blood

Prepublished online Oct 19, 2010;
doi:10.1182/blood-2010-07-295113

Dynamic combinatorial interactions of RUNX1 and cooperating partners regulates megakaryocytic differentiation in cell line models

Niv Pencovich, Ram Jaschek, Amos Tanay and Yoram Groner

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published semimonthly by the American Society of Hematology, 1900 M St, NW, Suite 200, Washington DC 20036.

Dynamic combinatorial interactions of RUNX1 and cooperating partners regulates megakaryocytic differentiation in cell line models

Niv Pencovich[1,#], Ram Jaschek[2.#], Amos Tanay[2] and Yoram Groner[1,*]

Departments of [1]Molecular Genetics and [2]Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot, 76100 Israel.

Running Title: RUNX1 regulates megakaryocytic gene expression

[*]Corresponding author: Y. Groner Department of Molecular Genetics, The Weizmann Institute of Science, Rehovot, 76100 Israel Tel: +972-8-9343972; Fax: +972-8-9344108; E-mail: yoram.groner@weizmann.ac.il

[#]These authors contributed equally to this work.

Abstract

Specific interactions of transcription factors (TFs) with their targets are crucial for specifying gene expression programs during cell differentiation. How specificity is maintained despite limited selectivity of individual TF-DNA interactions is not fully understood. *RUNX1* TF is among the most frequently mutated genes in human leukemia and an important regulator of megakaryopoiesis. We used megakaryocytic cell lines to characterize the network of RUNX1 targets and cooperating TFs in differentiating megakaryocytes and demonstrated how dynamic partnerships between RUNX1 and cooperating TFs facilitated regulatory plasticity and specificity during this process. Following differentiation onset RUNX1 directly activated a large number of genes through interaction with preexisting and *de-novo* binding sites. Recruitment of RUNX1 to *de-novo* occupied sites occurred at H3K4me1-marked preprogrammed enhancers. Significant number of these *de-novo* bound sites lacked RUNX motif, but were occupied by AP-1 TFs. Reciprocally, AP-1 TFs were upregulated by RUNX1 following TPA induction, and recruited to RUNX1 occupied sites lacking AP-1 motifs. At other differentiation stages, additional combinatorial interactions occurred between RUNX1 and its co-regulators, GATA1 and ETS. The findings suggest that in differentiating megakaryocytic cell-lines, RUNX1 cooperates

with GATA1, AP-1 and ETS to orchestrate cell-specific transcription programs through dynamic TF partnerships.

Introduction

The RUNX TFs are key regulators of cell lineage and differentiation in several important developmental pathways. They regulate transcription in a context-dependent manner through binding to the consensus core DNA sequence PyGPyGGT [1]. RUNX1 functions as key regulator in embryonic and adult hematopoiesis [2]. Consistent with its important roles, haploinsufficiency, due to heterozygous loss-of-function mutations, is associated with familial platelet disorder and predisposition to acute myeloid leukemia (FPD-AML) [3,4]. Sporadic heterozygous mutations in *RUNX1* are also leukemogenic [5,6]. *RUNX1* resides on human chromosome 21 and chromosomal translocations involving RUNX1, including 8;21, 3;21, and 12;21 are among the most frequent leukemia associated translocations [7]. In addition, patients with Down syndrome (DS), the phenotypic manifestation of trisomy 21, have 500 fold-increased risk of developing acute megakaryoblastic leukemia (DS-AMKL/AML-M7) relative to normal individuals [8].

RUNX1 plays an important role in megakaryopoiesis; the process leading to production of megakaryocytes, the polyploid precursors of platelets [9,10]. Megakaryocytes share a common precursor with erythrocytes known as the megakaryocyte erythroid progenitor (MEP) which gives rise to both megakaryocytic and erythroid lineages [9,10]. Overexpression of RUNX1 in myeloid cell lines induces megakaryocytic differentiation [11,12], while induced Runx1 deficiency in bone marrow results in impaired megakaryocytic maturation and reduced blood platelet number (thrombocytopenia) [13]. While the cellular differentiation stages of megakaryopoiesis are well characterized, the regulatory programs responsible for the implementation of this process are largely unknown, as are the global RUNX1-regulatory mechanisms and direct target genes that drive this differentiation process.

RUNX1, in conjunction with additional sequence-specific TFs regulates hematopoietic cell-differentiation programs through specific interaction with its target genes following developmental signals [14]. In complex metazoan genomes, sequence recognition of binding site motifs by TFs is by itself not sufficient to discriminate bona fide binding sites from background genomic sequences. Hence, additional parameters such as chromatin structure and interactions with cooperating TFs

determined the functionality of potential binding sites. In a typical scenario, only a fraction of the numerous potential TF binding site motifs in the genome is occupied at a given state, and even smaller subset directly regulate transcription. This flexible selectivity creates a dense network of TF-genome interactions, which is currently difficult to predict and/or understand. Most importantly, it is unclear how to discern functionally important TF-genome interactions from transient or spurious ones and hence define the interactions that play active role in transcriptional regulation [15]. Protein-protein interactions between TFs that simultaneously engage DNA [16], add another layer of complexity challenging our current understanding of transcriptional control.

