nature genetics

Triplication of a 21q22 region contributes to B cell transformation through HMGN1 overexpression and loss of histone H3 Lys27 trimethylation

Andrew A Lane[1], Bjoern Chapuy[1], Charles Y Lin[1], Trevor Tivey[1], Hubo Li[2], Elizabeth C Townsend[1], Diederik van Bodegom[1], Tovah A Day[1], Shuo-Chieh Wu[1], Huiyun Liu[1], Akinori Yoda[1], Gabriela Alexe[2], Anna C Schinzel[1,3], Timothy J Sullivan[4], Sébastien Malinge[5], Jordan E Taylor[3], Kimberly Stegmaier[2,3], Jacob D Jaffe[3], Michael Bustin[6], Geertruy te Kronnie[7], Shai Izraeli[8,9], Marian H Harris[10], Kristen E Stevenson[11], Donna Neuberg[11], Lewis B Silverman[2], Stephen E Sallan[2], James E Bradner[1], William C Hahn[1,3], John D Crispino[12], David Pellman[2,13] & David M Weinstock[1,3]

Down syndrome confers a 20-fold increased risk of B cell acute lymphoblastic leukemia (B-ALL)[1], and polysomy 21 is the most frequent somatic aneuploidy among all B-ALLs[2]. Yet the mechanistic links between chromosome 21 triplication and B-ALL remain undefined. Here we show that germline triplication of only 31 genes orthologous to human chromosome 21q22 confers mouse progenitor B cell self renewal *in vitro*, maturation defects *in vivo* and B-ALL with either the BCR-ABL fusion protein or CRLF2 with activated JAK2. Chromosome 21q22 triplication suppresses histone H3 Lys27 trimethylation (H3K27me3) in progenitor B cells and B-ALLs, and 'bivalent' genes with both H3K27me3 and H3K4me3 at their promoters in wild-type progenitor B cells are preferentially overexpressed in triplicated cells. Human B-ALLs with polysomy 21 are distinguished by their overexpression of genes marked with H3K27me3 in multiple cell types. Overexpression of HMGN1, a nucleosome remodeling protein encoded on chromosome 21q22 (refs. 3–5), suppresses H3K27me3 and promotes both B cell proliferation *in vitro* and B-ALL *in vivo*.

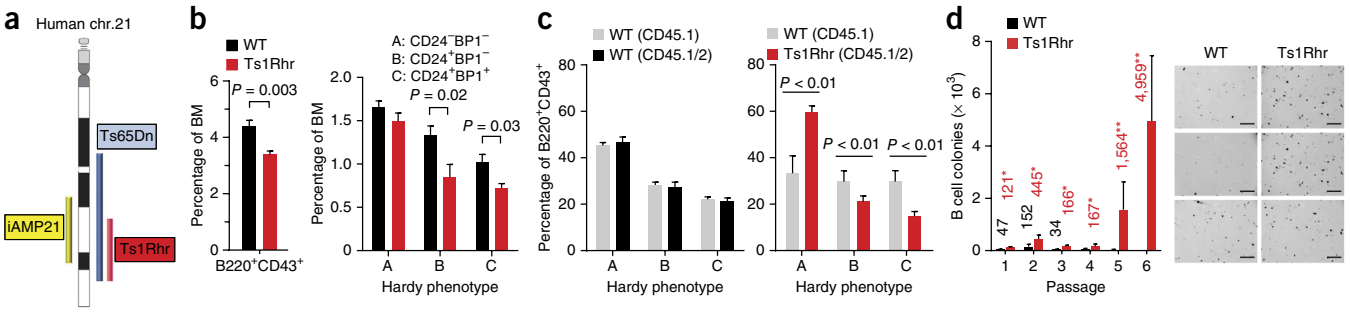
To interrogate the effects of polysomy 21 directly, we assayed B cell development in Ts1Rhr mice (**Fig. 1a**), which harbor a triplication of 31 genes and one noncoding RNA on mouse chromosome 16 orthologous to a segment of human chromosome 21q22 (ref. 6). The triplicated genes also overlap with a region of recurrent intrachromosomal amplification of chromosome 21q22 (iAMP21) in human B-ALL[7]. Bone marrow from 6-week-old Ts1Rhr mice had fewer total progenitor B ($B220^+CD43^+$) cells than that from wild-type littermates (**Fig. 1b**). Within the $B220^+CD43^+$ compartment, the Hardy B and C fractions but not the Hardy A fraction were reduced compared with wild-type littermates (**Fig. 1b** and **Supplementary Fig. 1a**)[8]. C57BL/6 Ts1Rhr, FVB × C57BL/6 F1 Ts1Rhr and Ts65Dn mice[9], which harbor a larger triplication (**Fig. 1a**), all had similar reductions in progenitor B cell fractions (**Supplementary Fig. 1b**). This differentiation defect essentially phenocopies human fetal livers with trisomy 21, which have reduced numbers of pre-pro-B ($CD34^+CD19^+CD10^-$) and pro-B ($CD34^+CD19^+CD10^+$) cells[10].

We performed competitive transplantation using equal mixtures of congenic $CD45.1^+$ wild-type bone marrow and $CD45.1^+CD45.2^+$ bone marrow from either Ts1Rhr or wild-type mice (**Supplementary Fig. 1c**). After 16 weeks, recipients of wild-type $CD45.1^+$ bone marrow mixed with wild-type $CD45.1^+CD45.2^+$ bone marrow had equal representations of both populations in the Hardy A, B and C fractions, as well as in whole bone marrow (**Fig. 1c** and **Supplementary Fig. 1d**). In contrast, mice that received wild-type $CD45.1^+$ bone marrow mixed with Ts1Rhr $CD45.1^+CD45.2^+$ bone marrow recapitulated the Ts1Rhr defect, with significant reductions in the $CD45.1^+CD45.2^+$ population in the Hardy B and C fractions (**Fig. 1c** and **Supplementary Fig. 1d**). Thus, the differentiation effect is independent of nonhematopoietic cells.

To address whether chromosome 21q22 triplication directly confers transformed phenotypes *in vitro*, we generated progenitor B cell colonies from unselected Ts1Rhr and wild-type bone marrow

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA. [2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA. [3]Broad Institute, Cambridge, Massachusetts, USA. [4]Microarray Core, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA. [5]Institut National de la Santé et de la Recherche Médicale (INSERM) U985, Institut Gustave Roussy, Villejuif, France. [6]Laboratory of Metabolism, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. [7]Department of Pediatrics, University of Padova, Padova, Italy. [8]Department of Pediatric Hemato-Oncology, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel. [9]Department of Human Molecular Genetics and Biochemsitry, Tel Aviv University, Tel Aviv, Israel. [10]Department of Pathology, Children's Hospital Boston, Boston, Massachusetts, USA. [11]Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA. [12]Division of Hematology/Oncology, Northwestern University, Chicago, Illinois, USA. [13]Howard Hughes Medical Institute, Chevy Chase, Maryland, USA. Correspondence should be addressed to A.A.L. (andrew_lane@dfci.harvard.edu) or D.M.W. (dweinstock@partners.org).

Received 13 January; accepted 13 March; published online 20 April 2014; doi:10.1038/ng.2949



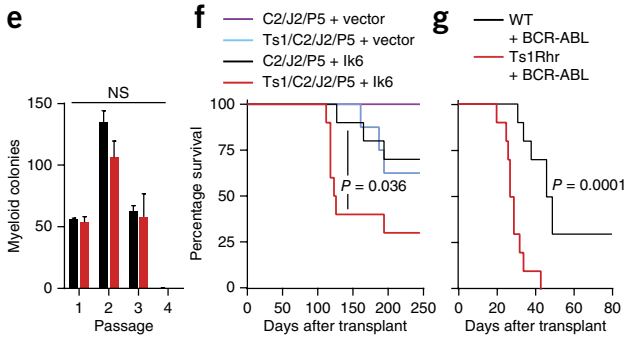


Figure 1 Segmental trisomy orthologous to human chromosome 21q22 promotes progenitor B cell transformation. (**a**) Regions orthologous to human chromosome (chr.) 21 that are triplicated in Ts1Rhr and Ts65Dn mice or amplified in iAMP21 B-ALL. (**b**) Progenitor B cells (B220$^+$CD43$^+$) and Hardy subfractions as percentages of bone marrow (BM) cells ($n = 7$ (WT, wild type) or 5 (Ts1Rhr) per group compiled from two independent experiments). (**c**) Subfractions from mixed populations in recipient BM 16 weeks after competitive transplantation ($n = 5$ per group). CD45.1/2 indicates CD45.1$^+$CD45.2$^+$. (**d**) B cell colonies across six passages ($n = 3$ biological replicates per genotype representative of three independent experiments; the values above the bars are the mean values; $^*P < 0.05$, $^{**}P < 0.01$) and bright field microscopy of three Ts1Rhr and three WT passage 2 cultures (scale bars, 4 mm). (**e**) Myeloid colonies across four passages ($n = 3$ mice per genotype; NS, not significant). Error bars (**b**–**e**), s.e.m. Groups (**b**–**e**) were compared by *t*-test. (**f**) Leukemia-free survival of recipient mice after transplantation of C2/J2/*Pax5*$^{+/-}$ (C2/J2/P5) with or without Ts1Rhr (Ts1) BM transduced with vector or Ik6 ($n = 8$ (C2/J2/P5 + vector), 8 (Ts1/C2/J2/P5 + vector), 10 (C2/J2/P5 + Ik6) or 10 (Ts1/C2/J2/P5 + Ik6) mice per group; curves compared by log-rank test). (**g**) Leukemia-free survival of recipient mice after transplantation of BM transduced with BCR-ABL ($n = 10$ mice per group, curves compared by log-rank test).

