c-Myb Binding Sites in Haematopoietic Chromatin Landscapes

Mads Bengtsen[1], Kjetil Klepper[2], Sveinung Gundersen[3], Ignacio Cuervo[1], Finn Drabløs[2], Eivind Hovig[3,4,5], Geir Kjetil Sandve[3], Odd Stokke Gabrielsen[1]*, Ragnhild Eskeland[1]*

1 Department of Biosciences, University of Oslo, Oslo, Norway, 2 Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway, 3 Department of Informatics, University of Oslo, Oslo, Norway, 4 Department of Tumour Biology, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, 5 Department of Medical Informatics, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway

* Ragnhild.Eskeland@ibv.uio.no (RE); o.s.gabrielsen@ibv.uio.no (OSG)

Abstract

Strict control of tissue-specific gene expression plays a pivotal role during lineage commitment. The transcription factor c-Myb has an essential role in adult haematopoiesis and functions as an oncogene when rearranged in human cancers. Here we have exploited digital genomic footprinting analysis to obtain a global picture of c-Myb occupancy in the genome of six different haematopoietic cell-types. We have biologically validated several c-Myb footprints using c-Myb knockdown data, reporter assays and DamID analysis. We show that our predicted conserved c-Myb footprints are highly dependent on the haematopoietic cell type, but that there is a group of gene targets common to all cell-types analysed. Furthermore, we find that c-Myb footprints co-localise with active histone mark H3K4me3 and are significantly enriched at exons. We analysed co-localisation of c-Myb footprints with 104 chromatin regulatory factors in K562 cells, and identified nine proteins that are enriched together with c-Myb footprints on genes positively regulated by c-Myb and one protein enriched on negatively regulated genes. Our data suggest that c-Myb is a transcription factor with multifaceted target regulation depending on cell type.

Introduction

c-Myb is a key regulatory transcription factor (TF) essential for normal adult haematopoiesis [1–4]. It is a TF highly expressed in haematopoietic stem cells and progenitors, and plays a direct role in lineage commitment where its downregulation is associated with haematopoietic maturation and differentiation of both myeloid and B and T lymphoid progenitor cells [5–8]. Clinical studies have revealed strong links between c-Myb aberrations and human cancer. The *MYB* gene is frequently rearranged in several human neoplasias, such as acute myelogenous leukaemia, melanoma, and breast, colon and pancreatic carcinoma [9–11]. In some cancers this involves amplification of the *MYB* gene and increased c-Myb expression. The expression level of c-Myb is also tightly controlled by specific miRNAs [12,13]. A recent report identified a group of tumour suppressor miRNAs with reduced abundance in leukaemia cells from

patients with T-cell acute lymphoblastic leukaemia (T-ALL) [14]. Since these miRNAs all converged on *MYB*, their downregulation caused increased c-Myb expression in the T-ALL patients. On the other hand, studies of a knockdown allele of *Myb* in mice have shown that reduced levels of c-Myb can also severely perturb haematopoiesis [6–8,15]. The emerging picture from these studies is that the level of c-Myb is critical for proper function in haematopoietic tissue, and that only a two-fold up- or down-regulation may have dramatic biological effects. In order to understand the biological effects of altered c-Myb levels, it is important to know the c-Myb binding sites and target genes in haematopoiesis and cancer.

Although some studies have identified potential target genes by knockdown or induced expression of c-Myb [1,5,16–25], very few genome-wide studies of c-Myb enrichment are available. Chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-seq) relies on good antibodies and this is where c-Myb may have had limitations. A ChIP-seq dataset mapping c-Myb binding sites of an ER-MYB fusion protein in myeloid progenitor cells has been reported [5]. However, a severely truncated c-Myb part was immunoprecipitated lacking important functional regions, and we cannot exclude that c-Myb binding could be sterically influenced by the large ER part of the fusion [5]. ENCODE has published one c-Myb ChIP-seq dataset from murine MEL cells from the Snyder laboratory. However no published study of this dataset is available [9,14]. A recent paper reported c-Myb ChIP-seq datasets from MOLT-3 and Jurkat cells, but the authors limited their analysis to studying an oncogenic super-enhancer [26].

Antibody independent methods offer an alternative way of mapping binding of proteins to chromatin, such as DamID or chromatin accessibility analysis that maps DNA occluding factors with nucleases. DNase I footprinting has been used as a method to study DNA protection for over 35 years [27]. With recent developments in sequencing technology, mapping of nuclease-protected DNA can be used genome-widely at single base pair resolution. Digital genomic footprinting (DGF) uses massively parallel sequencing of DNase I treated cells to map proteins associated with specific DNA sequences on a global scale [28–32]. The identity of the factors bound is deduced from comparing the DNA sequence within the footprint with known sequence recognition patterns of different TFs.

In this work, we have exploited this alternative DGF strategy to obtain a global picture of c-Myb occupancy in the human genome. We have investigated c-Myb binding in six different haematopoietic cell-types using DGF and biologically validated the c-Myb footprints using c-Myb knockdown data, reporter assays and DamID analysis. We show that the predicted c-Myb specific binding sites vary strongly among haematopoietic cell-types, but that there is a set of c-Myb footprints that are common to all cell-types analysed. We identify c-Myb footprints for both up- and down-regulated targets in K562 cells c-Myb is a TF of critical importance for correct haematopoietic development and our predictions show that c-Myb has differential occupancy depending on cell type reflecting its role in both lineage commitment and differentiation.

Results

Genome-wide prediction of c-Myb footprints

DGF is a powerful method to identify nucleotides protected by proteins at a genome-wide scale independent of antibodies [29–32]. To map changes in c-Myb occupancy during haematopoiesis, we used DGF to generate maps of c-Myb footprints with nucleotide resolution (Fig 1A). We selected haematopoietic cell-types where c-Myb is expressed at different levels: c-Myb is highly expressed in haematopoietic stem cells [33] and expressed at lower level in CD4+ T-helper cells [34] and B cells [35,36]. c-Myb is also highly expressed in most cases of leukaemia
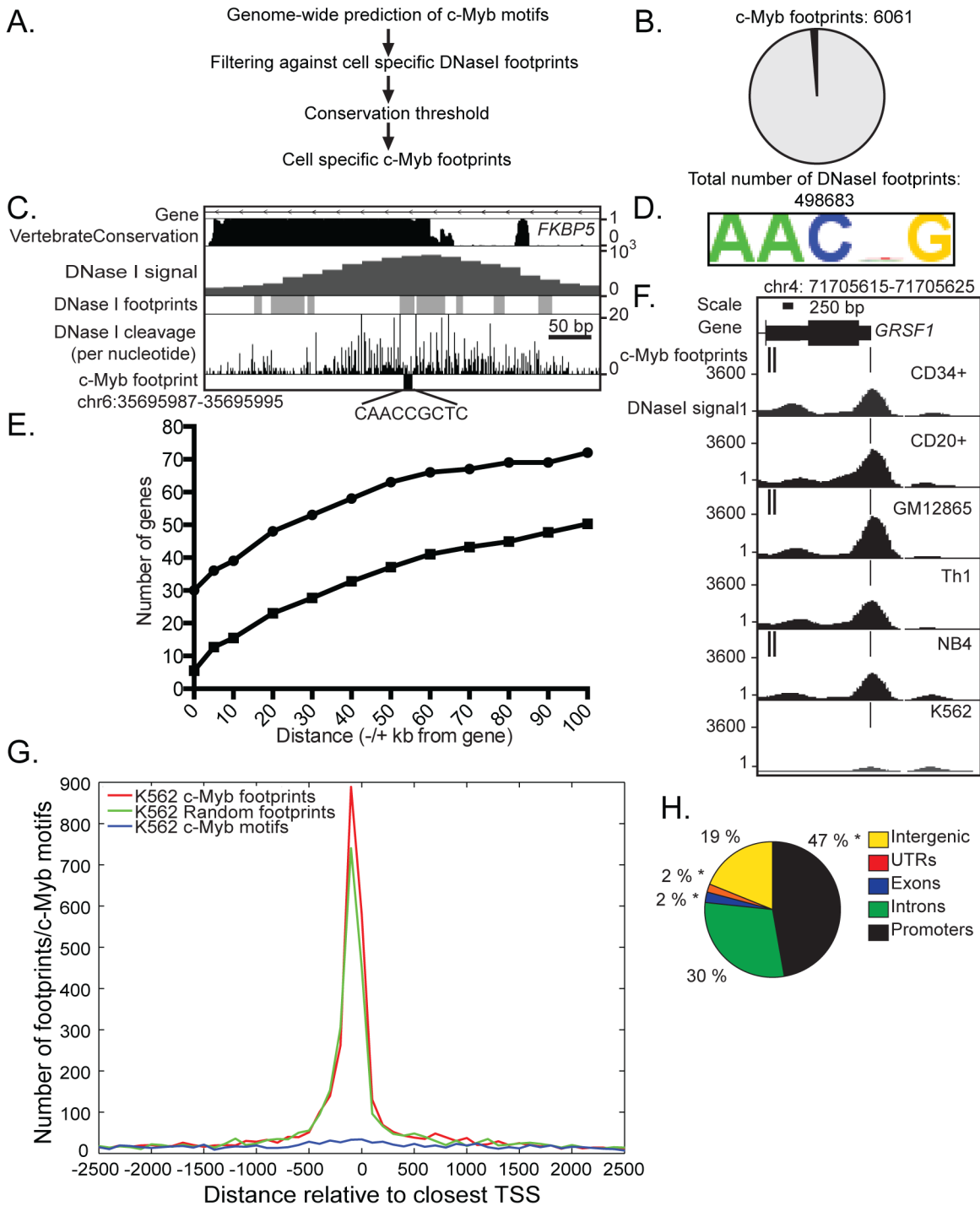
Fig 1. Identification of c-Myb footprints. (A) Workflow for identification of c-Myb footprints. (B) A pie chart representing the number of c-Myb footprints identified compared to the total number of footprints in K562 cells. (C) An illustration from Motiflab showing a c-Myb motif located at the start of the first intron of the gene *FKBP5* that overlaps with a DNase I footprint with a high conservation value in K562 cells, illustration modified. The coordinates for the c-Myb footprint is shown above the illustration, and to the right are the signal intensity for the DNase I datasets, in addition to a conservation score. (D) The binding motif enriched in c-Myb footprints in K562 cells. (E) Graph of the presence of c-Myb footprints and the distances to the 100 most regulated genes upon KD in K562 cells (dots) or a selection of 100 random genes (squares), an average of ten repetitions. Zero base pair indicates that the c-Myb footprints are found inside the gene body (F) A c-Myb footprint at the TSS of GRSF1 gene mapped in all six cell-types analysed. Coordinates for c-Myb footprint are shown above, and to the left are the signal intensity for DNase I datasets. (G) Position of c-Myb footprints, and random selections of DNase I footprints and c-Myb motifs, respectively, around ENSEMBL annotated TSS in K562 cells. (H) Distribution of c-Myb footprints at annotated genes, promoters and intergenic regions in K562 cells. *Overlapping significantly higher with c-Myb footprints than expected by random sampling of K562 DNase I footprints (p' < 5x10^{-2}).

[10]. We collected available DNase I footprint datasets in six different human cell-types from three healthy donors (CD34+ (mobilized), CD20+ and Th1 cells), transformed B-lymphocytes (GM12865) and two cancer cell-types where c-Myb is upregulated: erythroleukaemia (K562) and promyelocytic leukaemia (NB4) [31].

To predict potential c-Myb binding sites (c-Myb footprints), we first scanned the human genome with MotifLab [37] using four c-Myb motifs from the TRANSFAC database [38]. We identified more than 19 million c-Myb motif instances and filtered these against cell-specific DNase I footprints from the six different cell-types (Fig 1A) [31,39]. We decided that a c-Myb motif was regarded as occupied in each respective cell type if 90% of the motif overlapped a DNase I footprint. We found that between 0.14–0.3% of the total c-Myb motifs overlapped DNase I footprint signals in the six cell-types analysed (S1 Table).

