: 5′-CCATGCCCATTG GGAGAATAGCAGAATGCATACTTGGAATGAATCC, 5′-CCGCGACTAGTTCA GTAGGGCCTCCACACGGCCTC) [4], cloned into the expression vector pMC3 [5] and transfected into NALM-6 cells using DMRIE-C (Invitrogen, Darmstadt, Germany). Stable cell clones were selected using 400 μg/ml Hygromycin for 3 weeks and cultured under selective pressure according to the provider. Independent triplicates were investigated by ChIP.

Patient and controls

Four samples obtained from bone marrow of pediatric patients with TEL-AML1 positive precursor B-cell acute leukemia were collected at diagnosis from bone marrow after informed consent. CD19 + cells of two healthy donors were sorted from peripheral blood using standard immunomagnetic cell sorting (MACS, Qiagen) and used for comparison of patient gene expression profiles.

Furthermore results were compared to a previously published dataset of 132 primary specimen including 20 patients with TEL-AML1 [6]. Pre-processed data comparing TEL-AML1 positive and -negative patient samples was used as described by Fuka et al. [7].

Chromatin immunoprecipitation on microarray (ChIP-on-chip) data

Chromatin immunoprecipitations were carried out essentially as described elsewhere [8]. A TEL-AML1 antibody was derived and characterized for this study [1]. Mouse endogenous TEL and AML1 was precipitated with commercially available antibodies directed against TEL (N-19; Santa Cruz Biotechnology, Santa Cruz, CA, USA) and AML (Ab-1; Calbiochem/Merck, Darmstadt, Germany), respectively. Magnetic protein G beads (Dynal, Invitrogen, Darmstadt, Germany), pre-blocked with pre-immunization serum over night at 4 °C, were used since protein A/G agarose beads (Santa Cruz Biotechnology, Santa Cruz, CA) exhibited background binding of proteins to the beads in the induced cells (Fig. 2). Antibody binding to the beads was preformed with 5 μg antibody for 6 h at 4 °C. Washed beads with antibody were then incubated at 4 °C over night with DNA of 1×10^7 cross-linked cells, previously sheared to ~500 bp length with an ultrasound disintegrator UP 100H with sonotrode MS7 (Dr. Hielscher GmbH, Teltow, Germany). TEL-AML1, TEL and AML1 bound DNA fragments were amplified with GenomePlex® Complete Whole Genome Amplification Kit (Sigma-Aldrich, Munich, Germany) and labeled along with its respective input material for each ChIP. Quantitative PCR amplification of Granzyme B, a known target of TEL-AML1, was used as quality control for each experiment and an enrichment of at least 5-fold over input was required for further processing.

Labeled DNA fragments were co-hybridized to 385K RefSeq mouse promoter arrays (NimbleGen) by Source BioScience imaGenes (Berlin, Germany). Log$_2$ ratios of signals obtained from immunoprecipitated and input DNA were calculated and scaled using Tukey biweight scaling. Microarray and sample annotation data was deposited in GEO under the accession numbers GSE50730. Only genes identified with a peak false discovery rate ≤0.05 in either replicate were retained for further analysis.

Murine TEL and AML1 immunoprecipitations were performed in triplicates (GEO accession numbers GSE50731 and GSE50732, respectively). TEL-AML1 ChIP was performed in triplicates from stably transfected NALM6 cells (GEO accession number GSE50733). Here only genes identified with a peak false discovery rate ≤0.2 in at least two out of three replicate experiments were retained for further analysis.

Gene-expression data

Total RNA from patients was isolated following standard TRIZOL procedure (Invitrogen), whereas total RNA from cell lines was isolated with RNeasy mini columns (Qiagen, Hilden, Germany). RNA integrity
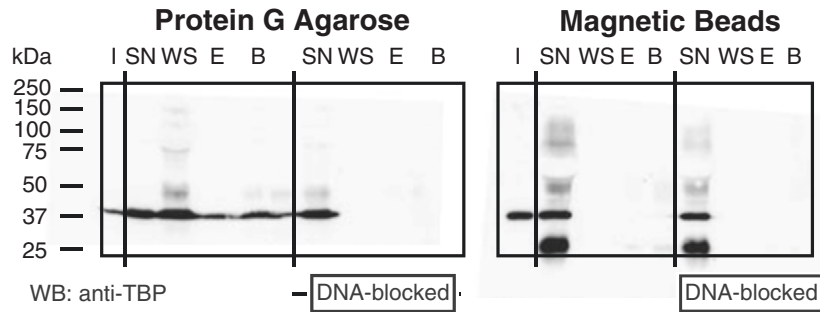
Protein G Agarose   **Magnetic Beads**

Fig. 2. SDS-page and Western blot analysis of ChIP experiments performed without antibody for specificity testing. Immunoprecipitation procedure was carried out with Protein A/G agarose (left) or magnetic beads (right). All beads were pre-cleared with pre-immunization serum. Agarose beads showed background binding of TBP to the beads, which was blocked when using salmon sperm DNA (as indicated). Abbreviations: 5% of input (I), supernatant (SN), wash (WS), and eluate (E). After elution, beads were boiled with SDS-sample buffer to elute all protein (B).

and quantity was assessed with RNA 6000 nano chip on a 2100 bioanalyzer (Agilent Technologies). RNA from 5×10^6 cells was labeled using a GeneChip 3′ IVT Express Kit (Affymetrix, Santa Clara, CA, USA) and hybridized to a GeneChip Mouse Genome 430 2.0 Array (Affymetrix) following standard procedure. Background correction and normalization of the probe signals were performed using the GCRMA package (v.2.28.0, [9]) in R 2.15 [10] and Bioconductor 2.16 [11] environment. Data was submitted to GEO (accession number GSE50735). Differentially expressed genes between TEL-AML1 induced cells and control were calculated using the linear model for microarray data as implemented in the LIMMA R-package [12] with Benjamini–Hochberg correction for multiple testing. Fold changes of means of $\log_2$ expression were calculated. Genes with a P-value ≤ 0.05 and a fold change of at least 1.5-fold were retained.

10 ng of patient-derived RNA was linear T7-amplified and Cy3 labeled with the Low Input Quick Amp Labeling Kit (Agilent Technologies) following the manufacturer's protocol. Yields of cRNA and the dye-incorporation rate were measured with the ND-1000 Spectrophotometer (NanoDrop Technologies). Labeled RNA was hybridized to Agilent Whole Human Genome Oligo Microarrays 8 × 60K. The Agilent Feature Extraction Software (FES) was used to read out and process the microarray image files. Normalized data was submitted to GEO (accession number GSE507349). For determination of differential gene expression between patients and controls, ratios of FES derived output data files were generated using the Rosetta Resolver® gene expression data analysis system (Rosetta Biosoftware).

Stable isotope labeling by amino acids in cell culture (SILAC) and mass spectrometric data

Inducible TEL-AML1 BA/F3 cell lines were grown in L-arginine-$^{13}C_6^{15}N_4$ and L-lysine-$^{13}C_6^{15}N_2$ supplemented growth media ("heavy weight") while mifepristone-treated vector control cell lines were grown in L-arginine-$^{13}C_6$ and L-lysine-D4 supplemented growth media ("medium weight") for at least 5 doublings. Induction with mifepristone was performed as described above. Non-induced cells bearing the inducible construct were labeled with supplemented "low weight" medium containing L-arginine and L-lysine without label. Pellets corresponding to 7×10^6 cells were flash frozen in liquid nitrogen. Mass spectrometric analysis of trypsin-digested size-fractionated proteins was performed as described elsewhere [13,14]. Peptides and proteins were identified with Mascot (Matrix Science, London, UK) and quantified with MSQuant (http://msquant.sourceforge.net) as described previously [14]. MSQuant analyzed data is provided as supplement. To account for effects of mifepristone treatment we only compared the middle and heavy weight peaks. Normalized peak ratios of heavy weight to medium weight peaks either below 0.85 or above 1.15 were retained for further analysis.

Integrative data analysis

All analyses were performed using R 2.15 software [10] and Bioconductor 2.16 [11]. Mapping of mRNA probe identifiers was first performed to either mouse or human entrez gene IDs using Bioconductor annotation package (mouse4302.db and hgug4112a.db, respectively). Entrez gene IDs for corresponding peak regions of mouse ChIP experiments were created by NimbleScan map peak analysis. MGI IDs of SILAC peptide data were also mapped to entrez gene IDs using biomaRt 2.12 [15]. Subsequently mapping of entrez gene IDs to the most recent Ensemble gene ID at the time of analysis (Ensemble v66) was carried out with biomaRt 2.12 to ensure uniformity of data version for integrative analysis. For comparison of datasets obtained from the inducible mouse cell line model, only genes queried by both, promoter arrays and mRNA expression array platforms, were retained for further analysis. Comparisons were carried out only on significant genes as described above. TEL-AML1 direct targets were defined by the binding of the fusion protein to promoter sequences and the simultaneous regulation of mRNA of the same gene and are thus reflected by the overlap of mRNA expression and ChIP datasets (Supplementary Table 1 of [1]). The indirect effects of TEL-AML1 overexpression were defined by changes in mRNA expression and protein expression as defined by microarray and SILAC analyses (Supplementary Table 2 of [1]).

The number of canonical AML1 binding motif "TG[CT]GGT[CT]" in promoter regions was normalized to promoter region length of direct and indirect target-genes and compared to the background AML1
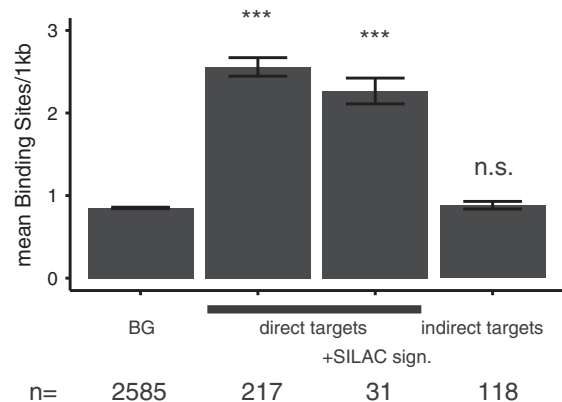

Fig. 3. AML1-binding sites in promoters. AML1 binding sites are enriched in all TEL-AML1 direct targets and in a subset of direct targets with significant changes in protein output, while indirect targets do not show enrichment over array background level. The mean number of target sites per gene normalized to promoter region ± 1 SEM is depicted. T-test statistics was performed; n.s., not significant; ***, p ≤ 0.001.

binding site count of the array (Fig. 3). P-values were calculated by comparison of these values to the distribution of the background. The background distribution of AML1 binding motif occurrence was simulated by taking the same number of random probes of the array platform as were identified in the respective TEL-AML1 ChIP experiments with 1000 repetitions and again calculating the occurrence of the AML1 binding motif as described above.

Discussion

We describe here a unique dataset obtained from a mouse early B-cell model with induced TEL-AML1 (ETV6-RUNX1) fusion protein in comparison to a treated control cell line. Using ChIP, mRNA expression arrays and SILAC we queried the occupancy of the fusion protein simultaneously to its impact on mRNA and protein output. Furthermore we generated occupancy profiles of the endogenous TEL and AML1 as well as mRNA profiles of a small patient cohort with TEL-AML1 positive early B-cell leukemia in comparison to CD19-sorted B-cells. This dataset has been recently used to identify early direct and indirect targets of TEL-AML1 [1] and reported persistent effects of directly regulated TEL-AML1 genes in our and an independent patient cohort [6].

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgment

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2014.05.010.

References

[1] Y. Linka, S. Ginzel, M. Kruger, A. Novosel, M. Gombert, et al., The impact of TEL-AML1 (ETV6-RUNX1) expression in precursor B cells and implications for leukaemia using three different genome-wide screening methods. Blood Cancer J. 3 (2013) e151.

[2] A.M. Ford, C. Palmi, C. Bueno, D. Hong, P. Cardus, et al., The TEL-AML1 leukemia fusion gene dysregulates the TGF-beta pathway in early B lineage progenitor cells. J. Clin. Invest. 119 (2009) 826–836.

[3] W. Strober, Trypan blue exclusion test of cell viability. Curr. Protoc. Immunol. 21 (2001) A.3B.1-A.3B.2.

[4] S.N. Ho, H.D. Hunt, R.M. Horton, J.K. Pullen, L.R. Pease, Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77 (1989) 51–59.

[5] R.M. Linka, S.L. Risse, K. Bienemann, M. Werner, Y. Linka, et al., Loss-of-function mutations within the IL-2 inducible kinase ITK in patients with EBV-associated lymphoproliferative diseases. Leukemia 26 (2012) 963–971.

[6] M.E. Ross, X. Zhou, G. Song, S.A. Shurtleff, K. Girtman, et al., Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. Blood 102 (2003) 2951–2959.

[7] G. Fuka, M. Kauer, R. Kofler, O.A. Haas, R. Panzer-Grumayer, The leukemia-specific fusion gene ETV6/RUNX1 perturbs distinct key biological functions primarily by gene repression. PLoS One 6 (2011) e26348.

[8] T.I. Lee, S.E. Johnstone, R.A. Young, Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. 1 (2006) 729–748.

[9] Z. Wu, R.A. Irizarry, Stochastic models inspired by hybridization theory for short oligonucleotide arrays. J. Comput. Biol. 12 (2005) 882–893.

[10] R. Ihaka, R. Gentleman, R: a language for data analysis and graphics. J. Comput. Graph. Stat. 5 (1996) 299–314.

[11] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, et al., Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5 (2004) R80.

[12] G.K. Smyth, Limma: linear models for microarray data. in: R. Gentleman, S. Dudoit, R.A. Irizarry, W. Huber (Eds.), Bioinformatics and Computational Biology Solutions using R and Bioconductor, Springer, New York, 2005, pp. 397–420.

[13] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. 68 (1996) 850–858.

[14] S.E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, et al., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics 1 (2002) 376–386.

[15] S. Durinck, P.T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4 (2009) 1184–1191.