Here we used TPA treated K562 [17] and CMK cells to model megakaryocytic differentiation and to explore cell immediate response to a differentiation signal. We found that RUNX1 acts as an essential regulator of immediate gene expression and characterized its genome-wide occupancy profile before and after induction of differentiation. A combination of genome-wide ChIP-seq occupancy and gene expression profiles was used to identify a subset of RUNX1 sites directly involved in regulatory response. Additional ChIP-seq and sequence analysis delineated the epigenomic landscape of H3K4me1/H3K27me3 and cooperating TFs that participate in RUNX1-mediated cell response to TPA. The data provide the first genomewide profile of RUNX1-occupancy before and during megakaryocytic differentiation and revealed a set of functional target genes downstream to a complex landscape of numerous RUNX1 binding sites. The analysis elucidated how the limited sequence-specificity of RUNX1 is diversified by the epigenomic makeup (H3K4me1 vs. H3K27me3) and the binding landscape of RUNX1 cooperating TFs. It shows that RUNX1 and its partners act in a coordinated manner affecting gene expression outcome. The data suggest that stage specific combinatorial interactions, in addition to epigenomic makeup, dynamically shape the transcriptional program during megakaryocytic differentiation.

Material and Methods

Cells

K562 and CMK cells were maintained in RPMI medium supplemented with 10%

fetal bovine serum (FBS) (Gibco, US), 2mM L-glutamine and penicillin/streptomycin at 37ºC and 5% CO_2. K562 cells were treated with 40nM 12-*O*-tetradecanoylphorbol-13-acetate (TPA) (Sigma-Aldrich, US) to induce megakaryocytic differentiation. For the generation of stable knockdown of RUNX1 in K562 cells ($K562^{RUNX1\text{-}KD}$), RUNX1 pGIPZ lentiviral shRNAmir vector V2LHS_150257 (Open Biosystems, US RHS4531-NM_001754) was transfected into K562 using the Lipofectamine reagent (Invitrogen, US) according to manufacturer instructions. For selection of RUNX1 knockdown cells, culture was supplemented with Puromycin (2µg ml^{-1}, Sigma, US) and medium replaced every 72h. Non-silencing lentiviral shRNAmir vector (Open Biosystems) was used for negative controls.

Further information on immunoprecipitation and Western blot analysis of RUNX1 in cell lysates of K562, K562-TPA and CMK cells as well as generation of primary fetal liver derived megakaryocytes is included in Supplementary Methods.

Chromatin Immunoprecipitation (ChIP)-sequencing (ChIP-seq)

ChIP was performed essentially as described [18]. Briefly, cross-linked chromatin from ~10^8 K562 cells, before or 24h after treatment with TPA (40nM) or from ~10^8 cells CMK or 10^7 mature FL derived megakaryocytes, was prepared and fragmented to an average size of ~200bp by 40 cycles of sonications (30sec each) in 15ml tubes using the Bioruptor UCD-200 sonicator (Diagenode, US). For immunoprecipitation, the following antibodies were added to 12ml of diluted, fragmented chromatin: 30ul of home-made anti-RUNX1[19] raised against the protein C-terminal fragment; anti-monomethyl-Histone H3(Lys4) and anti trimethyl-Histone H3(Lys27) (Millipore, US); anti C-FOS (Santa Cruz), anti FOS-B (Cell Signaling) and anti GATA1 (Abcam, US). Rabbit pre-immune serum was used as control. DNA was purified using QIAquick spin columns (QIAGEN, US) and sequencing performed using Illumina genome analyzer IIx, according to manufacturer instructions. Two biological repeats were conducted and separately sequenced with each cell line and/or physiological condition. For ChIP-seq analysis**,** Illumina sequencing short reads (36bp) were aligned to the human genome (hg18) using the Eland program (Illumina). Multiple reads were discriminated, and coverage profile generated by elongating reads to 200bp according to mapped strand. Coverage profile was analyzed in bins of 50bp unless otherwise noted. Non-immune serum ChIP was used to discard regions with higher than expected background coverage (>6 mapped elongated reads).

Further information on ChIP quantitative PCR and ChIP-seq data validation by reporter construct transfection assays is included in Supplementary Methods.

Microarray processing and analysis

RNA was isolated by EZ-RNA (Biological Industries, Beit Haemek Israel), according to manufacturer instructions. Purified RNA was reverse-transcribed, amplified and labeled with Affymetrix GeneChip whole transcript sense target labeling kit. Labeled cDNA from TPA treated or untreated K562 cells was analyzed using Affymetrix human exon ST 1.0 microarrays, according to manufacturer instructions. Microarrays were scanned using GeneChip scanner 3000 7G. Microarrays data was normalized using dChip model based expression. All microarray data are available in the GEO public database under accession number GSE24779.