It has previously been reported that factor specific DNase I footprints show a higher evolutionary conservation than immediately adjacent sequences and that these correspond with ChIP-seq signals [30,32,40]. We utilized information on weighted average conservation score (phastCons46wayPlacental) [41] to weigh each position in the footprint according to the information content of the corresponding column in the c-Myb motif. Sites that scored below 0.22 were discarded from further consideration. In total, we identified between 6061 and 12338 evolutionary conserved c-Myb footprints depending on the cell type (Fig 1B, S1A–S1F Fig and S1 Table). This is illustrated in Fig 1C where a c-Myb footprint in K562 cells fell within the first intron of c-Myb regulated *FKBP5* gene [1]and falls within an evolutionary conserved region. In all six cell-types, the weighted average conservation for each predicted motif instance are elevated for all genome-wide c-Myb footprints compared to all identified c-Myb motifs (S1A–S1F Fig).

We scanned the remaining 6061 c-Myb footprints in K562 cells with ChIPMunk [42] and identified a five nucleotide signature resembling the core c-Myb binding motif (Fig 1D). A similar c-Myb binding motif was identified in the other five cell-types (S1G–S1K Fig). This close resemblance of the five nucleotide signatures was expected as our analysis started with four c-Myb motifs from TRANSFAC database [38].

In order to evaluate the relevance of this collection of deduced c-Myb binding sites, we examined the correlation of the identified c-Myb footprints with a list of c-Myb target genes derived from c-Myb knockdown in K562 cells [1]. Seven of the ten most down-regulated genes (*KCNH2*, *LMO2*, *MYB*, *MYADM*, *STNM3*, *EPCAM* and *GRSF1*) had c-Myb footprints localized within the gene locus. For the gene GLUL, a c-Myb footprint was located 19 kilo bases (kb) downstream of the gene (S2 Table). Two target genes had no conserved c-Myb footprint present. Mapping c-Myb-footprints at the majority of these genes is consistent with c-Myb being involved in the activation of these. For genes being repressed by c-Myb, we identified c-Myb footprints in five of the ten most upregulated genes: within the gene locus for *GDF15*, *MKRN1*, *MRAP2*, *LEPR*, *CPEB4*. For two upregulated genes *SH3BGRL3* and *SLC30A10*, c-Myb footprints were identified 4 kb and 15 kb upstream respectively (S3 Table). The presence of conserved c-Myb footprints at a high fraction of gene loci that are most affected by c-Myb silencing suggests a role of c-Myb in direct regulation of these genes in K562 cells. We further extended this analysis to the 100 most up- or down-regulated genes upon c-Myb knockdown in K562 cells (Fig 1E) [1]. We find that 30% of these genes had conserved c-Myb footprints within the gene body. A total of 39% of the top 100 c-Myb target genes had a c-Myb footprint located +/- 10 kb from the gene body. Most cis-acting regulatory elements are found within 10–200 kb of their target genes [43]. By extending our analysis to +/-100 kb, we detected c-Myb footprints at 72% of the top 100 genes. The remaining 28% of genes had no c-Myb footprints and may not be direct targets of c-Myb, or these genes may be regulated by c-Myb at binding sites that are not conserved. Additional alternatives may be that c-Myb binds to a

DNA sequence motif different to the four TRANSFAC motifs used in this analysis, or indirect association of c-Myb with chromatin through interaction with another bond TF or co-factor. We also generated a graph displaying average of random sample of 100 genes repeated ten times which show a marked decrease in genes with c-Myb footprints (Fig 1E). For example, only 5.5% of these random genes had conserved c-Myb footprints within the gene locus, and 15.5% random genes had a c-Myb footprint located +/- 10 kb from the gene body.

We found that c-Myb footprints show a high degree of cell specificity, but there is also a common core of c-Myb footprints that could be detected in all six cell-types, suggesting that c-Myb may control both common functions and specific gene programs. One example is a c-Myb footprint that maps to the transcription start site (TSS) of the *GRSF1* gene in all six cell-types (Fig 1F). Nonetheless, two other c-Myb footprints in the first intron of *GRSF1* are only present in three cell-types (CD34+, GM12865 and NB4), suggesting a complex combination of general and cell type dependent control by c-Myb.

We analysed the global distribution of c-Myb footprints and found that between 10 and 15% (900–1300 footprints depending on cell type) map to the promoter directly upstream of TSS (Fig 1G and S2A–S2E Fig). In comparison, a random sample of the same number of predicted c-Myb motif hits in the respective cell-types showed far less preference for mapping close to the TSS. When we carried out the same analysis with the same number of randomly selected DNase I footprints, we found a similar TSS localization as with the c-Myb footprints, but with a slightly lower frequency directly upstream of TSS. Our analysis show that c-Myb footprints and randomly selected DNase I footprints follow a common pattern at TSS.

On a global level, we found that c-Myb footprints in K562 cells were located more in promoter regions (47%) and introns (30%) compared to intergenic regions (19%) (Fig 1H). When we compared this across the other cell-types, a large proportion of c-Myb footprints was present at promoters, with Th1 and CD20+ cells having over 60% of the c-Myb footprints located in these regions. In comparison, the percentages of c-Myb footprints at promoters were less (43–48%) in CD34+, GM12865, NB4, and K562 cells with more footprints in introns (28–31%) and intergenic sequences (18–21%) (Fig 1H and S2F Fig). However, when we compared our analysis with random sampling of DNase I footprints in K562 cells, c-Myb footprints overlapped significantly more with exons (with a normalised ratio, r, of 3.47), UTR regions ($r = 1.27$), and promoters ($r = 1.10$) than would be expected by random sampling of DNase I footprints (FDR-corrected p-value, $p' < 0.05$) (S4 Table). In the five other cell-types, c-Myb footprints were located significantly more in exon regions (with normalised ratios ranging from 2.98 to 3.96, $p' < 0.05$) (S4 Table) and 3'-UTR regions than expected by random sampling of DNase I footprints (with normalised ratios ranging from 1.10 to 1.44, $p' < 0.05$). For NB4 and GM12865 cells there was a slightly higher localization in promoter regions (normalised ratios 1.04 and 1.08, respectively, with $p' < 0.05$) (S2F Fig and S4 Table). Therefore, we conclude that c-Myb footprints differ from random sampling of DNase I footprints by locating more in exons than at promoters although the total number of c-Myb footprints in promoters is higher in all six cell-types.

Validation of predicted c-Myb footprints

To test whether a selection of c-Myb footprints is bound by c-Myb and in turn causes activation of the neighbouring gene, we performed transient reporter assays in CV-1 cells. We used a sumoylation-deficient c-Myb mutant (c-Myb-2KR) to ensure active c-Myb (Fig 2A) [1,44,45]. We selected nine regions containing c-Myb footprints that mapped to genes being activated by c-Myb (*KCNH2*, *LMO2*, *MYADM*, *GRSF1*, *IKZF1*, *SENP1*, *DUS3L*, *RABEPK* and *DCAF7*) (S2 and S5 Tables) in K562 cells [1]. Furthermore, we included four other K562 c-Myb footprints
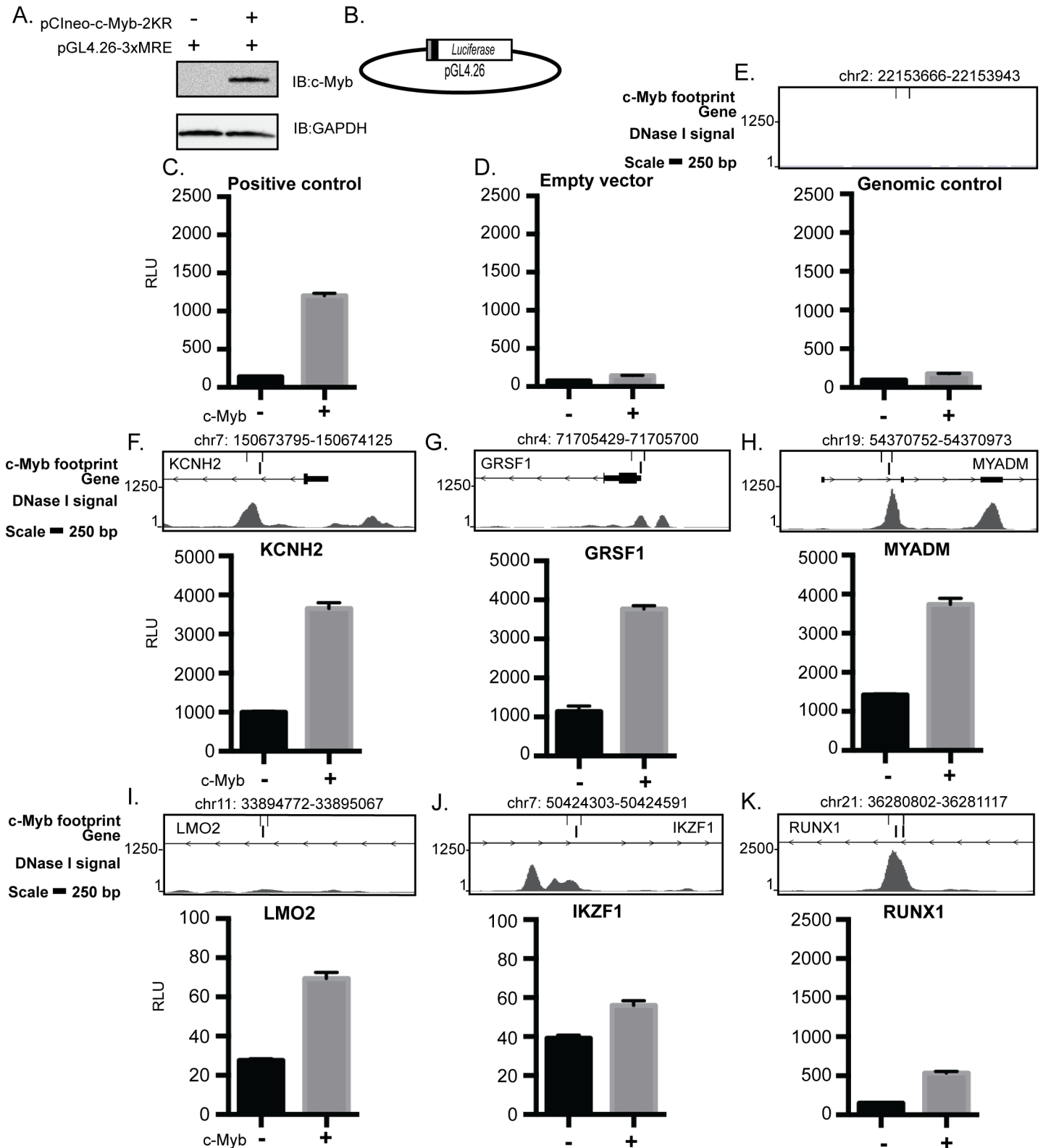
Fig 2. c-Myb enhances transcription from genomic elements with c-Myb footprints. Luciferase-based reporter assay to study the responsiveness of genomic regions, with one or more c-MYB footprints mapped. (A) Representative Immunoblot of CV1 cells transfected with reporter plasmid and-/+ c-Myb. (B) Map of the pGL4.26 vector used for the luciferase assays. The grey box illustrates the genomic fragment containing a c-Myb footprint or control region.

The black box illustrates the minimal promoter. (C) Positive control with three MREs. (D-E) Negative controls, (F-K) genomic loci identified to contain c-Myb footprints. The upper panels show the genomic region in the UCSC browser (hg19) presenting DNase I signals, c-Myb footprints and oligos for selected region. The coordinates for the c-Myb footprint are shown above the illustration. The values are the average from three independent experiments-/+ SEM.

located in proximity of or within the gene loci not known to be regulated by c-Myb in K562 cells (*RUNX1, RUNX2, KB-1458E12.1 and C10orf55*). Each amplified sequence (average 280 base pairs (bp)) spanning a c-Myb footprint was inserted into a luciferase reporter plasmid upstream of the minimal SV40-promoter (Fig 2B). As negative control we selected a genomic region on chromosome 2 that lacked c-Myb footprints. This control reporter showed only a marginal response similar to the empty vector (Fig 2D and 2E). Several of the selected regions (*KCNH2, MYADM, GRSF1, SENP1, RABEPK, DCAF7 and C10orf55*) showed a c-Myb response equal to or higher than the 3xMRE positive control (Fig 2F, 2G and 2H, S3A, S3B, S3D, S3E and S3G Fig). The base level differed largely between the constructs as expected since they span a larger segment than the just c-Myb footprint. A weaker response was measured for c-Myb footprints at the loci of *LMO2, IKFZ1, RUNX1, DUS3L, KB-1458E12.1* and *RUNX2* (Fig 2I, 2J, 2K and S3C, S3F and S3H Fig). These data confirmed that most of the selected c-Myb footprints, taken out of their normal context, confer c-Myb response consistent with c-Myb being capable of binding to the footprints and able to enhance transcription of the neighbouring gene (Fig 2 and S3 Fig).

In order to further validate the deduced c-Myb footprints, we performed a DamID analysis in K562 cells (S4A Fig) [46,47]. DNA adenine methyltransferase (Dam) was fused to full-length c-Myb, and we generated a pool of stably transfected cells that express trace amounts of Dam or c-Myb-Dam. It is critical to keep the Dam and Myb-Dam expression low to avoid too high background methylation. This precludes direct detection of the trace levels by normal Western blotting. We used an ecdysone-inducible promoter to detect the c-Myb-Dam expression and performed transient transfection together with the pVgRXR vector encoding the ecdysone receptor in K562 cells and induced expression by the ecdysone analog Ponasterone A [48]. A clear induction of the fusion protein was observed (Fig 3A). To rule out the effects of random integration of transgenes, we used two stable K562 pool cell lines for Dam and Dam-Myb derived at different time points. Finally, we used qPCR with oligos spanning selected c-Myb footprints to map c-Myb binding at these sites and compared the signals to those obtained with the Dam only cells.

To monitor c-Myb binding at c-Myb footprints in K562 cells, we monitored DamID signals by q-RT-PCR at six selected regions measured in the reporter assay (Fig 2), and in addition two controls and three other regions where we had detected c-Myb footprints. At two selected control loci without predicted c-Myb footprints we detected less c-Myb-Dam binding relative to Dam alone (Fig 2H and 2I). We detected c-Myb binding at five gene loci with c-Myb footprints that also showed response in the reporter assay (*KCNH2, LMO2, MYADM, GRSF1* and *RUNX1*) (Figs 2 and 3). Interestingly, we observed weak enrichment of the c-Myb footprint at the *IKZF1* locus, which showed only marginal response in the reporter assay (Figs 2J and 3F). We also detected binding of c-Myb-Dam over Dam alone at three other loci (*CBFA2T3, BHLHE40* and *PA2G4*) (S4B–S4D Fig). These results show that almost all loci with predicted c-Myb footprints that were tested by DamID are bound by c-Myb-Dam in K562 cells.

Histone modifications and transcription factors associated with c-Myb footprints

It has previously been reported that c-Myb acts as both a transcriptional activator and repressor and can influence the histone environment in the region it binds to [1,5,44,49]. To study
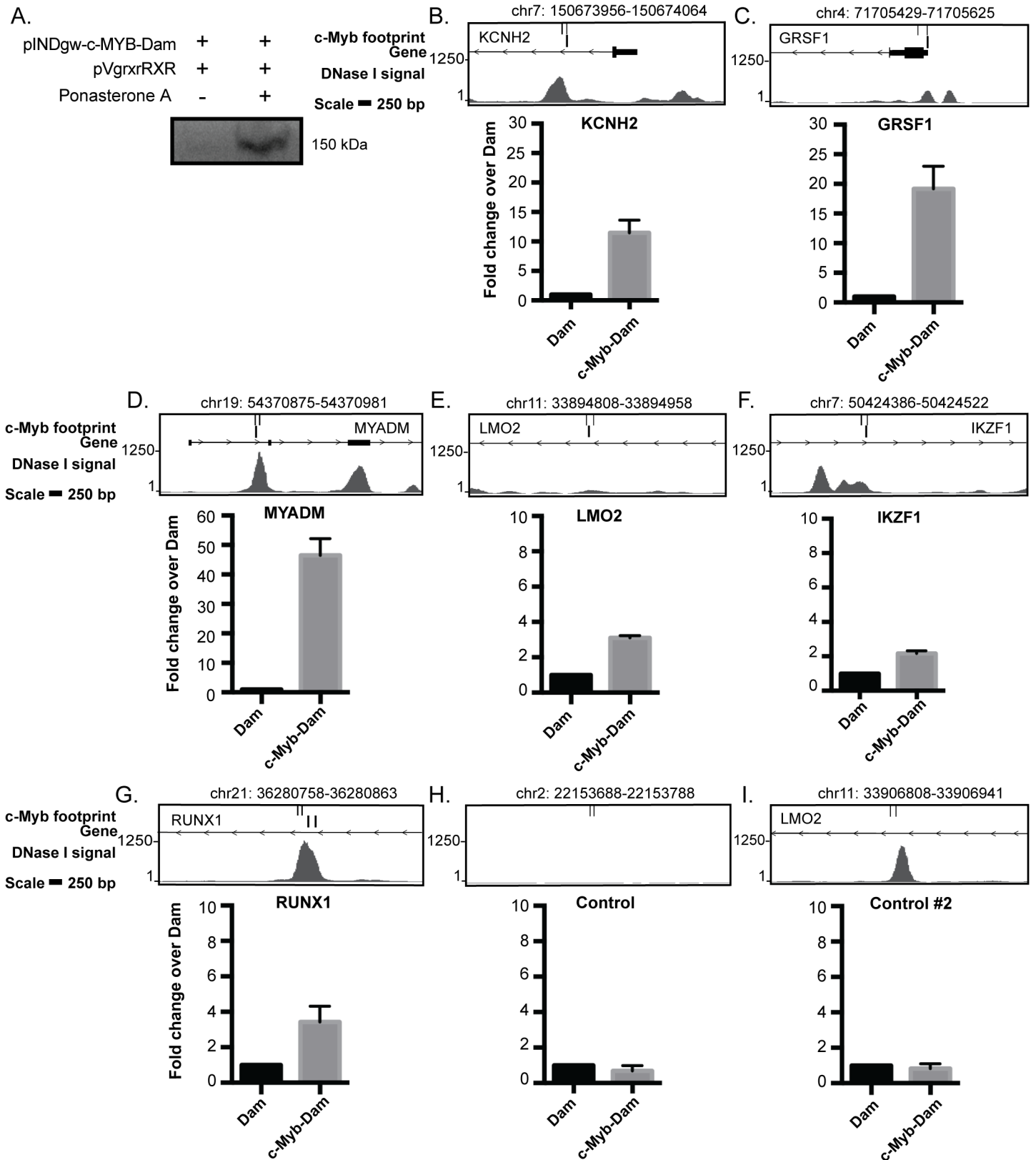
Fig 3. Validation of c-Myb footprints by DAMID. (A) Expression of the Flag-c-Myb-Dam construct. K562 cells were transfected with a plasmid encoding the c-Myb-Dam together with the activator plasmid pVgRXR and treated with 2 uM of Ponasterone A. Expression of the fusion construct was detected by

immunoblotting against Flag-tag. (B-J) Association of the control Dam and c-Myb-Dam at specific loci containing one or more c-Myb footprints quantified with qPCR and normalised to Dam. The upper panels show the genomic region in the UCSC browser (hg19) presenting DNase I signals, c-Myb footprints and oligos for qPCR. The coordinates for the c-Myb footprint are shown above the illustration. The values represent the average from two independent cell lines-/+ SEM.

doi:10.1371/journal.pone.0133280.g003

how c-Myb footprints and histone marks correlate on a genome-wide level, we compared the identified c-Myb footprints to ChIP-seq peak datasets for four different histone marks (H3K4me3, H3K4me1, H3K9ac and H3K27me3) in K562 cells, available from the ENCODE Consortium (Farnham and Snyder labs) [50,51] (Fig 4A–4D). We found that 36.9% of the c-Myb footprints in K562 cells overlapped with ChIP-seq peaks of H3K4me3, a mark generally associated with transcriptional initiation (Fig 4A) [52,53]. This overlap represents 10.7% of total H3K4me3 peaks (1863 of 18622 peaks). Similar enrichments were found for H3K4me1 and H3K9ac, both marks associated with "open" chromatin and being signatures of enhancers [54]. Here we found an overlap of 31.3% of the c-Myb footprints with H3K4me1 peaks (Fig 4B) and 40.6% overlap of the c-Myb footprints with H3K9ac (Fig 4C). Only 1.7% of total ChIP peaks for H3K4me1 overlapped with c-Myb footprints. The repressive mark H3K27me3 [55–57] showed a very low overlap with only 31 (0.02%) c-Myb footprints falling inside 134768 H3K27me3 peaks (Fig 4D).

We next tested whether the overlap between c-Myb footprints and histone marks were different than expected by chance. We found that DNase I containing c-Myb footprints overlapped significantly with H3K4me3 peaks (positively, with r = 1.10) and H3K4me1 peaks (negatively, r = 0.81) from what is expected from a null model based on random sampling of DNase I footprints (p' < 4x10^{-4}, Monte Carlo test) (S5A–S5D Fig and S6 Table). Furthermore, very few of a random sample of c-Myb motifs (same number as c-Myb footprints) overlapped with the different histone marks (S5E–S5H Fig). The general picture that emerges from this analysis is that c-Myb plays a role, both at enriched at TSS regions and exons, correlating with activating H3K4me3 marks. It also suggests that the repressive effects of bound c-Myb are achieved by other mechanisms than inducing repressive H3K27me3 marks.

The expression of a gene is often controlled by several TFs in concert through combinatorial control [58]. To obtain more information on how c-Myb exerts its function in synergy with other TFs in controlling gene expression of target genes, we analysed co-localisation of c-Myb footprints around the TSS and ChIP-seq peak datasets generated by the ENCODE Consortium [51] for 103 chromatin-associated proteins in K562 cells. We limited the analysis to the 467 genes positively or negatively regulated by c-Myb knockdown [1]. For each TF, we tested whether the ChIP-seq peaks overlapped c-Myb footprints around positively and negatively regulated genes, respectively, more than expected by random sampling of footprints. Based on certain thresholds (see Methods) we thus identified two sets of proteins that we suggest may co-regulate positively (9 factors) and negatively c-Myb regulated genes (1 factor), respectively (Fig 4E, S7 Table). Interestingly, c-Myb has previously been shown to interact with three of the proteins that we mapped to overlap on c-Myb target genes, either directly or as a part of complexes: a member of the mixed-lineage leukaemia (MLL) complex RBBP5 [49] and the two TFs ETS1 [59] and SIN3A [60]. Our analysis suggests that c-Myb may act together with these factors to modulate the expression of its target genes.

c-Myb footprints are present on a subset of genes across six haematopoietic cell-types

To understand how c-Myb exerts its function through downstream gene programs, we assigned molecular functions to the identified c-Myb footprints through the use of the Gene
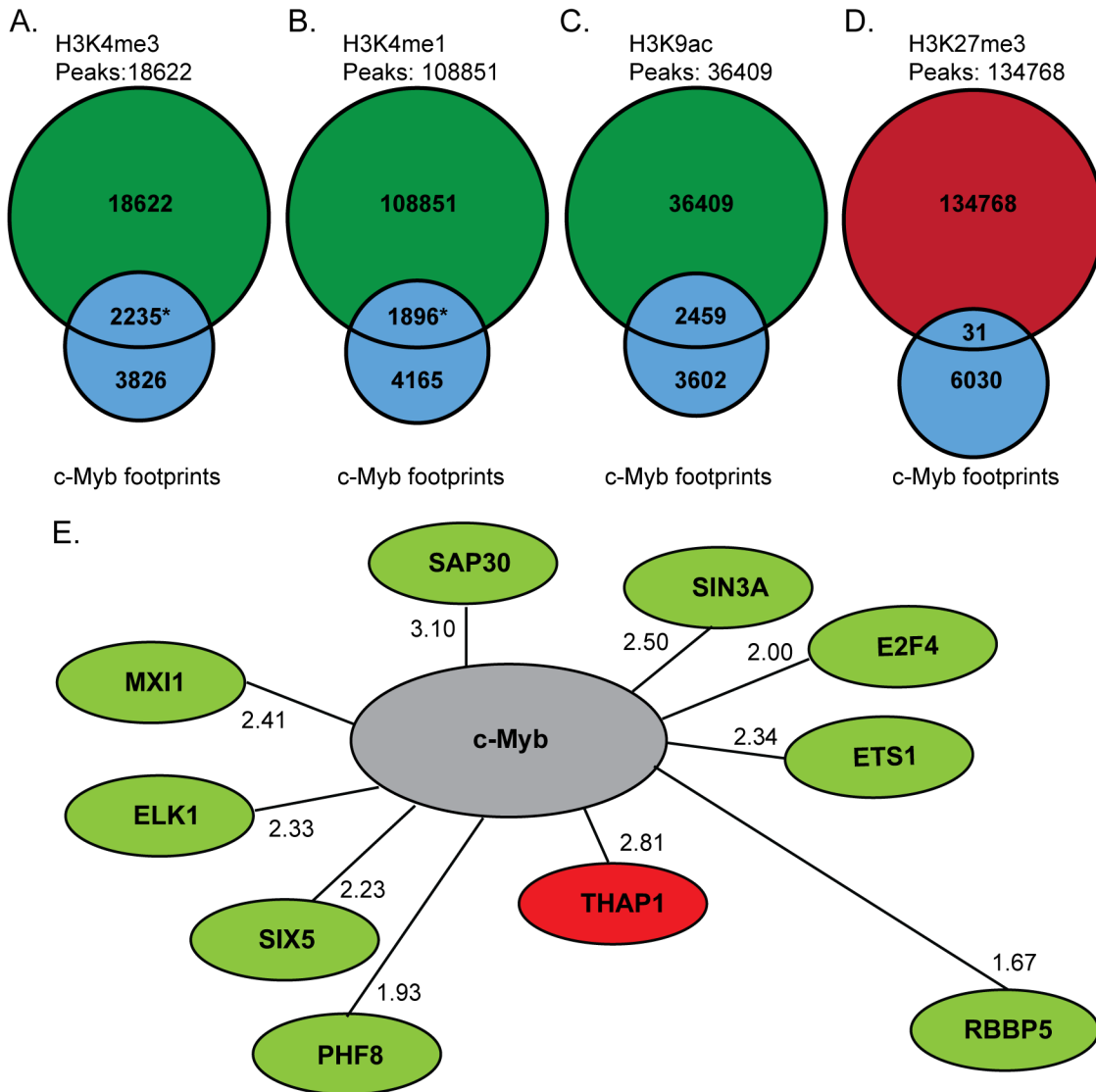
Fig 4. c-Myb footprints co-localises with active histone marks and co-regulatory TFs at c-Myb-regulated genes. (A-D) The number of c-Myb footprints (grey), which co-localize with ChIP-seq peaks for the active histone marks H3K4me3, H3K4me1, H3K9ac, (green) and the repressive mark H3K27me3 (red) in K562 cells. *Significantly different than expected from random sampling K562 DNase I footprints (p', > 4x10^{-4}). (E) Suggested co-regulatory TFs for c-Myb in K562 cells. Green and red denotes factors in the positive and negative set respectively. Next to each factor, the normalised ratio for the co-localisation with c-Myb footprints at positively or negatively regulated genes is displayed. The distance between each factor and c-Myb is calculated as a measure of the normalized ratio, as described in Methods.

doi:10.1371/journal.pone.0133280.g004

Ontology (GO) tool GREAT [61,62] (Fig 5, S6–S9 Figs). For K562 cells the top enriched functions were identified to be in three groups: RNA catabolic processes, regulation of gene expression and cell cycle regulation (S6 Fig). This result correlates well with previous conclusions after c-Myb knockdown in the same cell type [1]. The functional analysis of the five other cell-types showed genes involved in cellular maintenance and several cell-specific functions were enriched for each cell type (S6–S9 Figs). We repeated the analysis for the same number of randomly selected DNase I footprints in all six cell-types and obtained results showing different gene functions from those predicted from the c-Myb footprint gene list.
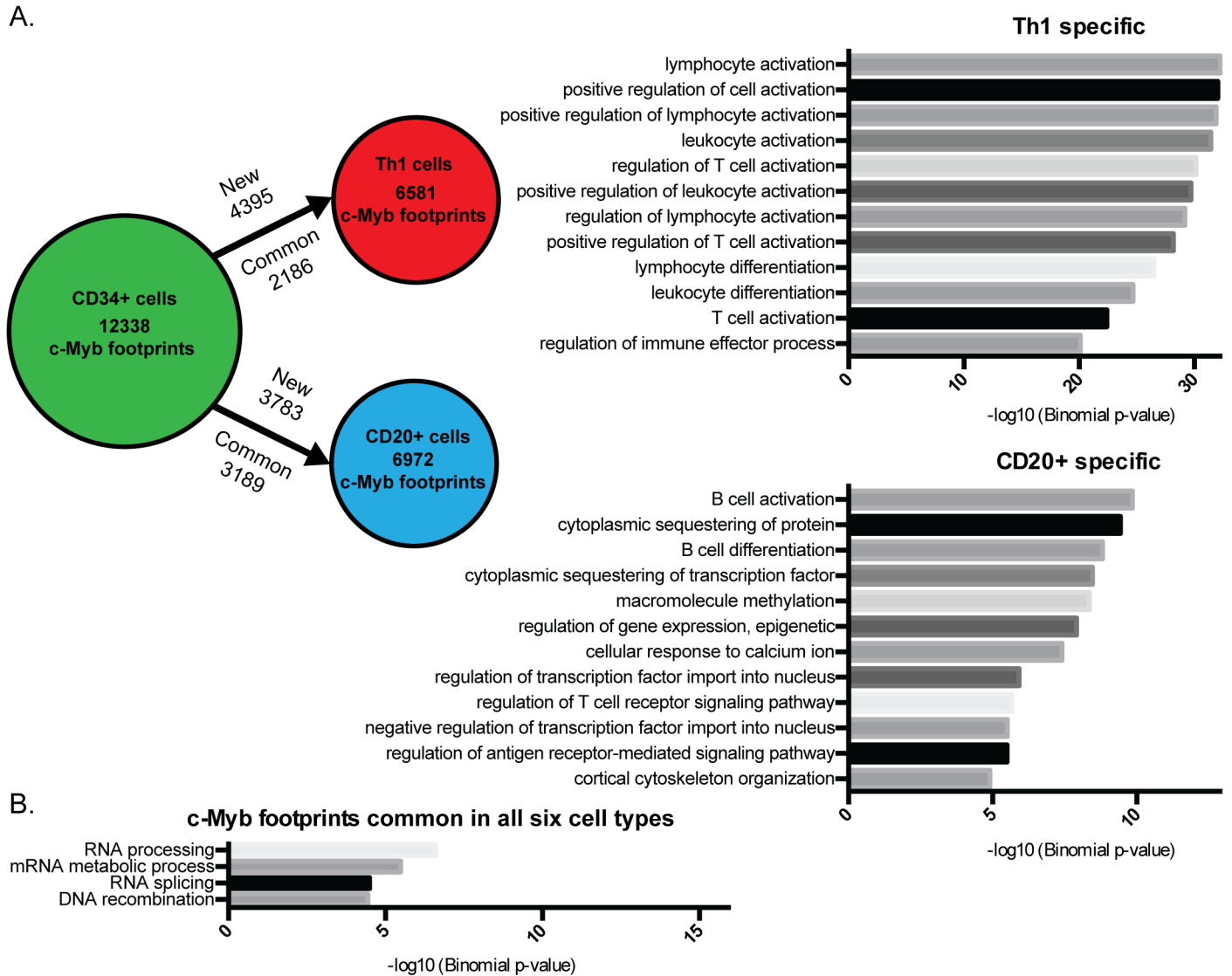
Fig 5. c-Myb controls differentiation and cell development. (A) Gain and loss of c-Myb footprints between CD34+ cells and CD20+ and Th1 cells, respectively. To the right are the top enriched functions for genes nearby c-Myb footprints specific for either CD20+ cells or Th1 cells as compared to CD34+ cells. For the full list of enriched functions, see S9 Fig. (B) Functional analysis of c-Myb footprints in all six cell-types. The functional analysis was made with GREAT [61].

To obtain more detailed information about the function of c-Myb in the different haematopoietic cells, we compared the c-Myb footprint genes from the haematopoietic progenitors CD34+ with c-Myb footprint genes from the more differentiated cell-types CD20+ and Th1 (Fig 5A). We found that a large number of c-Myb footprints are lost when the haematopoietic progenitors develop into each of the differentiated cell-types, while a small fraction of the c-Myb footprints is retained. However, an even larger fraction of the c-Myb footprints appear in this process and is specific for the differentiated cell type (Fig 5A). Functional analysis of the differentially mapped c-Myb footprint genes shows an enrichment of functions specific for the individual cell type, e.g. B cell activation and differentiation for CD20+ cells and T-cell activation and regulation for Th1 cells (Fig 5A).

A core of 406 common c-Myb footprints is present in all the six cell-types (e.g. *GRSF1*), and the functional analysis of this subset shows enrichment of genes involved in RNA processing and DNA recombination (Fig 5B, S10 Fig and S8 Table). We were concerned that these common c-Myb footprints could be driven by an overlap of DNase I footprints in all six cell-types and therefore we performed a random DNase I footprint control experiment ten times (S10 Fig). The random controls gave no common footprints, showing that there is a high degree of specificity for these common c- Myb footprints.

Four of the 65 common genes are listed among the genes regulated by c-Myb in K562 cells (*GRSF1*, *RUVBL2*, *UBE2N and SMNDC1*) (Figs 1G, 5B and S8 Table) [1]. Furthermore, when we analysed the list of common c-Myb footprints and compared overlap with ChIP-seq peaks for chromatin proteins that we identified as co-regulatory factors using c-Myb footprints in K562 cells (S7 and S9 Tables). A large fraction of the common c-Myb footprints (55–204) overlapped with ChIP-seq for the different factors.

To further validate our c-Myb footprints, we used the set of 406 common c-Myb footprints from our six cell-types and checked for overlap with c-Myb ChIP-seq peaks in human T-cell leukaemia cell lines (Jurkat and MOLT-3) [26]. The rationale is that if these footprints represent a common c-Myb signature, they should also be found among the c-Myb ChIP peaks in the two latter cell lines. We got an overlap of 65.2–75.8% in the Jurkat cell line and 79.6% in MOLT-3 the cell line. From this we can conclude that a large fraction of common c-Myb footprints from our analysis are also found in T-cell leukaemia cell lines. We illustrate an overlap of a common c-Myb footprint with the c-Myb ChIP-seq signal datasets at the GRSF1 promoter (S11A–S11B Fig).

Discussion

In this study we have predicted genome-wide c-Myb binding in six different cell-types using digital DNase I footprints, from the haematopoietic progenitor CD34+ to the more differentiated cell-types GM12865, CD20+ and Th1 and the cancerous cell-types K562 and NB4 (Fig 1) [31]. Our aim was to evaluate whether DGF was an approach that could compensate for the lack of available c-Myb ChIP-seq data. With the filters utilised, we ended up with about 6000 footprints sharing a c-Myb signature in K562 cells. Several validation experiments suggested that these predictions had a reasonable accuracy. We used our c-Myb knockdown dataset from K562 cells to validate the c-Myb footprint predictions. For the top 100 c-Myb regulated targets a large proportion (39%) had c-Myb footprints +/- 10 kb from TSS, whereas when we extended the analysis to +/-100 kb, we detected c-Myb footprints at 72% of top 100 genes. Furthermore, we used reporter assays and showed that thirteen selected c-Myb footprint regions that localized either within the gene locus or upstream of twelve genes were enhanced to different degrees in the presence of c-Myb compared to control (Fig 2 and S3 Fig). In addition to these functional assays, we directly tested c-Myb occupancy on a selection of c-Myb footprints in K562 cells with the antibody independent technique DamID and showed that they indeed are elements recruiting c-Myb in their chromatin context (Fig 3 and S4 Fig). It is noteworthy that the level of c-Myb-Dam expression is very low in DamID compared to the reporter assay, and we were unable to detect the c-Myb-Dam fusion protein by western in c-Myb-Dam stable cell lines. That we nevertheless find c-Myb enriched in nine out of nine selected regions with c-Myb footprints suggests that c-Myb recognizes and selectively binds these predicted footprints in chromatin under quite stringent conditions. The DamID validations, therefore, lend quite a strong support to the accuracy of the DGF predictions.

The vertebrate Myb family members consist of A-Myb (*MYBL1*), B-Myb (*MYBL2*) and c-Myb (*MYB*) and share a conserved DNA-binding domain [63]. Although the Myb family

members are very similar in overall structure and although they can be co-expressed in different cell-types, knockout studies of A-Myb, B-Myb or c-Myb show that they have differential roles in gene regulation during development and have distinct phenotypes [3,64,65]. The three MYB family members have their highest level of conservation in the DNA binding domain (DBD). They bind the same core Myb recognition element (MRE) (PyAACG/TG) [66–68] and the core MRE in c-Myb footprints in all six cell-types may therefore be bound by all three proteins (Fig 1D and S1G–S1K Fig). Our main focus has been on c-Myb footprints in K562 cells where c-Myb is the most highly expressed family member and overlapping binding of A-Myb to MREs is minimal, as *MYBL1* mRNA is approximately 900 times less abundant [1]. The *MYBL2* expression is four times lower than the expression of *MYB* in K562 cells and it is therefore a more likely candidate binder than *MYBL1* [1].While A- and c-Myb appear to have virtually identical DNA-binding properties, B-Myb forms complexes of significantly lower stability, which are rapidly dissociating under competitive conditions. It is therefore unlikely that B-Myb can form sufficiently stable enough complexes to generate clear DNase I footprints [69].

Another important aspect regarding prediction of specific TF footprint signatures is the residence time of the respective factor. A recent report by Hager and colleagues showed that DNase I "cleavage" signatures to a large extent depend on intrinsic properties of the DNase I and the DNA sequence in the factor-binding site [70]. However, the footprint depth seems to depend on the time the factor occupies and protects the target sequence. Many TFs with fast kinetics such as the glucocorticoid receptor (GR) gives poor overlap between GR footprints and ChIP-seq peaks compared to CTCF that has long residency time [70]. The *in vivo* dynamics of c-Myb binding is not known, but the intrinsic DNA binding properties of c-Myb has been extensively studied *in vitro*. Noteworthy, c-Myb binds to DNA in a two-step process— first the rapid formation of an unstable complex, followed by a slower transition to a stable complex, a process coupled with a conformational change in its DBD [71,72]. Therefore, c-Myb is expected to be able to bind more stably to chromatin than normal "tread milling" TFs. How this process is dependent on the DNA sequence in the factor-binding site remains to be elucidated.

Several methods for prediction of TF binding using DGF have been described in different cell-types from yeast to human [29–32,73–75]. Different computational prediction tools such as Wellington [75], CENTIPEDE [32], DNase2TF [70] and Footprint detection software [30] are available. We have devised an approach that uses DGF datasets from [31], in combination with MotifLab [37] and four c-Myb motifs from the TRANSFAC database [38] and weighted conservation using mammalian phastCons elements [41]. Our choice of conservation can be debated as regulatory elements may not necessarily be conserved across mammalian species [76]. A recent study showed that only about 22% of mouse TF footprints are conserved in human [77]. Even though several approaches have successfully identified active conserved regulatory regions across vertebrate species [78–81], many enhancers are poorly conserved and have species-specific TF binding [82,83]. Therefore, we cannot exclude the possibility that our filters will to a certain degree underestimate c-Myb binding sites in the six human cell-types. A recent report of an oncogenic super-enhancer formed by somatic mutation creating a novel c-Myb binding site shows that non-conserved enhancers can occur independently of evolution [26]. Our analysis, therefore, limits the prediction of c-Myb footprints to those that are evolutionarily conserved, and we may miss c-Myb regulatory elements only present in humans. We do, however, identify substantially more c-Myb footprints in our analysis as compared to the previously identified Myb footprints in seven lymphoblastoid cell lines [29].

Given these reservations, on a global level, our data show that c-Myb footprints differ from random sampling of DNase I footprints by locating more in exons than at promoters although

total number of c-Myb footprints in promoters is higher in all cell-types. An estimate of 51% of all enhancers are intragenic [54] and DNaseI HS sites in exons have been implicated in chromatin looping and possibly alternative splicing [84].The presence of c-Myb in exons and a role in such processes is very interesting and needs to be further characterized in future studies.

We identified factors that co-localize with c-Myb footprints at promoters of c-Myb regulated genes in K562 cells [1] (Fig 4E). Three of the co-regulatory proteins (RBBP5, ETS1 and SIN3A) have been found to interact directly or indirectly with c-Myb [49,59,60]. SIN3A, SAP30 and RBBP5 are part of the ALL-1 super complex identified in K562 cells [85]. This complex also contains two other known c-Myb co-factors p300 [44] and CHD3 [86] that are involved in the regulation of c-Myb activity. Both p300 and CHD3 enhance c-Myb activity, and may function together with the SIN3A/SAP30/RBBP5 and c-Myb. RBBP5 is also part of the MLL1/2 complex responsible for H3K4me3 [87] and MLL3/4 was recently described as the methyltransferases that monomethylates H3K4 [88]. We find that one-third of c-Myb footprints overlapped with H3K4me3, and that this overlap was statistically significantly different than expected by random sampling of DNase I footprints (Fig 4A). MLL1 interacts with c-Myb through Menin [49] and, therefore, c-Myb may play a role in directing MLL mediated H3K4 trimethylation to c-Myb target genes.

Besides a small core of c-Myb footprints that are common across cell-types (total 406) (Fig 5B and S10 Fig), our analysis shows that a large part of c-Myb binding sites are cell type specific. Performing GREAT for the c-Myb footprints indicates that c-Myb has specialized roles related to the function of the specific cell type (Fig 5A and S6–S9 Figs).

The gene *GRSF1* is an important mitochondrial regulator and is one of the most affected genes upon c-Myb knockdown in K562 cells (S2 Table). Interestingly, our analysis identifies a c-Myb footprint in the promoter region of *GRSF1* present in all six cell-types. Moreover, we show that c-Myb is capable of enhancing the expression of *GRSF1* and also binds to the locus in K562 cells. Taken together, the data indicates that c-Myb is important for the expression of the *GRSF1* gene in several stages of the haematopoiesis.

We used this dataset of common c-Myb footprints and found extensive overlap with c-Myb ChIP-seq peaks in Jurkat and MOLT-3 cells, with the rationale is that if these footprints represent a common c-Myb signature, they should also be found among the c-Myb ChIP-peaks in the T-cell leukaemia cell lines. This was indeed true, we found a marked overlap that indicates that the common c-Myb footprints are bound by c-Myb, and may function as a type of quality control of our footprint predictions.

In summary, our data show that DGF can be used to predict conserved functional binding sites for c-Myb and that c-Myb has specific binding sites depending on the haematopoietic cell type. We have compared the majority of our analysis results to a random control. Furthermore, we have validated a selection of predicted c-Myb footprints by two different methods, and we found that c-Myb was capable of binding and enhancing gene activity through these predicted elements. We also mapped predicted c-Myb footprints to top c-Myb regulated target genes in K562 cells. These results suggest that a compelling fraction of our identified c-Myb footprints indeed are true c-Myb binding sites.

Materials and Methods

Data source

Digital genomic footprints for the six cell-types: CD20+, CD34+ (mobilized), GM12865, K562, NB4 and Th1 were obtained from [31]. ChIP-seq peaks for factors in K562 generated from experiments as part of the ENCODE Consortium [51] were downloaded from the UCSC Table Browser (S12 Table). For the histone analysis, we used ChIP-seq peaks generated by the

Farnham and Snyder labs (S12 Table). For gene annotation data, we used ENSEMBL annotation GRCh37 [89].

Cell culture

Human K562 cells and African green monkey CV1 cells were obtained from ATCC and cultured as described in [86].

Constructs and Cloning

For luciferase constructs, genomic DNA was extracted from K562 cells using the DNeasy Blood & Tissue Kit (Qiagen). Selected genomic regions with the approximate size of 280 bp were amplified by PCR and cloned into the pGL.24.6 (Promega) vector using the restriction sites XhoI and NheI. For primers used, see S10 Table. To obtain the fusion construct 3xFLAG-c-Myb-V5-EcoDam, the c-Myb with an N-terminal 3xFLAG-tag was cloned into the pINDgw-RFA-V5-EcoDam using the Gateway technology (Invitrogen). The pINDgw RFA-V5-EcoDam, pIND-V5-EcoDam and pVgRXR vectors were a kind gift from Bas van Steensel [47]. c-Myb2KR is described in detail in [44]

Reporter assay

The day before transfection, CV-1 cells were plated in 24 micro-well plates at $2x10^4$ cells per well. Cells were transfected with a total of 0.3 micrograms of DNA per well using the TransIT-LT transfection reagent (Mirus Bio). For the reporter assay 0.2 micrograms of pCIneo-c-Myb-2KR [1] and 0.1 micrograms of the pGL.4.26 were used per well. Cells were lysed 18 hours after post transfections with Passive lysis buffer (Promega) and luciferase activity was measured in a luminometer (Turner Designs). Data from three biological and nine independent transfections are presented.

DNA adenine methyltransferase identification (DamID) assay

Stable K562 cell lines expressing either 3xFLAG-c-Myb-V5-EcoDam or EcoDam alone were generated by electroporation using the Amaxa Nucleofector system (Lonza Bioscience) with the pINDgw-3xFLAG-c-Myb-V5-EcoDam or pIND-V5-EcoDam, respectively. Following electroporation cell lines were selected with G418 (Invivogen). DamID libraries for EcoDam and c-Myb-V5-EcoDam were made as described in [47]. In brief: Genomic DNA was isolated using the DNeasy Blood & Tissue Kit (Qiagen) and processed to enrich for DNA methylated by either V5-EcoDam alone or 3xFLAG-c-Myb-V5-EcoDam. Purified DNA was analysed by qPCR using the same amount of DNA for EcoDam and c-Myb-V5-EcoDam [48] on a Lightcycler480 (Roche). For primers used, see S10 Table. To validate the expression of the full-length 3xFLAG-c-Myb-V5-EcoDam construct, K562 cells were transfected with pINDgw-3xflg-c-Myb-V5-EcoDam or pIND-V5-EcoDam respectively together with the pVgRXR ecdysone receptor-encoding vector. Next day, 2 μM of Ponasterone A (Invitrogen) was added to the cell media and after 24 hours cells were lysed in SDS loading dye and subjected to western blotting on a PVDF membrane with anti-FLAG (Sigma) and anti-GAPDH (Invitrogen) antibodies (S10 Table).

Identification of c-Myb footprints

To predict c-Myb footprints in the human genome (hg19), we used the MotifLab analysis workbench with MATCH motif scanning tool and *minSUM* cut-off threshold [37,90]. We scanned with four c-Myb motif models (M00004, M00183, M00773, M00913) from the

TRANSFAC database [38]. The overlap between the c-Myb motif instances and DNase I footprints was calculated for each of the six cell-types with a threshold of 0.9 (CD20+, CD34+ (mobilized), GM12865, K562, NB4 and Th1) [31]. For each predicted motif instance we calculated a weighted average conservation score across the site where the conservation score (phastCons46wayPlacental) (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/phastCons46way/placentalMammals/) in each position was weighted according to the information content of the corresponding column in the motif. Sites that scored below 0.22 were discarded from further consideration. The *de novo* search for motifs inside the c-Myb footprints was carried out with the motif identification tool ChIPMunk [42].

c-Myb regulated gene set

The list of 100 most up- or down-regulated genes upon c-Myb knockdown in K562 cells are obtained from [1]. In brief, we analyzed the global effects of c-Myb knockdown using microarray expression profiling by comparing genome wide patterns of gene expression between control and c-Myb-siRNA transfected K562 cells. The control K562 cells were transfected with a non-specific siRNA (siLuc; targeting the firefly luciferase gene). We performed a first profiling experiment using eight biological replicates and si323 RNA-mediated knockdown. A second expression profiling study with the si2992 RNA-mediated knockdown and four biological replicates was used to validate the regulated genes detected in the first dataset. After statistical analysis of the results from each of the experiments using permutation F2-tests, in which residuals were shuffled 5000 times, and family-wise error correction, and top 100 significantly regulated genes ($P < 0.05$) were selected.

Statistical analysis

For statistical analysis we used The Genomic HyperBrowser [91]. Hypothesis testing was performed using Monte Carlo simulation with 10000 repetitions, drawing random samples (of the same size as the number of c-Myb footprints) uniformly from the total population of DNase I footprints. As the test statistic, the difference in the overlap between the dataset in question and respectively the sampled footprints (case) and the rest (control) was used. The p-values were corrected for multiple testing using FDR correction over all tests, or in the case of the analysis of the cell-specific distribution of c-Myb footprints, over all tests per cell type [92]. As a measure of effect size, a normalised overlap ratio was used, defined as follows:

$$r = (Xm/Yn)$$

where X is the overlap between the query dataset and c-Myb footprints, Y is overlap between the query dataset and the rest of the DNase I footprints, n is the number of c-Myb footprints and m is the number of remaining DNase I footprints. For these analyses the middle point of the DNase I footprints were used.

Analysis of TF co-regulation

For the analysis of TF co-regulation, distance from c-Myb footprints to the closest gene regulated by c-Myb [1] was assigned using BEDOPS [93]. Footprints inside +/- 5 kb of TSS of regulated genes were isolated and compared with ChIP-seq datasets. Several thresholds were set: first, only factors with peaks overlapping the gene-regulating c-Myb footprints significantly more than expected by random sampling of DNase I footprints, were selected (FDR-corrected p-value, p', < 0.05); second, the threshold for normalised ratio was set to 1.05; third, there needed to be at least 20 genes with c-Myb regulating footprints (both positively and negatively) overlapping the ChIP-seq peaks of the factor; and fourth, the difference in normalised ratio for

the overlap between the peaks and the positively and negatively regulated genes, respectively, needed to be > 0.5. Factors thus selected were then assigned to either a positive or a negative set of co-regulating TFs according to the highest value of the normalised ratio. A distance measure between c-Myb and each protein was calculated as:

$$10^{\frac{a}{b}}$$

Where a is the highest normalised ratio for the factors in the set, and b is the normalised ratio of the factor in question.

Distribution of c-Myb footprints

To calculate the genomic distribution of the c-Myb footprints, Ensembl gene annotations were used. The annotations were divided into the following categories: promoters, exons, 3´-UTR, introns, and intergenic regions. The promoter regions were defined as -2500 bp upstream and 500 bp downstream of TSS. In cases where a footprint was found in more than one gene category, it was assigned to one category in the following order: promoters, exons, 3´-UTR, introns, and intergenic regions. For the distribution around TSS, c-Myb footprints, DNase I footprints and c-Myb motifs were grouped into 100 bp bins and summed. For all analyses, including histone marks and distance to regulated genes, the middle point of the DNase I footprints and motifs were used.

GREAT analysis

For the functional analysis of c-Myb footprints the GREAT tool with standard settings was used [61]. The middle point of either c-Myb footprints or a random selection containing the same number of cell- specific DNase I footprints was used as input. For a comparison of c-Myb specific footprints between cell-types, the middle point of c-Myb specific footprints were expanded with 12 bp on each side and an overlap between two footprints was set to require at least six bp. The promoter regions of the gene lists are defined as -2.5 kb upstream to +0.5 kb downstream of the TSS.

Analysis of c-Myb ChIP-seq data

For analysis of c-Myb ChIP-Seq data from [26], datasets were retrieved from NCBI Gene Expression Omnibus (GEO) (GSM1519643 and GSM1442006) and analysed with SraTailor [94] using the programs standard settings for Bowtie2 [95] and MACS [96]. ChIP-seq datasets for c-Myb were analysed for enrichment with corresponding control datasets. To calculate the fraction of common footprints in all six cell-types co-localising with ChIP-Seq peaks for c-Myb in Jurkat and MOLT-3 cells, the overlap between footprint and peaks was set to be a minimum of one bp.

Supporting Information

S1 Fig. Weighted average conservation for genome-wide c-Myb motifs and c-Myb footprints. (A-F) Weighted average conservation using mammalian phastCons elements for each predicted motif instance for all genome-wide were calculated for c-Myb motifs and the identified c-Myb footprints +/- SD, respectively, in the cell-types CD20+, CD34+, GM12865, K562, NB4, Th1. (G-I) The binding motif enriched in c-Myb footprints in CD20+, CD34+, GM12865, NB4 and Th1 cells.
(TIF)

S2 Fig. Genomic distribution of c-Myb footprints. (A-E) Localization of c-Myb footprints, a random selection of DNase I footprints and a random selection of c-Myb motifs for the cell-types CD20+, CD34+, GM12865, NB4 and Th1 around TSS. (F) Genomic distribution of c-Myb footprints for the cell-types CD20+, CD34+, GM12865, NB4 and Th1. *Overlapping significantly more with c-Myb footprints than with randomly selected K562 DNase I footprints (p' < 0.05, calculated by the Monte Carlo test).
(TIF)

S3 Fig. Additional luciferase assays. (A-H) Luciferase assay as described in Fig 2.
(TIF)

S4 Fig. Additional DamID assays. (A) Schematic overview of the DamID method. (B-D) DamID assay for the association of the control Dam and c-Myb-Dam as described in Fig 3.
(TIF)

S5 Fig. Co-localisation of DNase I and c-Myb motifs with histone marks. (A-H) Overlap between ChIP-seq peaks for the active histone marks H3K4me3, H3K4me1, H3K9ac (green) and the repressive mark H3K27me3 (red) in K562 cells, and K562 DNase I footprints or a random sample of c-Myb motifs. For DNase I footprints, the expected number of overlapping footprints when drawing random samples without replacement from the total set of K562 DNase I footprints (the hypergeometric distribution) are shown. For c-Myb motifs the overlaps of a single random sample are shown.
(TIF)

S6 Fig. Functional analysis of c-Myb footprints in K562 cells and CD20+ cells. GREAT GO-term annotations for c-Myb footprints and a random sample of DNase I footprints for K562 cells (A-B) and CD20+ cells (C-D).
(TIF)

S7 Fig. Functional analysis of c-Myb footprints in CD34+ cells and GM12865 cells. GREAT GO-term annotations for c-Myb footprints and a random sample of DNase I footprints for CD34+ cells (A-B) and GM12865 cells (C-D).
(TIF)

S8 Fig. Functional analysis of c-Myb footprints in NB4 cells and Th1 cells. GREAT GO-term annotations for c-Myb footprints and a random sample of DNase I footprints for NB4 cells (A-B) and Th1 cells (C-D).
(TIF)

S9 Fig. Functional analysis of cell-specific c-Myb footprints in CD20+ cells and Th1 cells. The full list of enriched functions identified with GREAT for cell specific c-Myb footprints for CD20+ cells (A) and Th1 cells (B) as compared to CD34+ cells.
(TIF)

S10 Fig. Analysis of overlap of c-Myb footprints in the six cell-types compared to random DNase I footprint controls. Graphs showing number of common c-Myb footprints or random selections of cell-specific DNase I footprints after subtraction of non-overlapping footprints between two cell-types at the time, and ending with the final number is a common set of footprints in all six cell-types. The analysis of a random selection of cell-specific DNase I footprints was repeated ten times starting with 12338 random footprints in CD34+ cells. The y-axis represents the number of c-Myb or DNase I footprints; the x-axis shows the six cell-types with total number of c-Myb footprints or number of random selection of cell-specific DNase I footprint used in the analysis (c-Myb footprints, red graph; random DNase I footprints, black bars). The

numbers to the right indicate common footprints for c-Myb (red) or a random selection of cell-specific DNase I footprints (black) footprints common in all the cell-types.
(TIF)

S11 Fig. Overlap of common c-Myb footprints and c-Myb ChIP-Seq data from Jurkat and MOLT-3 cells. A) Overlap between c-Myb ChIP-Seq peaks for Jurkat and MOLT-3 cells [26] and the c-Myb footprints common in all the six cell-types analysed in this study. ChIP-Seq data was processed with SraTailor [94] using the default settings. B) An illustration showing the identified c-Myb common footprints at the promoter for GRSF1 for the six cell-types analysed in this study (see also Fig 1F) and enriched c-Myb ChIP-Seq signals for the same region in Jurkat and MOLT-3 cells. Coordinates for c-Myb footprint are shown above, and to the left are the signal intensities for the ChIP-Seq data shown. UCSC version hg19 (http://genome.ucsc.edu).
(TIFF)

S1 Table. DNase I footprints and c-Myb footprints for the six cells types analysed. The total number of footprints, footprints overlapping with c-Myb motifs and predicted c-Myb footprints in all the six cell-types analysed.
(PDF)

S2 Table. The ten most downregulated genes in K562 cells upon knockdown of c-Myb. Gene name, ID number, degree of regulation, if the gene contains a c-Myb footprint and the position of the footprint for the ten most downregulated genes in K562 cells upon knockdown of c-Myb [1].
(PDF)

S3 Table. The ten most upregulated genes in K562 cells upon knockdown of c-Myb. Gene name, ID number, degree of regulation, if the gene contains a c-Myb footprint and the position of the footprint for the ten most upregulated genes in K562 cells upon knockdown of c-Myb [1].
(PDF)

S4 Table. Genomic localisation for c-Myb footprints in the six cell-types analysed. FDR-corrected p-values, p', and normalised ratio for the distribution of c-Myb footprints at annotated genes, promoters and intergenic regions for the six cell-types analysed.
(PDF)

S5 Table. Additional c-Myb target genes used in this study. Additional genes used in the study that are also regulated in K562 cells upon knockdown of c-Myb [1].
(PDF)

S6 Table. Co-localisation between c-Myb footprints and histone modifications. FDR-corrected p-values, p', and normalised ratios for co-localisation between c-Myb footprints and ChIP-Seq peaks for the four histone modifications analysed in K562 cells.
(PDF)

S7 Table. Co-regulatory transcription factors. Transcription factors with ChIP-seq peaks significantly overlapping c-Myb footprints at positively and negatively c-Myb-regulated genes. FDR-adjusted p-values, p', and normalised ratios are shown. For full list of ChIP-seq datasets analysed see S12 Table.
(PDF)

S8 Table. Genes enriched in the GO terms of c-Myb footprints common in all six cell-types.
Gene list enriched in the four GO-terms (Fig 5B) identified by GREAT for c-Myb footprints common in all the six cell-types analysed in this study. The promoter regions of the genes are defined as -2.5 kb upstream to +0.5 kb downstream of the TSS.
(PDF)

S9 Table. Overlap between suggested co-regulatory factors in K562 cells and c-Myb footprints common in all six cell-types.
(PDF)

S10 Table. List of primers used in this study.
(PDF)

S11 Table. Antibodies used in this study.
(PDF)

S12 Table. List of ChIP-Seq peak datasets used in this study.
(PDF)

S1 File. c-Myb footprints identified in the six cell-types analysed.
(ZIP)

Acknowledgments

Author Contributions

Conceived and designed the experiments: MB RE GKS OSG. Performed the experiments: MB KK SG IC GKS. Analyzed the data: MB KK SG IC GKS. Wrote the paper: MB RE OSG SG FD EH GKS.

References

1. Lorenzo PI, Brendeford EM, Gilfillan S, Gavrilov AA, Leedsak M, Razin SV, et al. (2011) Identification of c-Myb Target Genes in K562 Cells Reveals a Role for c-Myb as a Master Regulator. Genes & Cancer. doi: 10.1177/1947601911428224

2. Zhou Y, Ness SA (2013) Myb proteins: angels and demons in normal and transformed cells. Front Biosci 16: 1109.

3. Mucenski ML, McLain K, Kier AB, Swerdlow SH, Schreiner CM, Miller TA, et al. (1991) A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. Cell 65: 677–689. PMID: 1709592

4. Clarke D, Vegiopoulos A, Crawford A, Mucenski M, Bonifer C, Frampton J (2000) In vitro differentiation of c-myb(-/-) ES cells reveals that the colony forming capacity of unilineage macrophage precursors and myeloid progenitor commitment are c-Myb independent. Oncogene 19: 3343–3351. doi: 10.1038/sj.onc.1203661 PMID: 10918591

5. Zhao L, Glazov EA, Pattabiraman DR, Al-Owaidi F, Zhang P, Brown MA, et al. (2011) Integrated genome-wide chromatin occupancy and expression analyses identify key myeloid pro-differentiation transcription factors repressed by Myb. Nucleic Acids Res 39: 4664–4679. doi: 10.1093/nar/gkr024 PMID: 21317192

6. Sakamoto H, Dai G, Tsujino K, Hashimoto K, Huang X, Fujimoto T, et al. (2006) Proper levels of c-Myb are discretely defined at distinct steps of hematopoietic cell development. Blood 108: 896–903. doi: 10.1182/blood-2005-09-3846 PMID: 16597594

7. García P, Clarke M, Vegiopoulos A, Berlanga O, Camelo A, Lorvellec M, et al. (2009) Reduced c-Myb activity compromises HSCs and leads to a myeloproliferation with a novel stem cell basis. EMBO J 28: 1492–1504. doi: 10.1038/emboj.2009.97 PMID: 19360001

8. Akashi K, Traver D, Miyamoto T, Weissman IL (2000) A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature 404: 193–197. doi: 10.1038/35004599 PMID: 10724173

9. Consortium TEP (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74. doi: 10.1038/nature11247 PMID: 22955616

10. Ramsay RG, Gonda TJ (2008) MYB function in normal and cancer cells. Nat Rev Cancer 8: 523–534. doi: 10.1038/nrc2439 PMID: 18574464

11. Stenman G, Andersson MK, Andren Y (2014) New tricks from an old oncogene. Cell Cycle 9: 3058–3067. doi: 10.4161/cc.9.15.12515

12. Xiao C, Calado DP, Galler G, Thai T-H, Patterson HC, Wang J, et al. (2007) MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. Cell 131: 146–159. doi: 10.1016/j.cell.2007.07.021 PMID: 17923094

13. Zhao H, Kalota A, Jin S, Gewirtz AM (2009) The c-myb proto-oncogene and microRNA-15a comprise an active autoregulatory feedback loop in human hematopoietic cells. Blood 113: 505–516. doi: 10.1182/blood-2008-01-136218 PMID: 18818396

14. Sanghvi VR, Mavrakis KJ, Van der Meulen J, Boice M, Wolfe AL, Carty M, et al. (2014) Characterization of a set of tumor suppressor microRNAs in T cell acute lymphoblastic leukemia. Science Signaling 7: ra111. doi: 10.1126/scisignal.2005500 PMID: 25406379

15. Emambokus N, Vegiopoulos A, Harman B, Jenkinson E, Anderson G, Frampton J (2003) Progression through key stages of haemopoiesis is dependent on distinct threshold levels of c-Myb. EMBO J 22: 4478–4488. doi: 10.1093/emboj/cdg434 PMID: 12941699

16. Nakata Y, Shetzline S, Sakashita C, Kalota A, Rallapalli R, Rudnick SI, et al. (2007) c-Myb Contributes to G2/M Cell Cycle Transition in Human Hematopoietic Cells by Direct Regulation of Cyclin B1 Expression. MOLECULAR AND CELLULAR BIOLOGY 27: 2048–2058. doi: 10.1128/MCB.01100-06

17. Li L, Chang W, Yang G, Ren C, Park S, Karantanos T, et al. (2014) Targeting Poly(ADP-Ribose) Polymerase and the c-Myb–Regulated DNA Damage Response Pathway in Castration-Resistant Prostate Cancer. Science Signaling 7: ra47–ra47. doi: 10.1126/scisignal.2005070 PMID: 24847116

18. Zhao L, Ye P, Gonda TJ (2013) The MYB proto-oncogene suppresses monocytic differentiation of acute myeloid leukemia cells via transcriptional activation of its target gene GFI1. Oncogene. doi: 10.1038/onc.2013.419

19. Ye P, Zhao L, Gonda TJ (2013) The MYB oncogene can suppress apoptosis in acute myeloid leukemia cells by transcriptional repression of DRAK2 expression. Leuk Res 37: 595–601. doi: 10.1016/j.leukres.2013.01.012 PMID: 23398943

20. Bianchi E, Zini R, Salati S, Tenedini E, Norfo R, Tagliafico E, et al. (2010) c-myb supports erythropoiesis through the transactivation of KLF1 and LMO2 expression. Blood 116: e99–e110. doi: 10.1182/blood-2009-08-238311 PMID: 20686118

21. Deisenroth C, Thorner AR, Enomoto T, Perou CM, Zhang Y (2010) Mitochondrial Hep27 is a c-Myb target gene that inhibits Mdm2 and stabilizes p53. MOLECULAR AND CELLULAR BIOLOGY 30: 3981–3993. doi: 10.1128/MCB.01284-09

22. Hooper J, Maurice D, Argent-Katwala MJG, Weston K (2008) Myb proteins regulate expression of histone variant H2A.Z during thymocyte development. Immunology 123: 282–289. doi: 10.1111/j.1365-2567.2007.02697.x PMID: 17931383

23. Berge T, Matre V, Brendeford EM, Saether T, Lüscher B, Gabrielsen OS (2007) Revisiting a selection of target genes for the hematopoietic transcription factor c-Myb using chromatin immunoprecipitation and c-Myb knockdown. Blood Cells Mol Dis 39: 278–286. doi: 10.1016/j.bcmd.2007.05.007 PMID: 17587615

24. Maurice D, Hooper J, Lang G, Weston K (2007) c-Myb regulates lineage choice in developing thymocytes via its target gene Gata3. EMBO J 26: 3629–3640. doi: 10.1038/sj.emboj.7601801 PMID: 17641686

25. Shetzline SE, Rallapalli R, Dowd KJ, Zou S, Nakata Y, Swider CR, et al. (2004) Neuromedin U: a Myb-regulated autocrine growth factor for human myeloid leukemias. Blood 104: 1833–1840. doi: 10.1182/blood-2003-10-3577 PMID: 15187020

26. Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, et al. (2014) An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science. doi: 10.1126/science.1259037

27. Galas DJ, Schmitz A (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res 5: 3157–3170. PMID: 212715

28. Henikoff JG, Belsky JA, Krassovsky K, Macalpine DM, Henikoff S (2011) Epigenome characterization at single base-pair resolution. Proc Natl Acad Sci USA 108: 18318–18323. doi: 10.1073/pnas. 1110731108 PMID: 22025700

29. Boyle AP, Song L, Lee B-K, London D, Keefe D, et al. (2011) High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Research 21: 456–464. doi: 10.1101/ gr.112656.110 PMID: 21106903

30. Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, et al. (2009) Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Meth 6: 283–289. doi: 10.1038/nmeth.1313

31. Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, et al. (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 489: 83–90. doi: 10.1038/ nature11212 PMID: 22955618

32. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK (2011) Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Research 21: 447–455. doi: 10.1101/gr.112623.110 PMID: 21106904

33. Orlic D, Anderson S, Biesecker LG, Sorrentino BP, Bodine DM (1995) Pluripotent hematopoietic stem cells contain high levels of mRNA for c-kit, GATA-2, p45 NF-E2, and c-myb and low levels or no mRNA for c-fms and the receptors for granulocyte colony-stimulating factor and interleukins 5 and 7. Proc Natl Acad Sci USA 92: 4601–4605. PMID: 7538677

34. Nakata Y, Brignier AC, Jin S, Shen Y, Rudnick SI, Sugita M, et al. (2010) c-Myb, Menin, GATA-3, and MLL form a dynamic transcription complex that plays a pivotal role in human T helper type 2 cell development. Blood 116: 1280–1290. doi: 10.1182/blood-2009-05-223255 PMID: 20484083

35. Thomas MD, Kremer CS, Ravichandran KS, Rajewsky K, Bender TP (2005) c-Myb is critical for B cell development and maintenance of follicular B cells. Immunity 23: 275–286. doi: 10.1016/j.immuni.2005. 08.005 PMID: 16169500

36. Sakamoto Y, Watanabe S, Ichimura T, Kawasuji M, Koseki H, Baba H, et al. (2007) Overlapping roles of the methylated DNA-binding protein MBD1 and polycomb group proteins in transcriptional repression of HOXA genes and heterochromatin foci formation. J Biol Chem 282: 16391–16400. doi: 10.1074/jbc. M700011200 PMID: 17428788

37. Klepper K, Drabløs F (2013) MotifLab: a tools and data integration workbench for motif discovery and regulatory sequence analysis. BMC Bioinformatics 14: 9. doi: 10.1046/j.1471-4159.2002.00890.x PMID: 23323883

38. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34: D108– D110. doi: 10.1093/nar/gkj143 PMID: 16381825

39. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. (2012) The accessible chromatin landscape of the human genome. Nature 489: 75–82. doi: 10.1038/nature11232 PMID: 22955617

40. Samstein RM, Arvey A, Josefowicz SZ, Peng X, Reynolds A, Sandstrom R, et al. (2012) Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. Cell 151: 153–166. doi: 10.1016/j.cell.2012.06.053 PMID: 23021222

41. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Research 15: 1034– 1050. doi: 10.1101/gr.3715005 PMID: 16024819

42. Kulakovskiy IV, Boeva VA, Favorov AV, Makeev VJ (2010) Deep and wide digging for binding motifs in ChIP-Seq data. Bioinformatics 26: 2622–2623. doi: 10.1093/bioinformatics/btq488 PMID: 20736340

43. Dekker J (2014) Two ways to fold the genome during the cell cycle: insights obtained with chromosome conformation capture. Epigenetics Chromatin 7: 25. doi: 10.1016/j.cell.2013.02.028 PMID: 25435919

44. Molvaersmyr A- K, Saether T, Gilfillan S, Lorenzo PI, Kvaløy H, Matre V, et al. (2010) A SUMO-regulated activation function controls synergy of c-Myb through a repressor-activator switch leading to differential p300 recruitment. Nucleic Acids Res 38: 4970–4984. doi: 10.1093/nar/gkq245 PMID: 20385574

45. Dahle O, Andersen TØ, Nordgård O, Matre V, Del Sal G, Gabrielsen OS (2003) Transactivation properties of c-Myb are critically dependent on two SUMO-1 acceptor sites that are conjugated in a PIASy enhanced manner. Eur J Biochem 270: 1338–1348. PMID: 12631292

46. van Steensel B, Delrow J, Henikoff S (2001) Chromatin profiling using targeted DNA adenine methyltransferase. Nat Genet 27: 304–308. doi: 10.1038/85871 PMID: 11242113

47. Vogel MJ, Peric-Hupkes D, van Steensel B (2007) Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nat Protoc 2: 1467–1478. doi: 10.1038/nprot.2007.148 PMID: 17545983

48. Van Dessel N, Beke L, Görnemann J, Minnebo N, Beullens M, Tanuma N, et al. (2010) The phospha-tase interactor NIPP1 regulates the occupancy of the histone methyltransferase EZH2 at Polycomb tar-gets. Nucleic Acids Res 38: 7500–7512. doi: 10.1093/nar/gkq643 PMID: 20671031

49. Jin S, Zhao H, Yi Y, Nakata Y, Kalota A, Gewirtz AM (2010) c-Myb binds MLL through menin in human leukemia cells and is an important driver of MLL-associated leukemogenesis. J Clin Invest 120: 593–606. doi: 10.1172/JCI38030 PMID: 20093773

50. ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74. doi: 10.1038/nature11247 PMID: 22955616

51. Rosenbloom KR, Dreszer TR, Long JC, Malladi VS, Sloan CA, Raney BJ, et al. (2012) ENCODE whole-genome data in the UCSC Genome Browser: update 2012. Nucleic Acids Res 40: D912–D917. doi: 10.1093/nar/gkr1012 PMID: 22075998

52. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129: 823–837. doi: 10.1016/j.cell.2007.05.009 PMID: 17512414

53. Mikkelsen TS, Xu Z, Zhang X, Wang L, Gimble JM, Lander ES, et al. (2010) Comparative epigenomic analysis of murine and human adipogenesis. Cell 143: 156–169. doi: 10.1016/j.cell.2010.09.006 PMID: 20887899

54. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins D, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39: 311–318. doi: 10.1038/ng1966 PMID: 17277777

55. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125: 315–326. doi: 10.1016/j.cell.2006.02.041 PMID: 16630819

56. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, et al. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441: 349–353. doi: 10.1038/nature04733 PMID: 16625203

57. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, et al. (2006) Control of develop-mental regulators by Polycomb in human embryonic stem cells. Cell 125: 301–313. doi: 10.1016/j.cell.2006.02.043 PMID: 16630818

58. Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13: 613–626. doi: 10.1038/nrg3207 PMID: 22868264

59. Shapiro LH (1995) Myb and Ets proteins cooperate to transactivate an early myeloid gene. J Biol Chem 270: 8763–8771. PMID: 7721782

60. Tanikawa J, Nomura T, Macmillan EM, Shinagawa T, Jin W, Kokura k,et al. (2004) p53 suppresses c-Myb-induced trans-activation and transformation by recruiting the corepressor mSin3A. J Biol Chem 279: 55393–55400. doi: 10.1074/jbc.M411658200 PMID: 15509555

61. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. (2010) GREAT improves func-tional interpretation of cis-regulatory regions. Nat Biotech 28: 495–501. doi: 10.1038/nbt.1630

62. Tuteja G, Moreira KB, Chung T, Chen J, Wenger AM, Bejerano G (2014) Automated discovery of tis-sue-targeting enhancers and transcription factors from binding motif and gene function data. PLoS Comput Biol 10: e1003449. doi: 10.1371/journal.pcbi.1003449 PMID: 24499934

63. Lipsick JS (1996) One billion years of Myb. Oncogene 13: 223–235. PMID: 8710361

64. Tanaka Y, Patestos NP, Maekawa T, Ishii S (1999) B-myb is required for inner cell mass formation at an early stage of development. J Biol Chem 274: 28067–28070. PMID: 10497154

65. Toscani A, Mettus RV, Coupland R, Simpkins H, Litvin J, Orth J, et al. (1997) Arrest of spermatogenesis and defective breast development in mice lacking A-myb. Nature 386: 713–717. doi: 10.1038/386713a0 PMID: 9109487

66. Pinson B, Brendeford EM, Gabrielsen OS, Daignan-Fornier B (2001) Highly conserved features of DNA binding between two divergent members of the Myb family of transcription factors. Nucleic Acids Res 29: 527–535. PMID: 11139623

67. Golay J, Loffarelli L, Luppi M, Castellano M, Introna M (1994) The human A-myb protein is a strong acti-vator of transcription. Oncogene 9: 2469–2479. PMID: 8058310

68. Weston K (1992) Extension of the DNA binding consensus of the chicken c-Myb and v-Myb proteins. Nucleic Acids Res 20: 3043–3049. PMID: 1620600

69. Bergholtz S, Andersen TO, Andersson KB, Borrebaek J, Lüscher B, Gabrielsen OS (2001) The highly conserved DNA-binding domains of A-, B- and c-Myb differ with respect to DNA-binding, phosphoryla-tion and redox properties. Nucleic Acids Res 29: 3546–3556. PMID: 11522824

70. Sung M-H, Guertin MJ, Baek S, Hager GL (2014) DNase Footprint Signatures Are Dictated by Factor Dynamics and DNA Sequence. Mol Cell 56: 275–285. doi: 10.1016/j.molcel.2014.08.016 PMID: 25242143

71. Zargarian L, Le Tilly V, Jamin N, Chaffotte A, Gabrielsen OS, Toma F, et al. (1999) Myb-DNA recognition: role of tryptophan residues and structural changes of the minimal DNA binding domain of c-Myb. Biochemistry 38: 1921–1929. doi: 10.1021/bi981199j PMID: 10026273

72. Myrset AH, Bostad A, Jamin N, Lirsac PN, Toma F, Gabrielsen OS (1993) DNA and redox state induced conformational changes in the DNA-binding domain of the Myb oncoprotein. EMBO J 12: 4625–4633. PMID: 8223472

73. Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP, et al. (2014) Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat Biotech 32: 171–178. doi: 10.1038/nbt.2798

74. Chen X, Hoffman MM, Bilmes JA, Hesselberth JR, Noble WS (2010) A dynamic Bayesian network for identifying protein-binding footprints from single molecule-based sequencing data. Bioinformatics 26: i334–i342. doi: 10.1093/bioinformatics/btq175 PMID: 20529925

75. Piper J, Elze MC, Cauchy P, Cockerill PN, Bonifer C, Ott S (2013) Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Res 41: e201. doi: 10.1093/nar/gkt850 PMID: 24071585

76. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G (2013) Enhancers: five essential questions. Nat Rev Genet 14: 288–295. doi: 10.1038/nrg3458 PMID: 23503198

77. Stergachis AB, Neph S, Sandstrom R, Haugen E, Reynolds AP, Zhang M, et al. (2014) Conservation of trans-acting circuitry during mammalian regulatory evolution. Nature 515: 365–370. doi: 10.1038/nature13972 PMID: 25409825

78. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. Nature 444: 499–502. doi: 10.1038/nature05295 PMID: 17086198

79. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457: 854–858. doi: 10.1038/nature07730 PMID: 19212405

80. Erwin GD, Oksenberg N, Truty RM, Kostka D, Murphy KK, Ahituv N, et al. (2014) Integrating diverse datasets improves developmental enhancer prediction. PLoS Comput Biol 10: e1003677. doi: 10.1371/journal.pcbi.1003677 PMID: 24967590

81. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. (2014) An atlas of active enhancers across human cell types and tissues. Nature 507: 455–461. doi: 10.1038/nature12787 PMID: 24670763

82. Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. (2010) ChIP-Seq identification of weakly conserved heart enhancers. Nat Genet 42: 806–810. doi: 10.1038/ng.650 PMID: 20729851

83. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, et al. (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science 328: 1036–1040. doi: 10.1126/science.1186176 PMID: 20378774

84. Mercer TR, Edwards SL, Clark MB, Neph SJ, Wang H, Stergachis AB, et al. (2013) DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. Nat Genet 45: 852–859. doi: 10.1038/ng.2677 PMID: 23793028

85. Nakamura T, Mori T, Tada S, Krajewski W, Rozovskaia T, Wassell R, et al. (2002) ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol Cell 10: 1119–1128. PMID: 12453419

86. Saether T, Berge T, Ledsaak M, Matre V, Alm-Kristiansen AH, Dahle O, et al. (2007) The chromatin remodeling factor Mi-2alpha acts as a novel co-activator for human c-Myb. J Biol Chem 282: 13994–14005. doi: 10.1074/jbc.M700755200 PMID: 17344210

87. Wang P, Lin C, Smith ER, Guo H, Sanderson BW, Wu M, et al. (2009) Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. MOLECULAR AND CELLULAR BIOLOGY 29: 6074–6085. doi: 10.1128/MCB.00924-09

88. Hu D, Gao X, Morgan MA, Herz H-M, Smith ER, Shilatifard A (2013) The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. MOLECULAR AND CELLULAR BIOLOGY 33: 4745–4754. doi: 10.1128/MCB.01181-13

89. Hubbard TJP, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, et al. (2009) Ensembl 2009. Nucleic Acids Res 37: D690–D697. doi: 10.1093/nar/gkn828 PMID: 19033362

90. Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res 31: 3576–3579. PMID: 12824369

91. Sandve GK, Gundersen S, Johansen M, Glad IK, Gunathasan K, Holden L, et al. (2013) The Genomic HyperBrowser: an analysis web server for genome-scale data. Nucleic Acids Res 41: W133–W141. doi: 10.1093/nar/gkt342 PMID: 23632163

92. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing on JSTOR. Journal of the Royal Statistical Society 57: 289–300. Available: http://www.jstor.org/stable/2346101?seq=1#page_scan_tab_contents. Accessed 17 February 2015.

93. Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, et al. (2012) BEDOPS: high-performance genomic feature operations. Bioinformatics 28: 1919–1920. doi: 10.1093/bioinformatics/bts277 PMID: 22576172

94. Oki S, Maehara K, Ohkawa Y, Meno C (2014) SraTailor: graphical user interface software for processing and visualizing ChIP-seq data. Genes Cells 19: 919–926. doi: 10.1111/gtc.12190 PMID: 25324176

95. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Meth 9: 357–359. doi: 10.1038/nmeth.1923

96. Feng J, Liu T, Zhang Y (2011) Using MACS to identify peaks from ChIP-Seq data. Curr Protoc Bioinformatics Chapter 2: Unit2.14. doi: 10.1002/0471250953.bi0214s34