in three-dimensional cultures with IL-7 (**Supplementary Fig. 1e**,**f**). Wild-type bone marrow formed colonies (termed 'passage 1') that replated to form new colonies for two to three additional passages. In contrast, Ts1Rhr bone marrow generated more colonies in early passages and serially replated indefinitely (**Fig. 1d**), which indicates self-renewal capacity. There were no differences between Ts1Rhr and wild-type bone marrow in the number or repassaging potential of myeloid colonies (**Fig. 1e**).

Sixty percent of Down syndrome–associated B-ALLs harbor *CRLF2* rearrangements that commonly co-occur with activating JAK2 alterations[11–13]. To model this co-occurrence, we generated Eμ-CRLF2 (hereafter called 'C2') and Eμ-JAK2 p.Arg683Gly ('J2') transgenic mice, in which transgene expression is restricted to B cells. Mice harboring C2 and J2 (C2/J2) and *Pax5*$^{+/-}$ mice harboring C2 and J2 (C2/J2/*Pax5*$^{+/-}$) mice did not develop B-ALL by 18 months of age (data not shown). Transduction of C2/J2/*Pax5*$^{+/-}$ bone marrow with dominant-negative *IKZF1* (Ik6)[14] and transplantation into wild-type recipients resulted in CRLF2-positive B-ALL in all mice by 120 days after transplantation (**Supplementary Fig. 2a**,**b**). Control mice lacking C2, J2 or *Pax5* heterozygosity did not develop B-ALL with Ik6 (**Supplementary Fig. 2b**), establishing this transgenic combination as the first model of CRLF2- and JAK2-driven B-ALL. Mice transplanted with Ts1Rhr/C2/J2/*Pax5*$^{+/-}$ bone marrow transduced with a lower titer of Ik6-encoding virus developed B-ALL with greater penetrance and reduced latency compared to mice transplanted with C2/J2/*Pax5*$^{+/-}$ bone marrow alone (**Fig. 1f**). The same genotypes (C2/J2/*Pax5*$^{+/-}$/Ik6 with or without polysomy 21) occur in high-risk cases of human B-ALL[15], supporting the validity of the model.

Although BCR-ABL–rearranged ALL is uncommon in children with Down syndrome, polysomy 21 is the most common somatic aneuploidy in BCR-ABL–rearranged non-Down syndrome–associated B-ALLs[16]. Ts1Rhr and wild-type bone marrow had similar transduction efficiencies with p210 BCR-ABL[17] (**Supplementary Fig. 2c**), but mice (C57BL/6 and FVB × C57BL/6 F1 backgrounds) that received transduced Ts1Rhr bone marrow succumbed to B-ALL with shorter latency and increased penetrance (**Fig. 1g** and **Supplementary Fig. 2d**–**f**). Transplantation of BCR-ABL–transduced sorted Hardy B cells from Ts1Rhr or wild-type mice recapitulated the same effect (**Supplementary Fig. 2g**), indicating that chromosome 21q22 triplication confers leukemogenic effects that are progenitor B-cell autonomous.

Previous reports suggested that polysomy 21 contributes to leukemogenesis by promoting aberrant DNA double-strand break repair (DSBR)[18,19]. To address this possibility, we generated otherwise isogenic retinal pigment epithelial (RPE) cells with three or four copies of human chromosome 21 (**Supplementary Fig. 3a**–**c**). Using targeted DSBR reporters[20,21], we found that polysomy 21 had no effect on either homology-directed repair frequency or junction characteristics formed by nonhomologous end joining, whether DSBs were induced by the I-SceI meganuclease or the V(D)J recombinase (**Supplementary Fig. 3d**–**j**). Although a subtle defect or a defect specific to B cells remains possible, these results indicate that in an isogenic system, polysomy 21 does not drastically affect DSBR phenotype.

We next performed whole-transcriptome sequencing (RNA-seq) of passage 1 B cells. As expected, triplicated loci in Ts1Rhr cells were expressed at approximately 1.5-fold higher levels compared to wild-type cells (**Supplementary Fig. 4**). We defined a transcriptional Ts1Rhr gene set of the 150 most differentially expressed genes compared to wild type (**Supplementary Table 1a**,**b**). As expected, this signature was highly enriched by gene set enrichment analysis (GSEA)[22] for human chromosome 21q22 genes but not other human chromosomal segments (**Supplementary Table 1c**). The Ts1Rhr B cell signature was enriched among human Down syndrome–associated ALLs by GSEA (**Fig. 2a**,**b**; false discovery rate (FDR) = 0.019) in a data set of pediatric B-ALLs (Associazione Italiana Ematologia Oncologia Pediatrica (AIEOP))[19]. By hierarchical clustering, we defined a core Ts1Rhr set of 50 genes (**Fig. 2a** and **Supplementary Table 1a**) that contained none of the triplicated genes in Ts1Rhr cells but was highly enriched among Down syndrome–associated ALLs in