Further information on gene expression assay by quantitative RT-PCR is included in Supplementary Methods.

Analysis of genomic regions encompassing promoters and enhancers

Annotated transcription start sites were downloaded from the UCSC site (January 2010 version). For the analysis shown in Figures 2-5, each genomic locus was associated with the nearest TSS. Loci at a distance of up to 3K were categorized as "promoter regions", while loci at a distance between 3K and 200K were categorized "enhancer regions".

Distribution of RUNX1 genomic occupancy indicated numerous strong binding peaks under all tested conditions as well as a significant number of weaker ones (Figure S10). This behavior was even more pronounced when H3K4me1 profiles were considered. These observations and the notion that a flexible wide range of interaction specificities exist for certain transcription factors, suggested that it would be impossible (or in fact undesirable) to apply a single universal definition of a RUNX binding site. As the main goal of the analysis was to obtain data on the global behavior of RUNX1 and its potential cooperating TFs, we applied a simple coverage threshold to detect RUNX1 sites and candidate enhancers. Genomic regions with high binding coverage in the non-immune serum control ChIP-seq experiments (n>6) were discarded. Contiguous regions with high binding coverage were grouped together to form distinct binding sites or enhancer regions. The cut-off for RUNX1 was based on the top 0.05% of data in the K562 profile (weighted coverage >13). The cut-offs for

all other tracks were scaled proportionally to the number of reads in the track. The ChIP-seq analysis coverage statistics and derived cut-off values are shown in Supplementary Table S3. Further information about analysis of ChIP-seq data is described in the Supplementary Methods.

Results

RUNX1 expression in megakaryocytic cell lines

RUNX1 is highly expressed in megakaryocytic cell lines, including CMK and Meg01 (Supplementary Figure S1A and B). In the multipotent cell line K562, RUNX1 expression was upregulated upon induction of megakaryocytic differentiation by 12-*O*-tetradecanoylphorbol-13-acetate (TPA) (Figure 1A) as was also observed by Elagib *et al*. 2003[12] . These findings indicated that analysis of these cell lines under attenuated RUNX1 expression (Figure 1A) would furnish important information on the transcriptional program regulated by RUNX1 during megakaryopoiesis.

RUNX1 is a key gene expression regulator during megakaryocytic differentiation of K562 cells

K562 cells readily differentiate along the megakaryocytic lineage following TPA treatment [17] providing a well-characterized system for studying megakaryopoiesis[12] (and ref therein). Treatment with TPA induced a profound decrease in cell proliferation (Figure 1B) and changes in cell morphology (Figure 1C) as was previously reported [17]. These changes that were characteristic of megakaryocytic differentiation, include increased cell size and cytoplasm to nucleus ratio, reduced basophilic staining of cytoplasm, appearance of lobulated nuclei (Figure 1C) and increase in the expression of megakaryocytic markers (Figure 1D). Of note, knockdown (KD) of RUNX1 in K562 cells ($K562^{Runx1KD}$) (Fig 1A) resulted in marked diminution of the TPA effect on proliferation, cell morphology and expression of megakaryocytic markers (Figure 1B, C and D).

Gene expression analysis of K562 cells before and after TPA treatment (Figure 1E) revealed an extensive transcriptional response in the first 48 hours of treatment. Changes included repression of genes involved in growth-related pathways such as ribosomal proteins and DNA synthesis, and induction of numerous genes in pathways

involved in megakaryocytic differentiation (Supplementary Table S1). Significantly, comparable analysis in $K562^{Runx1KD}$ cells (Figure 1F) showed that ~ 80% of these induced megakaryopoitic genes displayed low response to TPA in the absence of RUNX1 (Figure 1F and G). In contrast, KD of RUNX1 did not systematically compromise the repression of immediate TPA responding genes (Figure 1G). These results identified a large set of TPA responsive genes whose transcriptional regulation was RUNX1-dependent (Supplementary Table S1) and established TPA treated K562 vs. $K562^{Runx1KD}$ cells as a unique system for analyzing the molecular events underlying RUNX1-mediated regulation during megakaryocytic differentiation in this cell line.

Induction of differentiation in K562 involves de-novo recruitment of RUNX1 to a large number of genomic sites

We used our highly specific anti RUNX1 antibodies (Figure 2A and Supplementary Methods) in ChIP–seq experiments to map RUNX1 binding in K562 before and after TPA treatment. The genome-wide RUNX1 binding profiles were then combined with genome-wide mapping of enhancer/promoter regions by H3K4me1/H3K27me3 ChIP-seq analysis. Prior to induction of megakaryocytic differentiation, RUNX1 occupied several thousand loci (3532 permissive threshold sites). Following induction, the original RUNX1 binding sites were largely preserved (Figure 2B and Supplementary Figure S2). But in addition, a large number of *de-novo* RUNX1 regions became occupied; increasing the number of RUNX1 bound sites by more than threefold, to a total of 12,507 bound sites (Figure 2B**).** These data support the finding that RUNX1 plays a pivotal role in regulating the TPA induced transcriptional program in K562 cells.

Analysis of RUNX1 occupancy sites locations, relative to the nearest transcription start sites (TSSs) of annotated genes, revealed that ~80% of RUNX1 bound sites were situated more than 5kb away from any TSS (Figure 2C), and ~25% were in “gene deserts” (over 100kb from the nearest TSS). The majority of *de-novo* RUNX1 occupied sites are therefore either not functional, or affect transcription through long-range promoter-enhancer interactions. The apparent plasticity and wide distribution of RUNX1 occupancy landscape, suggested that RUNX1 regulates gene expression via multiple interactions with genomic chromatin and other transcriptional regulators.

RUNX1 is preferentially recruited to sites of preprogrammed open chromatin

H3K4me1 marks chromatin of genomic regions associated with enhancer activity [20]. Using H3K4me1 ChIP-seq we analyzed the chromatin landscape, before and after the massive recruitment of RUNX1 to *de-novo* TPA-induced sites upon switch-on of the differentiation program. In K562 cells, RUNX1 binding is largely confined to regions displaying H3K4me1 occupancy (Figure 2D). Following induction of megakaryocytic differentiation, the genomic landscape of H3K4me1 regions expanded and changed, as a large group of loci (~25,000) acquired *de-novo* mono-methylation at H3K4 (Figure 2E blue), but fewer lost their existing marks (Figure 2E green). A third group, designated "constitutive" was marked with H3K4me1 in both pre- and post-TPA treated cells (Figure 2E gray). Importantly, the numerous *de-novo* (post induction) RUNX1 occupied sites belonged to this constitutively marked H3K4me1 group; sites that were already marked with H3K4me1 prior to induction (Figure 2F).

Analysis of H3K27me3 ChIP-seq readout in RUNX1 bound peaks indicated a general lack of overlap between RUNX1 occupied enhancers and the polycomb repressive histone marker (Figure 2G & H). Together, these results are consistent with the conclusion that RUNX1 recruitment, during the first 24 hours post induction, did not require extensive chromatin remodeling, and that the newly engaged enhancers were actually accessible prior to induction, but became occupied by RUNX1 only after the onset of the differentiation program.

Transcriptional activation of target genes is tightly correlated with RUNX1 recruitment to remote binding sites

As shown above, megakaryocytic differentiation of K562 cells was largely dependent on RUNX1-mediated gene expression (Figure 1) and was associated with a vast increase in *de-novo* occupied RUNX1 sites (Figure 2B and C). This would argue that *de-novo* RUNX1 bound genomic elements directly regulate TPA induced genes. Consistent with this possibility, the stringently selected group of 147 genes (Supplementary Table S1), which were TPA-induced and RUNX1-regulated, displayed a significant enrichment for *de-novo* RUNX1 occupancy within 250kb around their activated TSSs (Figure 3A). Interestingly, this defined subset of apparently direct RUNX1 targets (marked in Table S1) contained a preponderance of genes important for megakaryopoiesis. Together, the data establish a causal link

between *de-novo* RUNX1 occupancy and TPA induction of differentiation. Importantly, the finding that most (~80%) of these *de-novo* RUNX1 bound sites, in proximity to activated genes, were localized faraway from the TSSs (Figure 2C and Figure 3A), indicates that RUNX1 regulates its target genes primarily through long-range enhancer promoter interactions. Of note, a significant statistical dependency (Spearman = 0.07, $p<1.46e^{57}$) was observed between increased RUNX1 occupancy at gene promoters vs. their surrounding enhancers (Figure 3B**)**, underscoring the importance of remote enhancer-promoter interaction in RUNX1-mediated response to TPA.

This finding raised the possibility that some of the identified RUNX1 promoter-occupancy sites resulted from initial binding at remote enhancers followed by chromosomal looping [21,22]. This interpretation is illustrated by ChIP-seq readouts of several TPA induced RUNX1 regulated megakaryocytic genes encompassing remote newly occupied RUNX1 binding sites spanning H3K4me1 rich H3K27me3 poor regions (Figure 3C). RUNX1 occupancy along the regions shown in Figure 3C, was further evaluated using quantitative ChIP-PCR. RUNX1 binding was detected in all ChIP-seq peaks tested (Figure 3D), albeit it was higher at sites containing RUNX motifs. Moreover, Runx1 binding to several homologous mouse regions was detected by quantitative ChIP-PCR using fetal liver derived murine megakaryocytes (Figure 3D).

Sequence specificity of RUNX1 occupancy sites

To characterize the sequence specificity of constitutive RUNX1 occupancy sites we searched for DNA sequence motifs, within RUNX1 bound regions, prior to TPA treatment, in comparison to a background set consisting of H3K4me1-enriched regions lacking RUNX1 occupancy (for details see Supplementary Methods). This analysis confirmed the existence of a RUNX motif, which specified RUNX1 binding to a unique subset of enhancer and promoter elements distinguished from background enhancers (Figure 4A). Interestingly, while this RUNX1 specific motif was highly significant ($P<10^{-53}$), it was detected in only ~40% of occupied sites, when setting the background motif percentage at 5%. This occurred even when RUNX1 sites were selected from the H3K4me1-marked enhancers rather than considering the entire genome. Such loose specificity, previously found for other mammalian TFs [23-25], suggested that additional sequences and/or co-factors are required to specify RUNX1

binding. On the other hand, analysis of the correspondence between predicted RUNX1 binding potential (binding energy, see M&M and Supplementary Methods) and the actual level of RUNX1 ChIP-seq *in-vivo*, revealed a weak but statistically significant correlation (Pearson = 0.08, $p<10^{-300}$, Figure 4A), even for RUNX motifs of less than optimal sequence. This wide pattern of correlation suggested that weaker, sub-optimal RUNX motifs were still playing a role in specifying RUNX binding sites, probably in cooperation with additional TF motifs. Interestingly, the correlation between RUNX sequence motif binding energy and actual RUNX1 binding was weaker at promoter regions than at enhancer regions, (in promoters: Pearson=0.06, $p<10^{-137}$ in enhancers: Pearson=0.09, $p<10^{-300}$), supporting the thesis that some of the reported RUNX1 promoter occupancy resulted from chromatin looping.

GATA motifs enrichment and GATA1/RUNX1 co-occupancy at constitutive RUNX1 bound sites

Prior to induction of megakaryocytic gene expression program by TPA (Figure 1**)**, RUNX1 was bound at 3,538 genomic sites of which at least 2,504 were also occupied after TPA treatment (Figure 2B) and were stringently defined as constitutively occupied regions. Sequence analysis of these regions revealed significant enrichment for GATA box motifs (Figure 4B, $P<1e-^{24}$). Genome-wide analysis of distance distribution of RUNX-GATA motifs revealed that at constitutively occupied regions the two motifs were coupled, whereas in *de-novo* RUNX1 bound regions this coupling was weak (Figure 4C). This significant association between RUNX1-GATA at constitutively occupied RUNX1 regions was confirmed by analysis of previously published [26,33] GATA-1 ChIP-seq data in K562 cells. It showed that 25% of RUNX1 bound regions were co-occupied by GATA1 (Figures 4D & 4E and Supplementary Figure S3A**).**

This latter finding was in clear contrast to the limited co-occurrence of RUNX1 and GATA1 bound sites revealed by GATA1 ChIP-seq analysis of TPA treated K562 cells (Figure 4F). The ChIP-seq data was further confirmed by independent ChIP-PCR analysis of RUNX1 and GATA1 on several RUNX1 target genes (Supplementary Figure S3B), indicating lack of GATA1 binding at *de novo* occupied regions. The co-occurrence of RUNX and GATA motifs at constitutively bound RUNX1 sites and the ChIP-seq co-occupancy of GATA1 and RUNX1 prior to TPA treatment strongly indicated that combination of the two TFs plays a role in early

stages of the differentiating program. Supporting this idea are the findings that RUNX1 and GATA-1 have an essential role in megakaryopoiesis [6,13,27-30] and functionally cooperate in this process [12,14]. It suggests that RUNX1 modulates the broad regulatory activity spectrum of GATA-1, known to regulate development of other hematopoietic lineages [31-34].

AP-1 motifs enrichment and AP-1/RUNX1 co-occupancy at *de novo* RUNX1 bound sites

To further analyze the genomic characteristics underlying RUNX1 recruitment following TPA treatment, we examined the sequence compositions comprising *de-novo* RUNX1 occupancy regions. Motif analysis revealed that while RUNX motif was enriched at *de-novo* occupied regions, the motif alone was not sufficient to distinguish them from the similarly enriched constitutive sites. On the other hand, the analysis revealed a highly specific enrichment of the AP-1 motif (TGACTCA) at the *de-novo* RUNX1 sites (Figure 5A). Moreover, the estimated binding energy at the AP-1 motifs was positively correlated with differential ChIP-seq occupancy of RUNX1 in TPA treated versus non-treated cells (Figure 5A left), in contrast to the lack of such correlation to the binding energy of the RUNX motif itself (Figure 5A right)**.**

In addition, co-occurrence analysis revealed a significant coupling between AP-1 and RUNX motifs at *de-novo* RUNX1 occupancy regions (Supplementary Figure S4) and between AP-1 motif and RUNX1 binding to regions lacking proximal RUNX motif (Supplementary Figure S5). Importantly, using ChIP-seq we also demonstrated RUNX1/AP-1 co-occupancy of the FOS AP-1 component in K562-TPA cells (Supplementary Figure S4), found that FOS ChIP-seq peaks were highly enriched with an AP-1 motif (Supplementary Figure S6) and confirmed their significant co-occurrence with RUNX1 sites (Figure 5B). All in all, following induction FOS occupancy was found to be in high correlation with RUNX1 recruitment (Figure 5B and 5C).

Next we explored the nature of RUNX1/AP-1 co-occupancy by analyzing the relations between RUNX1 and AP-1 binding and their DNA motifs. For this purpose we used a data set combining the ENCODE-derived cFOS ChIP-seq occupancy in untreated K562 [26] and our FOS and FOSB ChIP-seq data in K562-TPA cells (Figure 5D). RUNX1 and AP-1 bound sites were highly enriched for their respective motifs.

However, *de-novo* RUNX1 bound sites lack RUNX motifs when recruited to constitutive AP-1 sites (group VIII in Figure 5D, only 5% have the motif compared to 25% of the stand alone RUNX1 sites). Conversely, *de-novo* AP-1 bound sites have a marked reduction in AP-1 motifs when recruited to constitutive RUNX1 sites (group VI in Figure 5D, 20% have the motif compared to 50% in stand alone sites). In joint AP-1/RUNX1 binding sites (either constitutive or *de-novo*), both motifs are enriched but to a lesser degree. According to this analysis RUNX1 and AP-1 are capable of recruiting each other to target sites. This conclusion is supported by finding that the two TFs physically interact [16]. Following TPA induction, levels of both TFs increased and facilitated *de novo* recruitment of AP-1/RUNX1 complexes to H3K4me1 marked sites, either new or previously occupied by only one of them (Supplementary Figure S11).

Interestingly, the finding that post TPA treatment RUNX1 bound to H3K4me1 marked regions upstream of *FOS, FOS-B* and *JUN* (Figure 5E) and up-regulated their expression (Figure 5F and Supplementary Table S1) raised the possibility that TPA induction triggered a regulatory cascade in which RUNX1 up-regulated *AP-1* expression, thereby facilitating recruitment of RUNX1-AP-1 modules to a new set of target genes.

Enrichment of ETS TF motif proximal to RUNX1 bound sites in CMK cells

The commonly used megakaryoblastic cell line CMK [35,36] is considered more differentiated than K562 as it expresses late markers of megakaryocytes and platelet differentiation [35,37]. Using this cell line we employed RUNX1 ChIP-seq to further address the plasticity of RUNX1 occupancy during megakaryopoiesis. Analysis revealed a substantial overlap between sites bound by RUNX1 in CMK and K562 cells, but also identified a significant number (~7000) of CMK-specific RUNX1 occupied sites (Figure 6A and Supplementary Figure S7). Sequence analysis revealed ETS TF motifs in close proximity to CMK specific RUNX1 bound sites, in clear distinction from the K562 sites (Figure 6B). ETS family members were previously shown to cooperate with RUNX1 [38-41]. Interestingly, analysis of RUNX1 occupancy patterns in loci of several genes expressed in CMK, revealed differential binding of RUNX1 to two ETS TFs, *ETS1* and *FLI1* in CMK compared to K562 cells (Figure 6C). Differential binding of RUNX1 in proximity to *PIK3R5/6* and *RAB27b* genes

was also noted (Figure 6C). These genes are known to play role in late stages of megakaryopoiesis and platelets formation [42,43],

To derive unbiased information regarding the relationship between sequence motifs and different modules of RUNX1 binding, we systematically calculated the fold enrichment of each motif associated with RUNX1 occupancy in the different binding modules (Figure 6C). The results correspond well to the experimental data indicating a common prevalence of RUNX motif in all classes and additional motifs, of RUNX1-cooperating TFs, including GATA, AP-1 and ETS that were biased towards class specificity. Importantly, their enrichment varied according to megakaryocytic differentiation stages; GATA at K562 constitutive sites, AP-1 at TPA-induced sites, and ETS at CMK-specific sites.

Interestingly, when RUNX1 ChIP-seq data for Jurkat T cells [40] was included in the co-occurrence analysis it was found that AP-1 motif was significantly under-represented, whereas a pronounced enrichment for the motif of TF PBX1B (GATGTG) was noted (Supplementary Figure S8) [44,45], raising the possibility that in T cells, RUNX1 also cooperates with PBX1B. Comparison of the overall RUNX1 binding profile showed that the highest overlap with T cells was found among CMK ChIP-seq data (Supplementary Figure S9).

We next assessed the functional cooperation between RUNX1 and its collaborating TFs using reporter assays in megakaryocytic cell lines. Regulatory regions of several biologically relevant RUNX1 target genes -identified as co-occupied by our Chip-seq experiments- were cloned into reporter constructs and tested (Figure 7). *HEMGN* promoter was co-activated by RUNX1 and GATA1 in non-induced K562 cells (Figure 7A), while intronic regulatory region of ITGB3 conferred RUNX1-AP-1 dependent reporter expression in K562-TPA cells (Figure 7B) and ITPR1 regulatory region, which was bound by RUNX1 at various differentiation stages, was cooperatively activated by RUNX1 and ETS TFs in Meg01 cells (Figure 7C). Collectively, the complementary outcome of these functional assays and the ChIP-seq occupancy data show that at different stages of megakaryocytic cell line differentiation, RUNX1 sequentially cooperates with GATA1, AP-1 and ETS TFs to drive the transcription program (Figure 6D).

Discussion

Cellular differentiation progresses through a cascade of coordinated transcriptional events involving dynamic interplay between TFs and epigenetic changes. This interplay affects chromatin structure and regulates gene expression by permitting or restricting transcription. Here we studied the mechanisms by which RUNX1 interacts with cellular genomic DNA sequences and the epigenomic makeup to regulate megakaryocytic transcriptional program. The data indicate that RUNX1 functions as key regulator mediating the differentiation process through stage-dependent cooperation with other TFs.

RUNX1 plays a pivotal role in megakaryopoiesis

Using differentiating megakaryocytic cell-line models we provided for the first time a systematic genome-wide flowchart of the RUNX1 occupancy patterns and regulatory targets in differentiating megakaryocytic cell lines. Our findings identified hundreds of previously unknown RUNX1 target genes based on their RUNX1-dependent response and its recruitment to sites proximal to their TSSs. The results delineated the molecular events underlying RUNX1 site selection specificity and its cooperation with co-regulators and underscored the pivotal role of RUNX1 in executing the megakaryopoietic gene expression program.

RUNX1-mediated gene expression is regulated by interactions of epigenetically preprogrammed enhancers and target promoters

TPA-induced gene expression in K562 cells was largely dependent on RUNX1, attesting to its crucial role in megakaryocytic differentiation. Interestingly, 24h post treatment, RUNX1 binds to numerous new occupancy sites. Analysis of H3K4me1 pattern prior and 24h post TPA treatment, demonstrated that the *de-novo* RUNX1 binding regions were pre-programmed with open chromatin prior to activation. This finding placed RUNX1 at the center of an ordered cell differentiation process in which the epigenomic landscape was pre-organized to meet subsequent regulatory requirements. The detailed RUNX1 genome-wide occupancy and associated gene expression patterns provided new important insights on the interaction of RUNX1 with the epigenome. RUNX1 binds preferentially to regions remote from its target genes. This was demonstrated by the correlation between the differential RUNX1 binding at target gene promoters and nearby enhancers and the better correspondence

between RUNX1 binding and RUNX motifs at remote enhancers compared to promoters.

GATA1, AP-1 and ETS emerge as key RUNX1 cooperators in megakaryopoiesis

Sequence analysis of regions occupied by RUNX1 prior to induction of megakaryocytic differentiation program indicated enrichment for the RUNX-GATA motif pair, suggesting that RUNX1 and GATA1 might cooperate during early stages of megakaryopoiesis. Experimental evidence in favor of this conclusion was granted by analysis of the recently reported data of GATA1 occupancy in K562 cells [26] GATA1 ChIP-seq analysis in K562-TPA cells and ChIP-PCR of RUNX1 target genes. These findings pertain to the possibility that impaired RUNX1-GATA1 cooperation in early FL hematopoiesis resulted in increased proliferation of megakaryocytic progenitors and contributes to the subsequent development of DS-AMKL. The data presented here underscores the involvement of GATA1 in regulating gene expression of several hematopoietic lineages, as was recently highlighted by the comprehensive analysis of its genome-wide occupancy in differentiating erythroid cells [31-33,46].

TPA induced, *de-novo* occupied RUNX1 sites were highly enriched in AP-1 sequence motifs, whereas constitutive RUNX1 sites were not. ChIP-seq and sequence analysis demonstrated coupling between RUNX1 and AP-1 binding and have indicated that constitutively bound RUNX1 recruited AP-1 to regions lacking AP-1 motifs while constitutively bound AP-1 recruited RUNX1 to sites lacking RUNX motifs. Interestingly, in the K562-TPA system, RUNX1 up-regulated *AP-1* genes, raising the intriguing possibility of a regulatory cascade in which RUNX1-mediated expression of AP-1 acts to facilitate its binding to a new set of target genes required for megakaryocytic differentiation (Figure 6D).

The CMK cell line expresses late megakaryocytic markers including platelet peroxidase and glycoprotein IIb/IIa [35,37] and is therefore considered more differentiated than 48h TPA treated K562 cells. Interestingly, analysis of RUNX1 occupancy regions in CMK cells revealed a pronounced enrichment of RUNX-ETS motif pair. This finding suggested that in CMK cells RUNX1 cooperated with members of the ETS TF family to drive megakaryocytic gene expression. Of note, both GATA1 and ETS family members were previously shown to cooperate with RUNX1 in various *in vitro* and cell transfection assays [9,12,39-41]. On the other hand, to

the best of our knowledge this is the first time that AP-1 is implicated as a major transcriptional collaborator of RUNX1, although its involvement in megakaryocyte differentiation was previously reported [47,48].

The methodologies introduced here may facilitate evaluation of RUNX1 function in other differentiation programs underlying the etiology of 8;21 leukemia as well as provide insights into the mechanisms underlying TF-DNA interaction specificity. The data suggest that genomic interactions of RUNX1 are highly dynamic, and specified by a combination of genomic inputs, which include RUNX binding sites and cell type-specific epigenetic makeup such as H3K4me1 marked enhancers, and protein-protein interactions with other sequence specific TFs. Importantly, the data show that the interaction of RUNX1 with its cooperating TFs is critically important for determining its occupancy profiles at different developmental stages. Protein-protein interactions of TFs generate an additional layer of complexity superimposed on the genomic sequence and epigenetic makeup, thereby enhancing the diversity of RUNX1 binding landscapes, and the repertoire of its regulated genes. The generality of this phenomenon will be clarified as more genomic occupancy data become available. A model delineating the plasticity and spatial dynamics of RUNX1 occupancy and interactions during the three-stage differentiation process is shown in Figure 6D.

Acknowledgments

We thank Dr. Daniela Amann-Zalcernstein and Dr. Shirely Horn-Saban for help in Ilumina sequencing, Dalia Goldenberg for technical assistance and Dr. Joseph Lotem and Dr. Ditsa Levanon for helpful comments and discussions throughout the work. This work was supported by grants from the European Union AnEUploidy project, the Jerome Lejeune Foundation and The Israel Science Foundation to YG as well as by Alon Fellowship and the Israel Science Foundation grant to AT.

Author Contributions

Conceived and designed the experiments: NP, AT, YG. Performed the experiments: NP. Analyzed the data: RJ, AT, NP, YG. Wrote the paper: NP, RJ, AT, YG.

Conflict of Interest Disclosure: the authors declare no competing financial interests.

References

1. Levanon D, and Groner Y. Structure and regulated expression of mammalian RUNX genes. *Oncogene*. 2004;23:4211-4219.
2. de Bruijn MF, Speck NA. Core-binding factors in hematopoiesis and immune function. *Oncogene*. 2004;23(24):4238-4248.
3. Look AT. Oncogenic transcription factors in the human acute leukemias. *Science*. 1997;278(5340):1059-1064.
4. Speck NA, Gilliland DG. Core-binding factors in haematopoiesis and leukaemia. *Nat Rev Cancer*. 2002;2(7):502-513.
5. Song WJ, Sullivan, M.G., Legare, R.D., Hutchings, S. et al. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukamia. *Nat Genet*. 1999;23(2):166-175.
6. Growney JD, Shigematsu H, Li Z, et al. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood*. 2005;106(2):494-504.
7. Asou N. The role of a Runt domain transcription factor AML1/RUNX1 in leukemogenesis and its clinical implications. *Crit Rev Oncol Hematol*. 2003;45(2):129-150.
8. Lange B. The management of neoplastic disorders of haematopoiesis in children with Down's syndrome. *Br J Haematol*. 2000;110(3):512-524.
9. Goldfarb AN. Megakaryocytic programming by a transcriptional regulatory loop: A circle connecting RUNX1, GATA-1, and P-TEFb. *J Cell Biochem*. 2009;107(3):377-382.
10. Goldfarb AN. Transcriptional control of megakaryocyte development. *Oncogene*. 2007;26(47):6795-6802.
11. Ben-Ami O, Pencovich N, Lotem J, Levanon D, Groner Y. A regulatory interplay between miR-27a and Runx1 during megakaryopoiesis. *Proc Natl Acad Sci U S A*. 2009;106(1):238-243.
12. Elagib KE, Racke FK, Mogass M, Khetawat R, Delehanty LL, Goldfarb AN. RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood*. 2003;101(11):4333-4341.
13. Ichikawa M, Asai T, Saito T, et al. AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat Med*. 2004;10(3):299-304.
14. Wilson NK, Miranda-Saavedra D, Kinston S, et al. The transcriptional program controlled by the stem cell leukemia gene Scl/Tal1 during early embryonic hematopoietic development. *Blood*. 2009;113(22):5456-5465.
15. Farnham PJ. Insights from genomic profiling of transcription factors. *Nat Rev Genet*. 2009;10(9):605-616.
16. Ravasi T, Suzuki H, Cannistraci CV, et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*;140(5):744-752.
17. Butler TM, Ziemiecki A, Friis RR. Megakaryocytic differentiation of K562 cells is associated with changes in the cytoskeletal organization and the pattern of chromatographically distinct forms of phosphotyrosyl-specific protein phosphatases. *Cancer Res*. 1990;50(19):6323-6329.
18. Ainbinder E, Revach M, Wolstein O, Moshonov S, Diamant N, Dikstein R. Mechanism of rapid transcriptional induction of tumor necrosis factor alpha-responsive genes by NF-kappaB. *Mol Cell Biol*. 2002;22(18):6354-6362.

19. Aziz-Aloya RB, Levanon D, Karn H, et al. Expression of AML1-d, a short human AML1 isoform, in embryonic stem cells suppresses in vivo tumor growth and differentiation. *Cell Death Differ*. 1998;5(9):765-773.
20. Heintzman ND, Stuart RK, Hon G, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 2007;39(3):311-318.
21. