Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL

Giusy Della Gatta1, Teresa Palomero1,2, Arianne Perez-Garcia1, Alberto Ambesi-Impiombato1, Mukesh Bansal3, Zachary W Carpenter1, Kim De Keersmaecker4,5, Xavier Sole6,7, Luyao Xu1, Elisabeta Paitetta8,9, Janis Racevskis8,9, Peter H Wiernik8,9, Jacob M Rowe10, Jules P Meijerink11, Andrea Califano1,3 & Adolfo A Ferrando1,2,12

The TLX1 and TLX3 transcription factor oncogenes have a key role in the pathogenesis of T cell acute lymphoblastic leukemia (T-ALL)1,2. Here we used reverse engineering of global transcriptional networks to decipher the oncogenic regulatory circuit controlled by TLX1 and TLX3. This systems biology analysis defined T cell leukemia homeobox 1 (TLX1) and TLX3 as master regulators of an oncogenic transcriptional circuit governing T-ALL. Notably, a network structure analysis of this hierarchical network identified RUNX1 as a key mediator of the T-ALL induced by TLX1 and TLX3 and predicted a tumor-suppressor role for RUNX1 in T cell transformation. Consistent with these results, we identified recurrent somatic loss-of-function mutations in RUNX1 in human T-ALL. Overall, these results place TLX1 and TLX3 at the top of an oncogenic transcriptional network controlling leukemia development, show the power of network analyses to identify key elements in the regulatory circuits governing human cancer and identify RUNX1 as a tumor-suppressor gene in T-ALL.

TLX1 and TLX3 encode highly related homeobox transcription factor oncogenes frequently activated by chromosomal translocations in T-ALL3-5. To interrogate the transcriptional programs associated with aberrant expression of TLX1 and TLX3, we analyzed gene expression data from 82 human T-ALL6. This analysis revealed that tumors with expression of TLX1 and TLX3 share a common expression signature that includes 319 upregulated and 450 downregulated gene transcripts compared with non–TLX1- and non–TLX3-expressing leukemias, respectively (greater than twofold change, P < 0.005) (Fig. 1a and Supplementary Table 1). Moreover, non-negative matrix factorization and a principal component analysis showed that leukemias with altered TLX1 and TLX3 are highly related in their global gene expression programs and cluster together in our analyses separate from the rest of the T-ALL samples in our series (Supplementary Fig. 1). These results support a broadly overlapping role of TLX1 and TLX3 in the induction of T-ALL, however, leukemias induced by TLX1 and TLX3 expression have been associated with different prognoses in some series1,7, suggesting key biological differences between these two groups. Consistently, comparative marker analyses identified numerous differentially expressed genes in TLX1 compared to TLX3 T-ALLs (Supplementary Fig. 2).

Next we analyzed TLX1 chromatin immunoprecipitation (ChIP)-chip data in ALL-SIL cells, a human T-ALL cell line with high expression of TLX1 as result of the t(10;14)(q24;q11) translocation3 and performed ChIP-chip analyses for TLX3 in cells from HPB-ALL, a human cell line positive for the TLX3-activating t(5;14)(q35;q32) translocation5. These analyses identified 2,236 promoters bound by TLX1 and 3,148 promoters occupied by TLX3 using a significance cutoff of P < 10−4 (Supplementary Table 2). Notably, 75% of the direct targets of TLX1 were also bound by TLX3 (χ2 P < 0.001) (Fig. 1b). A gene set enrichment analysis (GSEA) also showed a highly significant enrichment of genes whose promoters were bound by TLX1 and TLX3 in the expression signature associated with TLX1- and TLX3-expressing leukemias (P < 0.001) (Fig. 1c and Supplementary Table 3). Most notably, genes bound by TLX1 and TLX3 were characteristically downregulated in this group (Fig. 1c), strongly suggesting that TLX1 and TLX3 primarily function as transcriptional repressors in the pathogenesis of T-ALL.

We then used the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe) reverse-engineering algorithm9,10 to generate a genome-wide T-ALL transcriptional network, or T-ALL interactome (T-ALLi), using gene expression data from 228 T-ALLs (Online Methods). This analysis yielded a T-ALLi that included 19,689 genes (nodes) connected through 471,824 interactions (edges) with a hierarchical structure (Supplementary Fig. 3). Notably, v-myc myelocytomatosis
viral oncogene homolog (MYC) target genes inferred in the T-ALLi were markedly enriched in genes that are direct targets of MYC determined by ChIP-chip (74 out of 252 genes, $\chi^2 P = 2.5 \times 10^{-3}$), supporting the soundness of this approach (Supplementary Fig. 4). An analysis of TLX1- and TLX3-connected genes in this setting identified 325 predicted TLX target genes (Fig. 2a). Notably, these included 70 highly significant ($P < 0.0001$) target genes of TLX1 and TLX3 as determined by ChIP-chip ($\chi^2 P = 0.02$) (Fig. 2b) and 117 genes that are differentially expressed ($P < 0.0001$) in TLX1- and TLX3-expressing T-ALLs ($\chi^2 P < 0.001$) (Fig. 2c).

Next, we defined the TLX subnetwork (TLXi) as the space in the T-ALLi that encompasses the 445 genes that are direct targets of TLX1 and TLX3 (ChIP-chip $P < 0.0001$) and are also differentially expressed in TLX1- and TLX3-expressing T-ALLs ($P < 0.0001$) and their most direct interconnections (Fig. 3a). The TLXi retains the topological features of the T-ALLi. Accordingly, 92% (411/445) of these genes were involved in at least one ARACNe-predicted interaction, but less than 2% (8/445) of them had 50 or more direct network-predicted interactions (Fig. 3b and Supplementary Table 4). Notably, and consistent with the role of TLX1 as a transcriptional repressor, TLXi gene transcripts were also characteristically downregulated as determined GSEA in a transgenic mouse model of TLX1-induced T-ALL11 (Supplementary Fig. 5). Moreover, GSEA of the expression signatures induced by shRNA knockdown of TLX1 in ALL-SIL cells and of TLX3 in HPB-ALL cells showed high enrichment of genes in the TLXi among the transcripts upregulated on inactivation of TLX1 and TLX3, respectively (Supplementary Figs. 6 and 7).

Based on these results, we proposed that the hierarchical regulatory structure of the TLXi could reflect, at least in part, the functional
hierarchy of the TLX1 and TLX3 target genes involved in T cell transformation. In this context, RUNX1, a crucial transcription factor in hematopoietic development\(^{12}\) that is frequently mutated in acute myeloid leukemias\(^{13-15}\), was the single most highly interconnected hub in the TLXi (Fig. 3c). A ChIP analysis of TLX1 and TLX3 confirmed the binding of these transcription factors to the RUNX1 promoter (Supplementary Fig. 8). In addition, RUNX1 was significantly more interconnected in the TLXi than in the T-ALLi as a whole (\(\chi^2 P = 2.14 \times 10^{-133}\)) and was one of the top TLXi genes downregulated in mouse TLX1- and TLX3-expressing T-ALLs (Supplementary Fig. 8). Consistently, master regulator analyses\(^{16,17}\) identified RUNX1 as one of the top most significant master regulators of the transcriptional program associated with human TLX1- and TLX3-induced leukemias (Supplementary Table 5). The model that emerges from this analysis is a regulatory loop in which downregulation of RUNX1 by TLX1 and TLX3 would subsequently affect the expression of numerous other TLX target genes (Supplementary Fig. 9). To test this possibility, we performed a ChIP-chip analysis of the direct targets of RUNX1 in HPB-ALL cells. In this analysis, we identified 308 high-confidence RUNX1 target genes (\(P < 0.0001\)) (Supplementary Table 6). Notably, and in concordance with our network analysis, 50% of the RUNX1-occupied promoters were also bound by TLX1 and TLX3 (\(\chi^2 P < 10^{-15}\)). In addition, GSEA of genes that are direct targets of RUNX1 showed high enrichment of RUNX1 targets among the top transcripts downregulated in T-ALL cells with high expression of TLX1 or TLX3 (\(P = 0.05\)) (Fig. 3d).

These results suggest that RUNX1 could mediate, at least in part, some of the oncogenic effects of TLX1 and TLX3 overexpression. Consistent with this hypothesis, retroviral expression of RUNX1 in TLX1-positive (ALL-SIL) and TLX3-positive (HPB-ALL) cells resulted in impaired cell growth (Supplementary Fig. 10), indicating a possible tumor-suppressor role for RUNX1 in T-ALL. A mutation analysis of RUNX1 in T-ALL revealed the presence of RUNX1 mutations in 33.3% (4/12) of T-ALL cell lines (Online Methods) and in 4.4% (5/114) of T-ALL primary samples (Fig. 4a and Supplementary Tables 7 and 8). Notably, all ALLs identified in kindreds with familial platelet disorder with associated myeloid malignancy MIM#601399, a leukemia predisposition syndrome caused by mutations in RUNX1, are T-ALLs\(^{18-20}\).

The RUNX1 mutations we found in T-ALL were heterozygous frameshift truncating mutations (3/9) and missense single-nucleotide changes (6/9) (Fig. 4a,b). Notably, a DNA sequence analysis of samples from two individuals with T-ALL with available material obtained at clinical remission showed the somatic origin of the RUNX1 mutations (Fig. 4b). In addition, five of the RUNX1 mutant alleles (L29S, H58N, H78Y, S114fs and G138fs) were previously described as oncogenic mutations in myeloid tumors\(^{21-25}\). Notably, all four samples with mutations in RUNX1 with available immunophenotype data showed a CD4 and CD8 double-negative immunophenotype, which is indicative of a very early arrest in T cell maturation (Supplementary Table 9). Mapping of the T-ALL RUNX1 mutations onto the structure of the RUNX1 runt domain (protein Data Bank ID 1H9D) showed a clustering of these amino acid substitutions in the DNA-recognition interface of RUNX1 (Fig. 4c). Most notably, the RUNX1 His78 residue, which resides within a highly structurally conserved 16.9-Å-diameter cavity that is frequently targeted by RUNX1 mutant alleles in acute myeloid leukemia (AML), is adjacent to the DNA-binding interface and is predicted to be disrupted in the H78Y RUNX1 mutant in T-ALL (Fig. 4c). Next we tested the functional role of the RUNX1 mutants predicted to be the most structurally disruptive in luciferase reporter assays. In these...
experiments, the RUNX1 H78Y, RUNX1 S114fs and RUNX1 G138fs mutations showed marked (80%) reductions in their capacity to activate a RUNX1-responsive colony-stimulating factor (CSF) promoter reporter construct compared to wild-type RUNX1 (Fig. 4d).

Next we analyzed the transcriptional programs and disease kinetics of leukemias occurring in Lck-TLX1 transgenic[11] Runx1 wild-type mice (Runx1+/+ ) and in Lck-TLX1 Runx1 heterozygous knockout mice (Runx1+/−). This analysis revealed that the Lck-TLX1 Runx1+/+ and Lck-TLX1 Runx1+/− tumors shared a common gene expression program compared with non-Lck-TLX1 mouse leukemias consisting of 215 commonly differentially expressed genes (greater than twofold change, P < 0.001). However, and consistent with the presence of 50% nonoverlapping target genes between RUNX1 and TLX1, loss of one copy of Runx1 partially changed the transcriptional signature of TLX1-induced leukemias, resulting in 540 differentially expressed transcripts between Lck-TLX1 Runx1+/+ and Lck-TLX1 Runx1+/− tumors (greater than twofold change, P < 0.001) (Supplementary Fig. 11). Notably, and despite these transcriptional differences, Lck-TLX1 transgenic Runx1 wild-type and Lck-TLX1 Runx1 haploinsufficient mice developed T-ALL with identical kinetics (Supplementary Fig. 12), suggesting that, in agreement with the prediction of our network analysis, the oncogenic effects of TLX1 are overlapping with the tumor-suppressor activity of RUNX1.

Overall, the integrative analyses presented here (Supplementary Fig. 13) show a large amount of functional overlap between TLX1 and TLX3 in T cell transformation and identify RUNX1 as a tumor-suppressor gene in T-ALL. Notably, this work highlights the power of network analysis to decipher the structure of complex oncogenic circuitries and identify crucial genes and pathways involved in the pathogenesis of human cancer. Moreover, reverse engineering of signaling and transcriptional networks controlling the phenotypes associated with distinct gene expression signatures, such as cell transformation, metastatic potential or drug resistance, could be exploited to identify new therapeutic targets.

Two reports published during the preparation of this manuscript described convergent results with ours showing the presence of somatic mutations in RUNX1 in T-ALL[26,27].

**METHODS**

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemedicine/.

*Note: Supplementary information is available on the Nature Medicine website.*

**ACKNOWLEDGMENTS**

This work was supported by the US National Institutes of Health (grants R01CA120196 and R01CA129382 to A.A.F.; and U24 CA114737 to E.P.), the New York Community Trust (A.A.F.), the Innovative Research Award by the Stand Up to Cancer Foundation (A.A.F.), the Eastern Cooperative Oncology Group tumor bank, the Leukemia and Lymphoma Society Scholar Award (A.A.F.), the Stichting Kinderen Kankevrij (KiKa; grant 2007-012) (I.P.M.) and the Dutch Cancer Society (KWF-EMCR 2006-3500) (I.P.M.). A.P.-G. is a postdoctoral researcher funded by the Rally Foundation. G.D.G. was supported by a Marie Curie International Outgoing fellowship. We thank S. Nimer (Memorial Sloan Kettering Cancer Center) for the pCDNA3 RUNX1 expression vector, D. Zhang (University of California, San Diego) for the pM-CSF-R-luc promoter reporter and the pCMV CBFB plasmids and J. Downing (St. Jude Children’s Research Hospital) for the Runx1 knockout mice.

**AUTHOR CONTRIBUTIONS**

G.D.G. performed expression and network analysis of human and mouse leukemias, identified RUNX1 mutations in T-ALL and wrote the manuscript; T.P. performed ChIP-chip analysis of TLX1 and TLX3 and the RNA interference experiments; A.P.-G. analyzed the tumor activity of RUNX1; A.A.-I. and M.B. performed network analysis; Z.W.C. performed structure modeling of RUNX1 mutant protein; K.D.K. performed mouse studies; X.S. analyzed ChIP-chip data; L.X. performed mouse studies; E.P., J.R., P.H.W. and J.M.R. provided clinical specimens and correlative data on adult T-ALL samples; I.P.M. contributed clinical specimens and performed microarray analysis; A.C. supervised research and A.A.F. designed the study, supervised research and wrote the manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturemedicine/. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Ferrando, A.A. *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* 1, 75–87 (2002).

2. Atlantis, I., Raetz, E. & Buonomics, S. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nat. Rev. Immunol.* 8, 380–390 (2008).
3. Hatano, M., Roberts, C.W., Minden, M., Crist, W.M. & Korsmeyer, S.J. Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia. Science 253, 79–82 (1991).
4. Kennedy, M.A. et al. HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24. Proc. Natl. Acad. Sci. USA 88, 8900–8904 (1991).
5. Bernard, O.A. et al. A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. Leukemia 15, 1495–1504 (2001).
6. Van Vlierberghe, P. et al. The recurrent SET-NUP214 fusion as a new HOXA13-PEBP2A fusion in T-cell acute lymphoblastic leukemia. Blood 111, 4668–4680 (2008).
7. Ferrando, A.A. et al. Prognostic importance of TLX1 (HOX11) oncogene expression in adults with T-cell acute lymphoblastic leukemia. Lancet 363, 535–536 (2004).
8. Su, X.Y. et al. Various types of rearrangements target TLX3 locus in T-cell acute lymphoblastic leukemia. Genes Chromosom. Cancer 41, 243–249 (2004).
9. Basso, K. et al. Reverse engineering of regulatory networks in human B cells. Nat. Genet. 37, 382–390 (2005).
10. Margolin, A.A. et al. ARA-CNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7 (suppl. 1), S7 (2006).
11. De Keersmaecker, K. et al. The TLX1 oncogene drives aneuploidy in T cell transformation. Nat. Med. 16, 1321–1327 (2010).
12. Speck, N.A. & Gilliland, D.G. Core-binding factors in haematopoiesis and leukaemia. Nat. Rev. Cancer 2, 502–513 (2002).
13. Song, W.J. et al. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukemia. Nat. Genet. 23, 166–175 (1999).
14. Osato, M., Yanagida, M., Shigesada, K. & Ito, Y. Point mutations of the RUNX1/AML1 gene in sporadic and familial myeloid leukemias. Int. J. Hematol. 74, 245–251 (2001).
15. Osato, M. Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia. Oncogene 23, 4284–4296 (2004).
16. Lefebvre, C. et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol. Syst. Biol. 6, 377 (2010).
17. Carro, M.S. et al. The transcriptional network for mesenchymal transformation of brain tumours. Nature 463, 318–325 (2010).
18. Preudhomme, C. et al. High frequency of RUNX1 biallelic alteration in acute myeloid leukemia secondary to familial platelet disorder. Blood 113, 5583–5587 (2009).
19. Owen, C.J. et al. Five new pedigrees with inherited RUNX1 mutations causing familial platelet disorder with propensity to myeloid malignancy. Blood 112, 4639–4645 (2008).
20. Nishimoto, N. et al. T cell acute lymphoblastic leukemia arising from familial platelet disorder. Int. J. Hematol. 92, 194–197 (2010).
21. Osato, M. et al. Biallelic and heterozygous point mutations in the runt domain of the AML1/PEBP2αB gene associated with myeloblastic leukemias. Blood 93, 1817–1824 (1999).
22. Rocquain, J. et al. Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNXI, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias. BMC Cancer 10, 401 (2010).
23. Langabeer, S.E., Gaie, R.E., Rollinson, S.J., Morgan, G.J. & Linch, D.C. Mutations of the AML1 gene in acute myeloid leukemia of FAB types M0 and M7. Genes Chromosom. Cancer 34, 24–32 (2002).
24. Auewarakul, C.U. et al. AML1 mutation and its coexistence with different transcription factor gene families in de novo acute myeloid leukemia (AML): redundancy or synergism. Haematologica 92, 861–862 (2007).
25. Christiansen, D.H., Andersen, M.K. & Pedersen-Bjergaard, J. Mutations of the HOX11 locus in T-cell acute lymphoblastic leukemia. Blood 94, 8900–8904 (1999).
26. Grossmann, V. et al. Prognostic relevance of RUNX1 mutations in T-cell acute lymphoblastic leukemia. Blood 110, 1474–1481 (2004).
27. Zhang, J. et al. The genetic basis of early T-cell precursor acute lymphoblastic leukemia. Nature 481, 157–163 (2012).
Online Methods

Clinical samples. Leukemic DNA and cryopreserved lymphoblast samples were provided by the collaborating institutions in the United States (the Eastern Cooperative Oncology Group and the Pediatric Oncology Group). All samples were collected under the supervision of local institutional review board committees. Informed consent was obtained from all subjects at entry into the trial, according to the declaration of Helsinki.

Master regulator analysis. A master regulator analysis was carried out as previously described19. Briefly, each set of transcription factor targets (regulons) was partitioned as ‘positive’ or ‘negative’ based on the correlation of the transcription factor and target. Positive and negative regulons were tested for enrichment in the TLX1 and TLX3 signature. Redundancy in the inferred master regulators that have a large number of common targets was corrected for by removing ‘shadowed transcription factors’, which are identified as those master regulators whose enrichment is significantly reduced when the common targets are disregarded.

ChIP and ChIP-chip analysis. A ChIP-chip analysis of the TLX3 and RUNX1 target genes was performed in the HPB-ALL cell line. Briefly, 1 × 10⁸ cells were used for ChIP using A-17 goat polyclonal (sc-23397) and H-55 rabbit polyclonal (sc-30185) antibodies to TLX3 (Santa Cruz Biotechnology) or two rabbit polyclonal antibodies to RUNX1 (Ab980 from Abcam and 43365 from Cell Signaling Technologies). ChIP-chip was performed following the standard protocols provided by Agilent Technologies using Agilent Human Proximal Promoter Microarrays (244,000 features per array), as previously described20. This platform analyzes ~17,000 of the best-defined human genes sourced from UCSC hg18 (NCBI Build 36.1, March 2006) and covers regions ranging from −5.5 kb upstream to +2.5 kb downstream of their transcriptional start sites. We scanned the arrays with an Agilent scanner and extracted the data using Feature Extraction 8 software. Genes that were direct targets of TLX3 and RUNX1 were identified using a ChIP-chip significance analysis, as previously described20. A ChIP-chip analysis of MYC and TLX1 in T-ALL has been previously reported11,20.

Relative real-time PCR quantification of the RUNX1 promoter sequences was normalized to the ACTB gene in chromatin immunoprecipitates performed with antibodies to TLX1 (C-18 rabbit polyclonal antibody (sc-880), Santa Cruz Biotechnology) and TLX3 (A-17 goat polyclonal antibody (sc-23397), Santa Cruz Biotechnology). The primer sequences used are listed in Supplementary Table 10.

Reverse engineering of the T-ALL transcriptional networks. To generate a T-ALL transcriptional network, we processed Human U133 Plus2.0 Affymetrix microarray gene expression data from a series of 228 T-ALL primary samples using the GC robust multi-array average normalization ARACNe algorithm, as previously described19, and named the resulting global T-ALL transcriptional network the T-ALLi. Given the high amount of overlap between TLX1- and TLX3-regulated direct target genes, and to avoid the elimination of the connections between genes showing high amounts of mutual information with both TLX1 and TLX3 by ARACNe during the data processing inequality step (which is aimed to filter out indirect connections), the expression of these two transcription factors was averaged as a single node by assigning the same gene label (TLX) to both the TLX1 and TLX3 probes.

In a separate analysis, we defined the genes experimentally identified by ChIP-chip as being direct targets of TLX1 and TLX3 (P < 0.0001) and the genes differentially expressed in TLX1- and TLX3-expressing tumors (differential expression P < 0.0001) as being the core of the oncogenic program that is controlled by TLX1 and TLX3 in T-ALL. We then defined the TLXii as the subspace within the T-ALLii containing all of these differentially expressed TLX1 and TLX3 direct target genes and their shortest path interconnections. The significance of the TLXii was tested for by performing in silico simulations of 10,000 random networks characterized by the same TLXi features (48 transcription factors and 1,655 connections). The significance of the TLXii compared to randomly generated networks was determined by calculating a nonparametric P value.

RUNX1 mutation analysis. All RUNX1 exon sequences were amplified from genomic DNA by PCR and analyzed by direct dideoxynucleotide sequencing. The PCR and sequencing primer sequences used are listed in Supplementary Table 11.

Structural depiction and analysis. Structural coverage of the RUNX1 protein was identified through use of the PSI-Blast and SKAN algorithms; viable structures were subsequently mapped to all RUNX1 isoforms and analyzed with the MarkUS web annotation server29. The Protein Data Base structures 1EAN, 1EAQ, 1H9D, 1H4O, 1HJB, 1HJC and 2J6W were structurally aligned along the RUNX1 runt domain–DNA interface, and the resulting composite structure was subsequently analyzed to assess conformational flexibilities30.

Potential effects for the RUNX1 mutants in T-ALL were investigated with SCREEN and VASP for cavity prediction and volumetric rendering, ConSurf for analysis of structural conservation, PredUS for protein-protein interface prediction and DelPhi to highlight potential alterations in electrostatic potential25. A probabilistic classification of mutations through physical and evolutionary comparative considerations was conducted through the use of the PolyPhen-2 batch servers and algorithms31. RUNX1 mutants in AML were extracted from the COSMIC database, filtered and mapped to runt domain structures30. All structural images were created using UCSF Chimera30.

Statistical analyses. Significant overlapping between different groups of genes was calculated using a χ² test.

Additional methods. Detailed methodology is described in the Supplementary Methods.