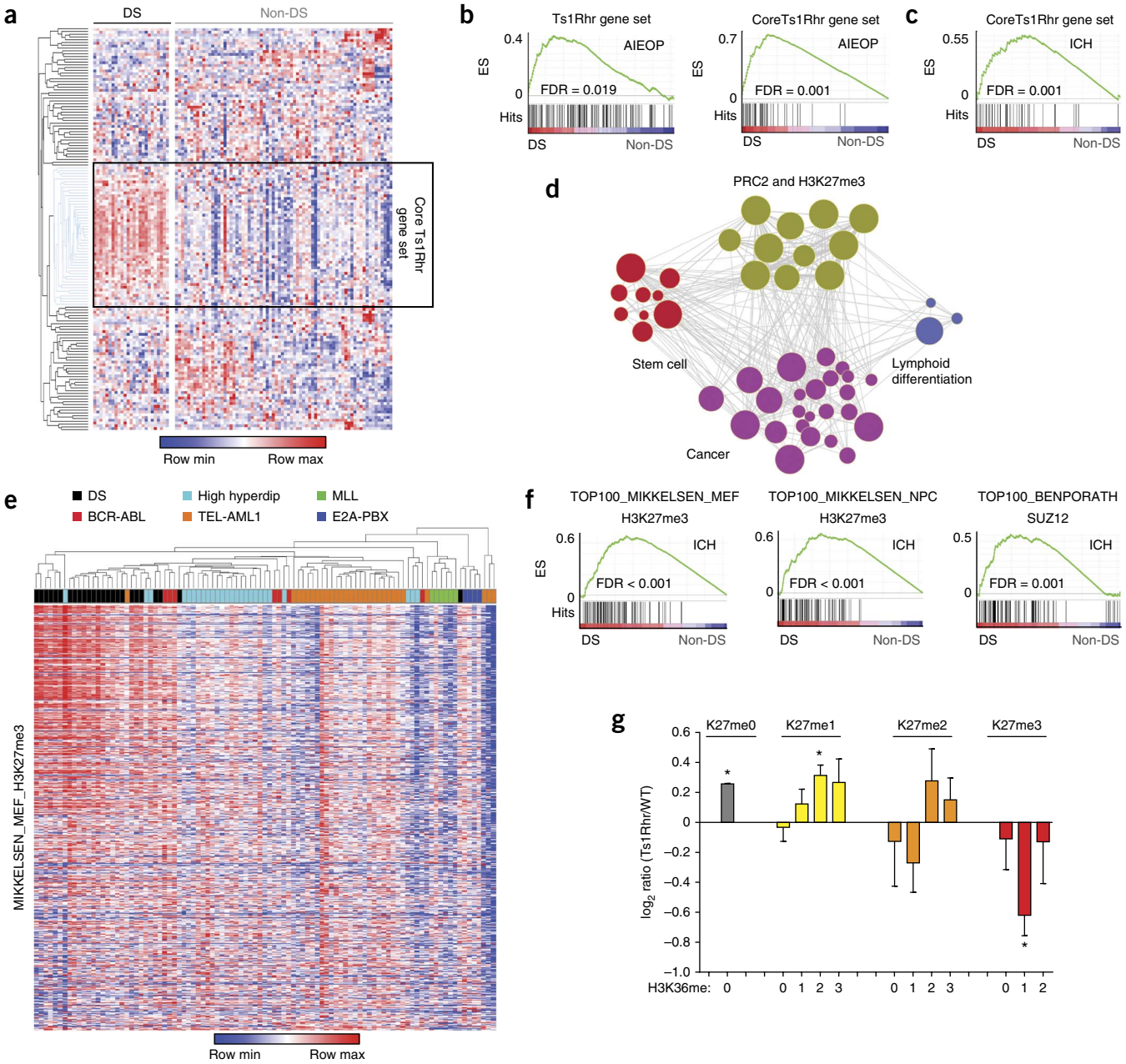


Figure 2 Polysomy 21 B-ALL is associated with the overexpression of PRC2 targets. (**a**) Heat map of human genes orthologous to the 150 most upregulated genes from Ts1Rhr B cells in primary human pediatric B-ALLs (DS indicates Down syndrome–associated ALL; non-DS indicates non–Down syndrome–associated ALL). Unsupervised hierarchical clustering by gene revealed the core Ts1Rhr gene set (boxed). (**b**) GSEA plots for the full and core Ts1Rhr gene sets in the AIEOP data set. ES, enrichment score. (**c**) GSEA plot of the core Ts1Rhr gene set in an independent ICH validation cohort. (**d**) Network enrichment map of MSigDB gene sets enriched (FDR < 0.05) in the Ts1Rhr expression signature. (**e**) Unsupervised hierarchical clustering of H3K27me3-marked genes from the MIKKELSEN_MEF_H3K27me3 gene set in the AIEOP pediatric B-ALL cohort (karyotype shown). High hyperdip, high hyperdiploid. (**f**) GSEA plots of the top 100 genes from three PRC2 and H3K27me3 gene sets (as defined in the AIEOP patient cohort) in the ICH validation cohort. (**g**) Quantitative histone mass spectrometry for H3K27–H3K36 peptides (* $P < 0.05$; $n = 3$ samples per group per genotype; each group compared to $\log_2$ ratio of zero by *t*-test). Error bars, s.e.m.

both the AIEOP data set (**Fig. 2c**; FDR = 0.001) and an independent validation data set (Institute of Child Health, London (ICH)) (**Fig. 2c**; FDR = 0.001).

To identify pathways perturbed by chromosome 21q22 triplication, we queried the Ts1Rhr gene set against over 3,000 functionally defined gene sets in the MSigDB 'c2' chemical and genetic perturbations and 'c6' oncogenic signatures repositories[22]. Arranging the significant gene sets in a network enrichment map[23] defined four clusters (**Fig. 2d**). The most highly enriched cluster consisted of polycomb repressor complex 2 (PRC2) targets and sites of H3K27me3, the repressive mark added by PRC2, that were defined across multiple lineages (**Supplementary Table 1d**,**e**). The additional clusters consisted of gene sets that distinguish stem cells from lineage-matched differentiated cells, cancer cells from nonmalignant cells or less differentiated from more differentiated lymphoid cells (**Supplementary Table 1d**).

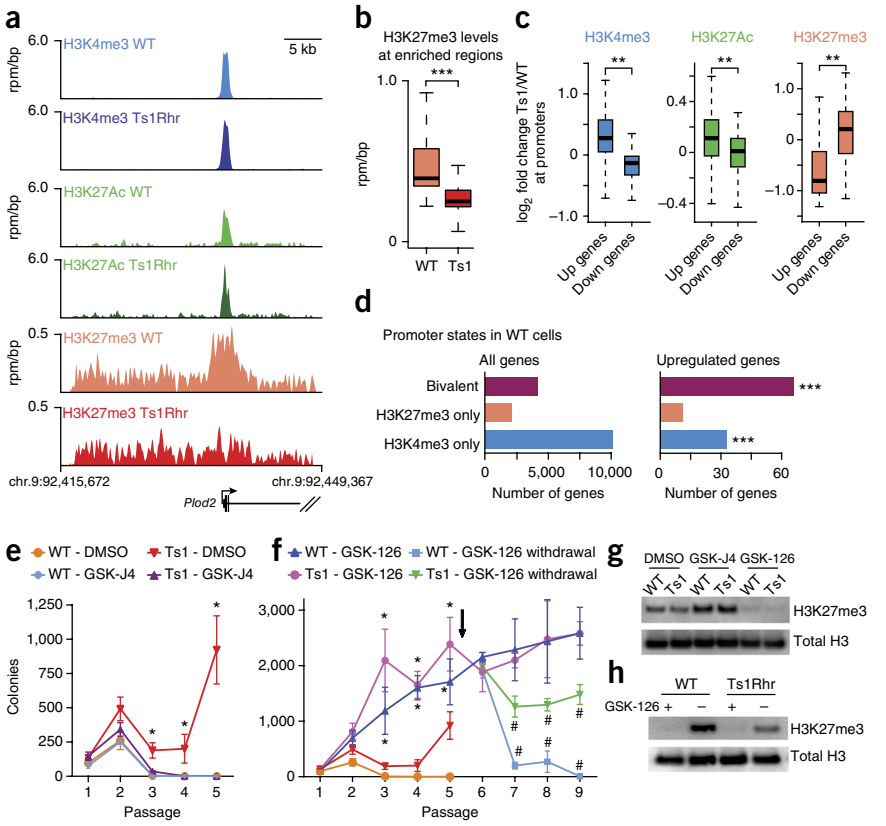


Figure 3 Ts1Rhr B cells have reduced H3K27me3 that results in overexpression of bivalently marked genes. (**a**) Gene tracks showing occupancy of histone marks at the *Plod2* promoter (1 of the 50 core Ts1Rhr genes) in reads per million per base pair (rpm/bp). (**b**) Levels of H3K27me3 in Ts1Rhr and WT B cells at regions enriched for H3K27me3 in WT cells (***$P < 1 \times 10^{-16}$; $n = 15{,}323$ loci per group; groups compared by *t*-test) (**c**) Histone marks at the promoters of genes that are upregulated (Up) or downregulated (Down) in Ts1Rhr compared to WT (Ts1/WT) cells (**$P < 1 \times 10^{-5}$; $n = 132$ (Up) or 163 (Down) genes; groups compared by *t*-test). The box plots in **b** and **c** show the median (horizontal line), 1 s.d. (box) and 2 s.d. (whiskers). (**d**) Chromatin marks in WT B cells present at promoters of all genes (left) or genes that are upregulated in Ts1Rhr B cells (right; ***$P < 0.0001$ compared to all genes by χ^2 with Yates' correction). (**e**) Colony counts in the presence of dimethyl sulfoxide (DMSO) or GSK-J4 ($n = 6$ (WT – DMSO), 6 (Ts1 – DMSO), 3 (WT – GSK-J4) or 3 (Ts1 – GSK-J4) biological replicates per genotype; *$P < 0.05$ compared to GSK-J4 for the same genotype; groups compared by *t*-test). (**f**) Colony counts in the presence of GSK-126 or after withdrawal at passage 5 ($n = 6$ (WT – DMSO), 6 (Ts1 – DMSO), 3 (WT – GSK-126), 3 (Ts1 – GSK-126), 3 (WT – GSK-126 withdrawal) or 3 (Ts1 – GSK-126 withdrawal) biological replicates per genotype; *$P < 0.05$ compared to DMSO for the same genotype, #$P < 0.05$ compared to the other genotype or no withdrawal; groups compared by *t*-test). The arrow indicates GSK-126 withdrawal. Error bars (**e**,**f**), s.e.m. (**g**) Western blotting of passage 2 colonies after 14 total days in culture with DMSO, 1 μM GSK-J4 or 1 μM GSK-126. (**h**) Western blotting of colonies one passage (7 days) after continuation (+) or removal (–) of GSK-126.

We next asked whether differential expression of PRC2- and H3K27me3-classified genes would distinguish Down syndrome–associated ALLs from other B-ALLs. A previous effort using genome-wide expression in the AIEOP cohort failed to define a transcriptional signature that was specific to Down syndrome–associated ALL[19]. Strikingly, expression of H3K27me3-marked genes defined in mouse embryonic fibroblasts[24] distinguished Down syndrome–associated ALLs from non–Down syndrome–associated ALLs (**Fig. 2e**). To validate these findings, we determined the 100 most differentially expressed genes between Down syndrome–associated ALLs and non–Down syndrome–associated ALLs in the AIEOP cohort across three different PRC2 and H3K27me3 signatures (**Supplementary Fig. 5a** and **Supplementary Table 1e**)[24,25]. All three signatures were significantly enriched (FDR ≤ 0.001) among Down syndrome–associated ALLs in the ICH validation cohort (**Fig. 2f**). In a third cohort of non–Down syndrome–associated ALLs (AIEOP-2), leukemias with either polysomy 21 or iAMP21 clustered on the basis of expression of PRC2 targets (**Supplementary Fig. 5b**; $P = 0.001$ by Fisher's exact test), and the Ts1Rhr and H3K27me3 gene sets were enriched among cases with polysomy 21 or iAMP21 by GSEA (**Supplementary Fig. 5c**).

Genes from PRC2 and H3K27me3 gene sets that distinguish Down syndrome–associated ALLs are predominantly overexpressed in Down syndrome–associated ALL (**Fig. 2e** and **Supplementary Fig. 5a**), suggesting that Down syndrome–associated ALL is associated with PRC2 target derepression through reduced H3K27me3. Histone H3 mass spectrometry confirmed a global reduction in H3K27me3 peptides in passage 1 Ts1Rhr B cells compared to wild-type cells (**Fig. 2g**). BCR-ABL B-ALLs from Ts1Rhr bone marrow also had reduced H3K27me3 by both mass spectrometry and immunoblotting (**Supplementary Fig. 5d**,**e**). Thus, triplication of only 31 genes orthologous to chromosome 21q22 is sufficient to suppress H3K27me3.

Chromatin immunoprecipitation sequencing (ChIP-seq) of passage 1 Ts1Rhr B cells demonstrated a genome-wide reduction of H3K27me3 at regions enriched for this mark in wild-type cells (**Fig. 3a**,**b**) that was confirmed at multiple loci by ChIP quantitative PCR (ChIP-qPCR) (**Supplementary Fig. 6a**). Within Ts1Rhr B cells, H3K27me3 was present almost exclusively at regions enriched for H3K27me3 in wild-type cells, suggesting little or no redistribution but rather a global reduction in H3K27me3 density (**Supplementary Fig. 6b–d**). As expected, we observed reciprocal changes in activating (H3K4me3 and acetylation of H3 Lys27 (H3K27ac)) and repressive (H3K27me3) marks at promoters of genes differentially expressed in Ts1Rhr B cells (**Fig. 3c**).

Of note, genes bivalently marked with both H3K27me3 and H3K4me3 in wild-type cells were highly enriched among those overexpressed in Ts1Rhr B cells (**Fig. 3d**; $P < 0.0001$). Bivalent marks may indicate genes that are modulated during lineage-specific differentiation[26]. Thus, global loss of H3K27me3 from chromosome 21q22 triplication could selectively drive a progenitor B cell–specific developmental program. In support of this hypothesis, the Ts1Rhr, PRC2 and H3K27me3 gene sets were highly enriched for predicted binding sites of the master B cell transcription factors E2A (also called TCF3) and LEF1 (**Supplementary Fig. 6e**). Genes within the Ts1Rhr gene set had increased proximal occupancy by E2A (**Supplementary Fig. 6f**) based on a previous data set from wild-type and E2A-deficient mouse B cell progenitors[27]. In addition, the expression of genes within both the Ts1Rhr gene set

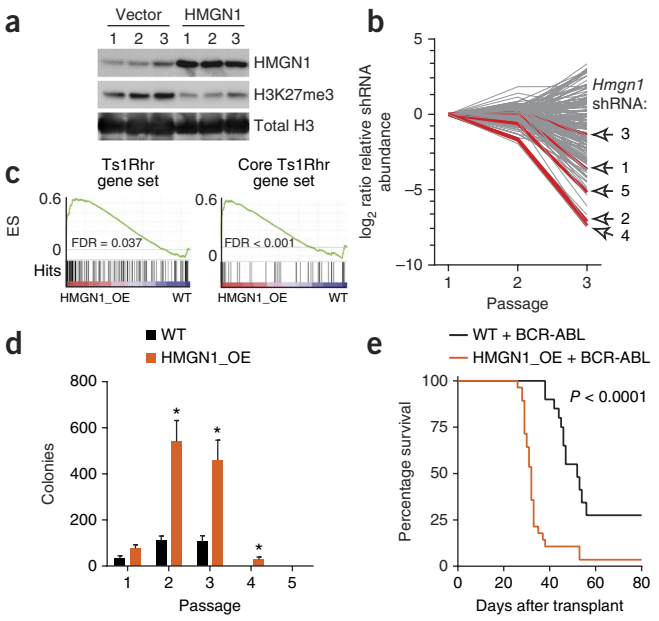


Figure 4 HMGN1 overexpression decreases H3K27me3 and promotes transformed B cell phenotypes. (**a**) Western blotting of Ba/F3 cells transduced with empty virus or mouse HMGN1 ($n = 3$ independent biological replicates). (**b**) Relative shRNA representation over passages 1–3. Each line represents an individual shRNA ($n = 155$ total). The five shRNAs targeting *Hmgn1* are indicated. (**c**) GSEA plots for the full and core Ts1Rhr gene sets in HMGN1_OE transgenic B cells. (**d**) B cell colonies during repassaging of WT and HMGN1_OE BM ($n = 4$ (WT) or 5 (HMGN1_OE) biological replicates per genotype compiled from two independent experiments; $^*P < 0.05$; groups compared by *t*-test). Error bars, s.e.m. (**e**) Leukemia-free survival of recipient mice after transplantation of WT or HMGN1_OE BM transduced with BCR-ABL (aggregate of three independent experiments (**Supplementary Fig. 8**), $n = 20$ (WT) or $n = 28$ (HMGN1_OE) per group, curves compared by log-rank test).

and the core Ts1Rhr set was preferentially increased in the presence of E2A (**Supplementary Fig. 6g**).

We hypothesized that pharmacologic restoration of H3K27me3 with GSK-J4 (ref. 28), a selective inhibitor of H3K27 demethylases, would block Ts1Rhr B cell repassaging. GSK-J4 increased H3K27me3 in Ts1Rhr B cells, decreased colony-forming activity and blocked indefinite repassaging (**Fig. 3e**,**g**). Previous studies demonstrated that 10 μM GSK-J4 reduces lipopolysaccharide-induced proinflammatory cytokine production by human primary macrophages[28]. In comparison, half-maximum inhibitory (IC_{50}) values for GSK-J4 across a panel of Down syndrome–associated ALLs ranged from only 1.4 to 2.5 μM (**Supplementary Fig. 6h**). Strikingly, treatment with GSK-126 (ref. 29), a selective inhibitor of the PRC2 catalytic subunit EZH2, decreased H3K27me3 and was sufficient to confer indefinite repassaging in wild-type B cells that was reversible by drug withdrawal (**Fig. 3f–h**).

Among the 31 triplicated genes in Ts1Rhr cells is *Hmgn1*, which encodes a nucleosome binding protein that modulates transcription and promotes chromatin decompaction[3,5]. Modest increases in HMGN1 expression induce changes in histone H3 modifications and gene expression[4,30]. Overexpression of HMGN1 in mouse Ba/F3 B cells suppressed H3K27me3 in a dose-dependent fashion (**Fig. 4a** and **Supplementary Fig. 7a**). By RNA-seq, *Hmgn1* was one of only seven triplicated genes that maintained >70% of its passage 1 expression level at passages 3 and 6 in all Ts1Rhr replicates (**Supplementary Fig. 7b**), suggesting it may be necessary for serial repassaging. To address this possibility, we transduced five short hairpin RNAs (shRNAs) targeting each of the 31 triplicated genes and controls individually into Ts1Rhr and wild-type passage 1 B cells (**Supplementary Fig. 7c**), pooled and serially passaged the transduced cells.

As expected, positive control shRNAs were equally depleted at later passages from Ts1Rhr and wild-type backgrounds (**Supplementary Fig. 7d** and **Supplementary Table 1g**). Among shRNAs against triplicated genes, two of the top four that most selectively depleted Ts1Rhr B cells targeted *Hmgn1* (**Fig. 4b** and **Supplementary Table 1h**). The remaining three shRNAs targeting *Hmgn1* were also preferentially depleted in Ts1Rhr B cells. By passage 6, all five shRNAs against *Hmgn1* were depleted >99%, averaged across replicates of Ts1Rhr B cells (data not shown). All five shRNAs also reduced HMGN1 protein in Ba/F3 cells (**Supplementary Fig. 7e**).

We then analyzed mice with transgenic overexpression of human HMGN1 (HMGN1_OE) at levels comparable to mouse HMGN1 (**Supplementary Fig. 7f**)[31]. HMGN1_OE passage 1 B cells had a gene expression signature that was highly enriched for the Ts1Rhr and core Ts1Rhr gene sets (**Fig. 4c**). Compared to control bone marrow, HMGN1_OE bone marrow had reduced numbers of Hardy C cells *in vivo* (**Supplementary Fig. 7g**), generated more B cell colonies in passages 1–4 *in vitro* (**Fig. 4d**) and resulted in greater penetrance and shorter latency of BCR-ABL–induced B-ALL (**Fig. 4e**). Thus, HMGN1 overexpression recapitulates many transcriptional and phenotypic alterations observed from triplication of all 31 Ts1Rhr genes.

In conclusion, we have shown that triplication of chromosome 21q22 genes confers cell-autonomous differentiation and transformation phenotypes in progenitor B cells. By first delineating these biologic consequences of chromosome 21q22 triplication, we were able to more effectively interrogate human B-ALL data sets and demonstrate that Down syndrome–associated ALLs are distinguished by overexpression of H3K27me3-marked genes. Our data also highlight the therapeutic potential of H3K27 demethylase inhibitors for B-ALLs with extra copies of chromosome 21q22. At the same time, EZH2 inhibitors may be useful for *in vitro* or *in vivo* expansion of precursor B cells. We also provide evidence that overexpression of HMGN1 suppresses global H3K27me3 and promotes B-ALL *in vivo*. Further studies will be needed to determine how HMGN1 modulates transcription at differentially expressed loci in cells with polysomy 21, as well as the contributions from other triplicated chromosome 21q22 loci.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. RNA-seq, ChIP-seq and microarray expression data are available through Gene Expression Omnibus (GEO) accession code GSE48555.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank N. Kopp and A. Schlauch for technical assistance, M. Busslinger (Research Institute of Molecular Pathology, Vienna) for the *Pax5*$^{+/-}$ mice and M. Oshimura (Tottori University) for A9 cells carrying human chromosome 21. This research was supported by the Conquer Cancer Foundation (A.A.L.), the Lauri Strauss Leukemia Foundation (A.A.L.), the Leukemia and Lymphoma Society (A.A.L.), the Alex Lemonade Stand Foundation (A.A.L., H. Li and D.P.), the US Department of Defense (C.Y.L.), the Israel Science Foundation (S.I.), the US Israel Binational Foundation (J.D.C. and S.I.), the Stellato Fund (D.M.W.),

US National Institutes of Health/National Cancer Institute R01 awards CA15198-01 and CA172387-A01 (D.M.W.) and a Translational Research Award from the Leukemia and Lymphoma Society (J.D.C. and D.M.W.).

AUTHOR CONTRIBUTIONS

A.A.L., T.T., H. Li, E.C.T., D.v.B., T.A.D., S.-C.W., H. Liu, A.Y., S.M., J.E.T., J.D.J. and G.t.K. designed and performed experiments. A.A.L., B.C., C.Y.L., G.A., T.J.S., K.S., K.E.S., D.N., J.D.J., S.I., J.D.C., D.P. and D.M.W. analyzed data. A.C.S., J.E.B., W.C.H. and J.D.J. developed analytical tools. M.B., M.H.H., L.B.S., S.E.S. and D.P. provided essential reagents. A.A.L. and D.M.W. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Rabin, K.R. & Whitlock, J.A. Malignancy in children with trisomy 21. *Oncologist* **14**, 164–173 (2009).
2. Heerema, N.A. *et al.* Specific extra chromosomes occur in a modal number dependent pattern in pediatric acute lymphoblastic leukemia. *Genes Chromosom. Cancer* **46**, 684–693 (2007).
3. Catez, F., Brown, D.T., Misteli, T. & Bustin, M. Competition between histone H1 and HMGN proteins for chromatin binding sites. *EMBO Rep.* **3**, 760–766 (2002).
4. Lim, J.H. *et al.* Chromosomal protein HMGN1 enhances the acetylation of lysine 14 in histone H3. *EMBO J.* **24**, 3038–3048 (2005).
5. Rattner, B.P., Yusufzai, T. & Kadonaga, J.T. HMGN proteins act in opposition to ATP-dependent chromatin remodeling factors to restrict nucleosome mobility. *Mol. Cell* **34**, 620–626 (2009).
6. Olson, L.E., Richtsmeier, J.T., Leszl, J. & Reeves, R.H. A chromosome 21 critical region does not cause specific Down syndrome phenotypes. *Science* **306**, 687–690 (2004).
7. Moorman, A.V. *et al.* Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial. *Lancet Oncol.* **11**, 429–438 (2010).
8. Hardy, R.R., Carmack, C.E., Shinton, S.A., Kemp, J.D. & Hayakawa, K. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J. Exp. Med.* **173**, 1213–1225 (1991).
9. Reeves, R.H. *et al.* A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nat. Genet.* **11**, 177–184 (1995).
10. Roy, A. *et al.* Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy 21. *Proc. Natl. Acad. Sci. USA* **109**, 17579–17584 (2012).
11. Mullighan, C.G. *et al.* Rearrangement of CRLF2 in B-progenitor– and Down syndrome–associated acute lymphoblastic leukemia. *Nat. Genet.* **41**, 1243–1246 (2009).
12. Russell, L.J. *et al.* Deregulated expression of cytokine receptor gene, *CRLF2*, is involved in lymphoid transformation in B-cell precursor acute lymphoblastic leukemia. *Blood* **114**, 2688–2698 (2009).
13. Yoda, A. *et al.* Functional screening identifies CRLF2 in precursor B-cell acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. USA* **107**, 252–257 (2010).
14. Iacobucci, I. *et al.* Expression of spliced oncogenic Ikaros isoforms in Philadelphia-positive acute lymphoblastic leukemia patients treated with tyrosine kinase inhibitors: implications for a new mechanism of resistance. *Blood* **112**, 3847–3855 (2008).
15. Mullighan, C.G. *et al.* JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. USA* **106**, 9414–9418 (2009).
16. Wetzler, M. *et al.* Additional cytogenetic abnormalities in adults with Philadelphia chromosome-positive acute lymphoblastic leukaemia: a study of the Cancer and Leukaemia Group B. *Br. J. Haematol.* **124**, 275–288 (2004).
17. Krause, D.S., Lazarides, K., von Andrian, U.H. & Van Etten, R.A. Requirement for CD44 in homing and engraftment of BCR-ABL–expressing leukemic stem cells. *Nat. Med.* **12**, 1175–1180 (2006).
18. Cabelof, D.C. *et al.* Mutational spectrum at *GATA1* provides insights into mutagenesis and leukemogenesis in Down syndrome. *Blood* **114**, 2753–2763 (2009).
19. Hertzberg, L. *et al.* Down syndrome acute lymphoblastic leukemia, a highly heterogeneous disease in which aberrant expression of CRLF2 is associated with mutated JAK2: a report from the International BFM Study Group. *Blood* **115**, 1006–1017 (2010).
20. Smith, J.R. *et al.* Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem Cells* **26**, 496–504 (2008).
21. Weinstock, D.M. & Jasin, M. Alternative pathways for the repair of RAG-induced DNA breaks. *Mol. Cell. Biol.* **26**, 131–139 (2006).
22. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
23. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G.D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, e13984 (2010).
24. Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
25. Ben-Porath, I. *et al.* An embryonic stem cell–like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.* **40**, 499–507 (2008).
26. Bernstein, B.E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
27. Lin, Y.C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.* **11**, 635–643 (2010).
28. Kruidenier, L. *et al.* A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488**, 404–408 (2012).
29. McCabe, M.T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108–112 (2012).
30. Rochman, M. *et al.* Effects of HMGN variants on the cellular transcription profile. *Nucleic Acids Res.* **39**, 4076–4087 (2011).
31. Bustin, M. *et al.* Characterization of transgenic mice with an increased content of chromosomal protein HMG-14 in their chromatin. *DNA Cell Biol.* **14**, 997–1005 (1995).

ONLINE METHODS

Mice. All animal experiments were performed with approval of the Dana-Farber Cancer Institute (DFCI) Institutional Animal Care and Use Committee. All experiments were performed in an FVB × C57BL/6 F1 background unless otherwise specified. Ts1Rhr (B6.129S6-Dp(16Cbr1-ORF9)1Rhr/J; #005838) and Ts65Dn (B6EiC3Sn.BLiA-Ts(17^{16})65Dn/DnJ; #005252) mice were from Jackson Laboratories. HMGN1_OE mice were previously described[31]. *Pax5*$^{+/-}$ mice[32] backcrossed to C57BL/6 mice were obtained from M. Busslinger (Research Institute of Molecular Pathology, Vienna, Austria). Eμ-CRLF2 and Eμ-JAK2 p.Arg683Gly mice were generated by subcloing cDNAs expressing human CRLF2 or mouse JAK2 p.Arg683Gly[11,13] downstream of the immunoglobulin heavy chain enhancer (Eμ) and generating transgenic founders in FVB fertilized eggs as previously described[33]. Controls for Ts1Rhr mice were wild-type littermates from crosses with either C57BL/6 (Jackson; #000664) or FVB (Jackson; #001800) mice as indicated. Controls for Ts65Dn mice were littermates from the colony (B6EiC3Sn.BLiAF1/J; Jackson #003647). HMGN1_OE mice[31] had been backcrossed over ten generations to C57BL/6 mice[34]. Controls for HMGN1_OE were wild-type littermates after crossing with FVB mice. Donors for competitive transplantation were congenic CD45.1$^+$ B6.SJL-*Ptprc*^{a}*Pepc*^{b}/BoyJ mice (Jackson; stock #002014) crossed with FVB mice (CD45.1$^+$), C57BL/6 × FVB F1 (CD45.1$^+$CD45.2$^+$) mice or Ts1Rhr (C57BL/6) mice crossed with FVB F1 (CD45.1$^+$CD45.2$^+$) mice. Recipients for competitive transplant and for BCR-ABL and Ik6 bone marrow transplants were C57BL/6 × FVB F1 female mice. No randomization was performed for experiments involving mice or samples collected from animals.

Antibodies. Western blotting antibodies were to HMGN1 (Aviva Systems Biology, #ARP38532_P050 and Abcam, #ab5212; both at 1:1,000), mouse HMGN1 (affinity purified rabbit polyclonal; 1:1,000)[35,36], H3K27me3 (Cell Signaling Technologies, #9733, rabbit polyclonal; 1:1,000), total H3 (Cell Signaling Technologies, #9715, rabbit polyclonal; 1:2,000) and α-tubulin (Sigma, #T9026, mouse monoclonal; 1:2,000). Flow cytometry antibodies were B220-Pacific Blue (BD Pharmingen, #558108, clone RA3-6B2; 1:100), CD43-allophycocyanin (APC) (BD, #560663, clone S7; 1:200) or CD43-FITC (BD, #561856, clone S7; 1:200), CD24-phycoerythrin (PE)-Cy7 (BD, #560536, clone M1/69; 1:200), BP1-PE (eBiosciences, 12-5891, clone 6C3; 1:50) or BP1-FITC (eBiosciences, 11-5891, clone 6C3; 1:50), CD45.1-PE-Cy7 (eBiosciences, 25-0453, clone A20; 1:200) and CD45.2-APC (eBiosciences, 17-0454, clone 104; 1:200). ChIP-seq antibodies were to H3K27me3 (Cell Signaling Technologies, #9733; 10 μg/ChIP), H3K4me3 (Abcam, #ab8580; 5 μg/ChIP) and H3K27ac (Abcam, #ab4729; 5 μg/ChIP).

Flow cytometry for bone marrow B cells. Whole bone marrow was harvested from femurs and tibias of 6- to 8-week-old mice. After red blood cell lysis (Qiagen, #158904), B cell progenitors were stained using antibodies, and flow cytometry was performed as previously described[8]. Analysis was performed on a BD FACSCanto II.

Competitive bone marrow transplantation. Whole bone marrow was pooled from femurs and tibias of 8-week-old donor mice. Donor cells were wild-type or Ts1Rhr CD45.1$^+$CD45.2$^+$ C57BL/6 × FVB F1 (test) and CD45.1$^+$ B6.SJL × FVB F1 (competitor) and were mixed 1:1. Recipients were lethally irradiated (550 cGy twice, >4 h apart). B6.SJL × FVB F1 mice received 10^6 total cells by lateral tail vein injection. Bone marrow was harvested 16 weeks after transplantation and analyzed by flow cytometry.

Methylcellulose colony-forming assays. Whole bone marrow was harvested from 6-to 8-week-old mice, and red blood cells were lysed. Cells were plated in B cell (Methocult M3630, Stem Cell Technologies) or myeloid (Methocult M3434) methylcelluose medium in gridded 35-mm dishes. Myeloid colonies were plated at 2×10^4 cells/ml per passage. B cell colonies were plated at 2×10^5 cells/ml in passage 1 and at 5×10^4 cells/ml per subsequent passage. Colonies were counted at 7 d, and colonies were then pooled and replated in the same manner.

Bone marrow transplantation models. For BCR-ABL transplantations[17], 10^5 transduced cells were transplanted with 10^6 wild-type untransduced bone marrow cells for radioprotection. For generation of BCR-ABL B-ALLs derived from Hardy B cells, 5×10^4 Hardy B cells from 6-week-old mice were sorted on a BD FACSAria II SORP, spinoculation was performed as described above, and 10^3 cells were transplanted into lethally irradiated wild-type recipients with 10^6 bone marrow cells for radioprotection. Dominant-negative Ikaros experiments were performed similarly, except that 10^6 cells spinfected with a murine stem cell virus (MSCV) retrovirus either expressing GFP alone or coexpressing GFP and Ik6 (refs. 14,37) were transplanted. Mice were followed daily for clinical signs of leukemia and were euthanized when moribund. Investigators were not blinded to the experimental groups. Ten mice were used per arm for 80% power to detect a 60% difference in survival at a specific time point with α of 0.05. No animals were excluded from analysis.

Cell culture. Ba/F3 cell experiments were performed as previously described[13]. shRNAs targeting *Hmgn1* are described below (competitive shRNA assay section), and cDNA expressing HMGN1 was described previously[30]. One week after selection in puromycin, retroviral cDNA or lentiviral shRNA-transduced cells were harvested for western blotting. hTERT-RPE1 cells were cultured in DMEM/F-12. Mouse A9 cells containing a single human chromosome 21 tagged with a neomycin resistance gene (a gift from M. Oshimura, Tottori University, Japan) were cultured in DMEM. All media were supplemented with 10% FBS, 100 IU/ml penicillin and 100 μg/ml streptomycin.

Immunoblotting and quantification. Western blotting was performed as previously described[13]. ImageJ (http://imagej.nih.gov/ij) was used for quantification of immunoblots, with band intensities normalized to total H3.

Microcell-mediated chromosome transfer. Microcell-mediated chromosome transfer was performed as described previously[38] with modifications. A9 cells were cultured to ~70% confluence and treated with 75 ng/ml colcemid for 48 h. Cells were collected and resuspended in 1:1 DMEM:Percoll (GE Healthcare Biosciences) with 10 μg/ml cytochalasin B (Sigma-Aldrich) and spun at 17,000 r.p.m. for 75 min in a Beckman JA17 rotor. Supernatant was collected and filtered through 10- and 5-μm filters. Approximately 2×10^6 RPE1 cells were collected and mixed with filtered microcells, treated with 100 μg/ml PHA-P (Sigma-Aldrich) for 30 min and fused by PEG 1500 (Sigma-Aldrich) in solution. Hybrid cells were plated and cultured for 48 h and selected with 500 μg/ml Geneticin (Life Technologies) for 12–14 d. Standard G-band analysis was performed at Karyologic, Inc. SNP array was performed at the DFCI microarray core using the Human Mapping 250k-Nsp platform. Fluorescent *in situ* hybridization was performed with the Vysis LSI 21 SpectrumOrange probe (Abbott Molecular) according to the manufacturer's instructions.

DR-GFP and DR-GFP-CE reporter targeting. Generating and screening of targeted clones were performed as described[39] with the following modifications. 10^6 RPE1 cells with two, three or four copies of chromosome 21 were nucleofected with 2 μg pAAVS1-DR-GFP or pAAVS1-DR-GFP-CE plasmid together with 2 μg pZFN-AAVS1 using program X-001 of the Amaxa nucleofector II (Lonza). Targeting of individual clones was confirmed by PCR using the Accuprime GC-rich DNA polymerase (Life Technologies). The presence of a single integrant was determined by qPCR (data not shown).

DNA repair assays using DR-GFP reporter cell lines. Assays for homologous recombination and imprecise nonhomologous end joining were performed as described previously[40] with the following modifications. Transfections were performed with the Neon transfection system (Life Technologies) using 1,600 V, 20 ms and 1 pulse. 4×10^5 DR-GFP cells were transfected with 10 μg I-SceI expression vector (pCBASce) or empty vector (pCAGGS) and plated in six-well plates. pmCherry-C1 vector (Clontech) was transfected in parallel to confirm equal transfection efficiency. Cells were cultured for 7 d and analyzed by fluorescence-activated cell sorting (FACS) using FACSCalibur (BD Biosciences) for homology-directed repair. The remaining cells were used to extract genomic DNA. 1 μg DNA was digested with 20 U I-SceI (Roche) overnight, purified and amplified with a two-step PCR protocol. Accuprime GC-rich polymerase was used for the first-step PCR (20 cycles), and Taq polymerase (Qiagen) was used for the second-step PCR (20 cycles).

doi:10.1038/ng.2949

PCR products were cloned with the TOPO TA cloning kit for sequencing (Life Technologies). For DR-GFP-CE cells, pCAGGS-RAG1 and pCAGGS-RAG2 vectors were co-transfected. 1 μg genomic DNA was digested with 10 U MfeI and 10 U NdeI (NEB) overnight to exclude templates that had not been cleaved by RAG-1 and RAG-2 before PCR amplification. The primer sequences are available in **Supplementary Table 2**.

Competitive shRNA assay in primary B cells. shRNAs targeting triplicated Ts1Rhr genes and controls were obtained from The RNAi Consortium (http://www.broadinstitute.org/rnai/trc) as pLKO lentiviral supernatants[41] (n = 185; see **Supplementary Table 1** for clone ID numbers and target sequences). Wild-type or Ts1Rhr passage 1 B cell colonies were collected and plated at 5×10^4 cells per well of a 96-well plate in 100 μl of RPMI with 20% fetal bovine serum (FBS) and 10 ng/mL each of mouse IL-7, stem cell factor and FLT3 ligand (all from R&D Systems) with 8 μg/ml polybrene. 10 μl of lentiviral supernatant was added, and the plate was centrifuged at 1,000g for 30 min and then placed in a 37 °C incubator for 24 h. Wells were pooled, 10^6 cells were saved for input shRNA analysis and 2×10^6 cells were plated in 6 ml M3630 methylcellulose with 0.05 μg/ml puromycin in a 10 cm non–tissue culture treated dish. At this density of plating, after 7 d of growth there were at least 4×10^4 colonies per plate, which would represent >200 colonies per individual shRNA on average. After each passage, genomic DNA was harvested from 10^6 cells (Qiagen QIAmp kit), and 2×10^6 cells were replated in the same manner. Repassaging continued until cultures stopped forming new colonies (three to four passages for wild type) or until six passages were completed. The entire assay was repeated in n = 3 (wild type) or n = 4 (Ts1Rhr) independent biological replicates.

The shRNA encoded in the genomic DNA was amplified using two rounds of PCR. Primary PCR reactions were performed using up to 10 μg of genomic DNA in 100 μl reactions consisting of 10 μl Takara Ex Taq buffer, 8 μl dNTPs (2.5 mM each), 10 μl of 5 μM primary PCR primer mix (**Supplementary Table 2**) and 1.5 μl Takara ExTaq. For the secondary PCR amplification, the reaction was performed as described previously[41] using modified forward primers, which incorporated Illumina adaptors and 6-nucleotide barcodes. Secondary PCR reactions were pooled and run on a 2% agarose gel. The bands were normalized and pooled on the basis of relative intensity. An equal amount of sample was run on a 2% agarose gel and gel purified. Samples were sequenced using a custom sequencing primer on an Illumina Hi-Seq and quantified as previously described[41]. The primary, secondary and sequencing primers are listed in **Supplementary Table 2**.

RNA sequencing and data processing. Total RNA was harvested from B cell colonies (n = 3 independent biologic replicates per genotype per passage). RNA sequencing was performed at the DFCI Center for Cancer Computational Biology. Quality control of total RNA was performed using the RNA Qubit Assay (Invitrogen) and the Bioanalyzer RNA Nano 6000 Chip Kit (Agilent). At least 100 ng of total RNA and a Bioanalyzer RNA Integrity Number of >7.0 were required. Library construction was performed using the TruSeq RNA Library Prep Kit (Illumina). Final library quality control was performed using the DNA High Sensitivity Qubit Kit (Invitrogen), the Bioanalyzer High Sensitivity Chip Kit (Agilent) and the 7900HT Fast qPCR machine (Applied Biosystems). qPCR was performed using the Illumina Universal Library Quantification Kit from KAPA Biosystems. RNASeq libraries were normalized to 2 nM, pooled for multiplexing in equal volumes and sequenced at 10 pM on the Illumina HiSeq 2000. Sequencing was performed as 2 × 50 paired-end reads using the 100 cycles per lane Sanger/Illumina 1.9 deep sequencing protocol. The raw sequence data were subjected to data quality control checks based on per-base sequence quality scores, per-sequence quality scores, per-sequence GC content, sequence length distribution and over-represented sequences, which are implemented in the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads that passed quality control filters were aligned against the mouse reference genome by using the ultra high-throughput long read aligner Bowtie2 (ref. 42) available through TopHat 2.0.7 (http://tophat.cbcb.umd.edu)[43]. Mapping results were further analyzed with TopHat to identify splice junctions between exons. Genomic annotations in gene transfer format (GTF) were obtained from Ensembl mouse genome GRCm38 (http://useast.ensembl.org/Mus_musculus/Info/Index). Gene-level expression measurements for 23,021 Ensembl mouse genes were reported in fragments per kilobase per million reads (FPKM) by Cufflinks 2.0.0 (http://cufflinks.cbcb.umd.edu/)[44]. An FPKM filtering cutoff of 1 in at least one sample was used to determine expressed transcripts.

Differential analysis for RNA-seq transcript expression. Differential analysis was performed by applying the EdgeR method[45] implemented in Bioconductor v2.11 (http://www.bioconductor.org/). EdgeR uses empirical Bayes estimation and exact tests based on the negative binomial distribution model of the genome-scale count data. EdgeR estimates the genewise dispersions by conditional maximum likelihood, conditioning on the total count for that gene. The genewise dispersion is 'normalized' by shrinking toward a consensus value based on an empirical Bayes procedure[46]. The differential expression is estimated separately for each gene on the basis of an exact test analogous to Fisher's exact test adopted for overdispersed data[47].

Gene expression profiling and GSEA. The series matrix files for two DS ALL data sets (AIEOP and ICH) were downloaded from GEO (GSE17459)[19], as were those for the *Rag1*$^{-/-}$ and *E2A* (*Tcf3*)$^{-/-}$ B cell progenitors (GSE21978)[27]. RNA from HMGN1 transgenic (HMGN1_OE) or wild-type littermate B cell colonies was processed and hybridized to the Affymetrix Mouse Gene 2.0 ST array at the DFCI Microarray Core per the manufacturer's instructions. Raw probe-level data from the AIEOP-2 non-DS ALL cohort and the mouse HMGN1_OE gene expression profiling were summarized using the Robust Multiarray Average (RMA)[48] and Brainarray custom chip identification files based on Entrez IDs (version 17)[49] using the ExpressionFileCreator module in Gene Pattern[50]. For GSEA, the expression file was converted to human gene orthologs using BioMart[51]. GSEA of the Ts1Rhr, the core Ts1Rhr and the PRC2 gene sets was performed as described previously using GSEA v2.0.10 (http://www.broadinstitute.org/gsea/)[22]. The Ts1Rhr gene set was tested for its enrichment in the c1 (positional), c2.cgp (chemical and genetic perturbation), c3.tft (transcription factor targets) and c6 (oncogenic signatures) gene sets deposited in the Molecular Signature Database MSigDB v3.1 (Broad Institute; http://www.broadinstitute.org/gsea/msigdb). The analysis was performed by applying the two-tailed Fisher test method as implemented in the Investigate_GeneSets module at MSigDB. To define the Ts1Rhr B cell gene set, the top 150 most differentially expressed protein-coding genes with an adjusted P value below 0.25 were selected. Hierarchical clustering of this signature in DS ALL compared to non-DS ALL revealed a subset of genes most contributing to the distinguishing phenotype, and this branch defined the 'core' Ts1Rhr gene set. Full gene sets for BENPORATH_SUZ12_TARGETS, MIKKELSEN_MEF_HCP_WITH_H3K27ME3 and MIKKELSEN_MEF_NPC_WITH_H3K27ME3 were obtained from MSigDB v3.1. The 100 most differentially expressed genes between the DS ALLs and the non-DS ALLs were determined using the MarkerSelectionModule in Gene Pattern. For E2A target gene expression, we compared *Rag1*$^{-/-}$ pro-B cells to *E2A*$^{-/-}$ pre-pro-B cells to generate probesets with >1.5-fold change and $P < 0.05$ between conditions, exactly as had been done by previous authors[27]. The Ts1Rhr and core gene sets were compared to all probesets for their relative expression in *E2A* wild-type (*Rag1*$^{-/-}$ pro-B) compared to *E2A*$^{-/-}$ cells.

Network enrichment mapping. The gene sets with significant enrichment in genes upregulated in Ts1Rhr cells by GSEA were selected on the basis of the maximum cutoff value 0.05 for P value and FDR and visualized with Enrichment Map software[23]. This software organizes the significant gene sets into a network, where nodes correspond to gene sets and the edges reflect significant overlap between the nodes according to a Fisher's test. The size of the nodes is proportional to the number of genes in the gene set. The hubs correspond to collections of genes sets with significant pairwise overlap that have a unifying functional description according to Gene Ontology (GO) biological processes. The node color is associated to the functional description of the hub. The clusters provided by the Enrichment Map are described in **Supplementary Table 1d**.

Visualization of gene expression and mass spectrometry data. RNA-seq–derived expression data from Ts1Rhr and wild-type B cells, B-ALL gene expression data and histone mass spectrometry data were visualized as heat

maps using GENE-E (Broad Institute; http://www.broadinstitute.org/cancer/software/GENE-E/).

Column purification of mouse B-ALLs. For western blotting of mouse B-ALLs, we enriched cryopreserved B-ALL splenocytes using antibody to CD19 conjugated to magnetic microbeads (#130-052-201) and an MS MACS column (#130-042-201), both from Miltenyi Biotec.

Histone mass spectrometry. Mass spectrometry for global histone H3 post-translational modifications was performed as described previously[52] using wild-type or Ts1Rhr passage 1 B cells and BCR-ABL B-ALLs. H3K27 modifications are presented in conjunction with H3K36, as both are present in the same measured peptides because of their close proximity. Significance was calculated as difference from the log2 ratio of zero by *t*-test.

Drug treatment. GSK-J4 (#M60063-2) and GSK-126 (#M60071-2) were purchased from Xcessbio. For methylcellulose experiments, at each passage, DMSO, GSK-J4 or GSK-126 was added to cultures at a final concentration of 1 μM. DS ALLs (deidentified specimens obtained with informed consent under DFCI Institutional Review Board protocol 05-001) were treated *in vitro* in quadruplicate with GSK-J4 at twofold dilutions from 40 nM to 10 μM in RPMI with 20% calf serum supplemented with 10 ng/mL IL-3, IL-7, stem cell factor (SCF), FLT3 ligand and 50 μM β-mercaptoethanol. After 3 d, viability was measured using CellTiter-Glo reagent and normalized to DMSO control (Promega).

ChIP analyses. B cell colonies (>5,000 colonies per genotype) from three wild-type and three Ts1Rhr animals were pooled after 7 d in methylcellulose culture. ChIP was performed as described previously[53]. Libraries for sequencing were prepared following the Illumina TruSeq DNA Sample Preparation v2 kit protocol. After end repair and A tailing, immunoprecipitated DNA (10–50 ng) or DNA from whole-cell extracts (50 ng) was ligated to a 1:50 dilution of Illumina Adaptor Oligo Mix assigning 1 of 24 unique indexes in the kit to each sample. After ligation, libraries were amplified by 18 cycles of PCR using the HiFi NGS Library Amplification kit from KAPA Biosystems. Amplified libraries were then size selected using a 2% gel cassette in the Pippin Prep system from Sage Science set to capture fragments between 200 and 400 bp. Libraries were quantified by qPCR using the KAPA Biosystems Illumina Library Quantification kit according to the kit protocols. Libraries with distinct TruSeq indexes were multiplexed by mixing at equimolar ratios and running together in a lane on the Illumina HiSeq 2000 for 40 bases in single-read mode. Alignment to mouse genome assembly NCBI37/mm9 and normalization were performed as described previously[54]. Regions of modified histones enriched in wild-type and Ts1Rhr cells were identified using a MACS peak calling algorithm at a P value of 1×10^{-9} (ref. 55). Location analysis of ChIP-target enriched regions was performed using the CEAS software suite developed by the Liu lab at DFCI[56]. Promoter states were classified by the presence of H3K4me3, H3K27me3 or both (bivalent) ChIP-seq enriched regions in the ±1 kb region relative to the transcriptional start site (TSS). ChIP-qPCR was performed on two independent sets of pooled B cell colonies from three wild-type and three Ts1Rhr mice (primers available upon request). For analysis of upregulated genes in Ts1Rhr B cells, we excluded the 31 triplicated genes in Ts1Rhr mice. Data are presented as boxplots designating the median (black line), 1 s.d. (box) and 2 s.d. (whiskers). E2A ChIP-Seq data from *Rag1*$^{-/-}$ pro-B cells were obtained from GEO (GSE21978)[27] and mapped to the genome as described above. We defined regions of enriched E2A genomic occupancy using the MACS algorithm as described above. Genes were considered to be associated with E2A if their gene body overlapped an E2A enriched region or if their TSS was within 50 kb of an E2A enriched region, as was performed previously[57].

Statistical analyses. Pairwise comparisons are represented as means ± s.e.m. by two-tailed *t*-test, except where otherwise specified. Categorical variables were compared using a Fisher's exact test. Kaplan-Meier survival curves were compared using the log-rank test.

32. Urbánek, P., Wang, Z.Q., Fetka, I., Wagner, E.F. & Busslinger, M. Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* **79**, 901–912 (1994).
33. Dildrop, R. *et al.* IgH enhancer-mediated deregulation of N-myc gene expression in transgenic mice: generation of lymphoid neoplasias that lack c-myc expression. *EMBO J.* **8**, 1121–1128 (1989).
34. Abuhatzira, L., Shamir, A., Schones, D.E., Schaffer, A.A. & Bustin, M. The chromatin-binding protein HMGN1 regulates the expression of methyl CpG-binding protein 2 (MECP2) and affects the behavior of mice. *J. Biol. Chem.* **286**, 42051–42062 (2011).
35. Birger, Y. *et al.* Chromosomal protein HMGN1 enhances the rate of DNA repair in chromatin. *EMBO J.* **22**, 1665–1675 (2003).
36. Bustin, M., Crippa, M.P. & Pash, J.M. Immunochemical analysis of the exposure of high mobility group protein 14 and 17 surfaces in chromatin. *J. Biol. Chem.* **265**, 20077–20080 (1990).
37. Trageser, D. *et al.* Pre-B cell receptor–mediated cell cycle arrest in Philadelphia chromosome–positive acute lymphoblastic leukemia requires IKAROS function. *J. Exp. Med.* **206**, 1739–1753 (2009).
38. Tomizuka, K. *et al.* Functional expression and germline transmission of a human chromosome fragment in chimaeric mice. *Nat. Genet.* **16**, 133–143 (1997).
39. Fung, H. & Weinstock, D.M. Repair at single targeted DNA double-strand breaks in pluripotent and differentiated human cells. *PLoS ONE* **6**, e20514 (2011).
40. Weinstock, D.M., Nakanishi, K., Helgadottir, H.R. & Jasin, M. Assaying double-strand break repair pathway choice in mammalian cells using a targeted endonuclease or the RAG recombinase. *Methods Enzymol.* **409**, 524–540 (2006).
41. Ashton, J.M. *et al.* Gene sets identified with oncogene cooperativity analysis regulate *in vivo* growth and survival of leukemia stem cells. *Cell Stem Cell* **11**, 359–372 (2012).
42. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
43. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
44. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
45. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
46. Robinson, M.D. & Smyth, G.K. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**, 2881–2887 (2007).
47. Robinson, M.D. & Smyth, G.K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**, 321–332 (2008).
48. Irizarry, R.A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
49. Dai, M. *et al.* Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **33**, e175 (2005).
50. Reich, M. *et al.* GenePattern 2.0. *Nat. Genet.* **38**, 500–501 (2006).
51. Kinsella, R.J. *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011**, bar030 (2011).
52. Peach, S.E., Rudomin, E.L., Udeshi, N.D., Carr, S.A. & Jaffe, J.D. Quantitative assessment of chromatin immunoprecipitation grade antibodies directed against histone modifications reveals patterns of co-occurring marks on histone protein molecules. Molecular & cellular proteomics. *Mol. Cell. Proteomics* **11**, 128–137 (2012).
53. Verzi, M.P. *et al.* Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. *Dev. Cell* **19**, 713–726 (2010).
54. Lin, C.Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).
55. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
56. Shin, H., Liu, T., Manrai, A.K. & Liu, X.S. CEAS: *cis*-regulatory element annotation system. *Bioinformatics* **25**, 2605–2606 (2009).
57. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).

doi:10.1038/ng.2949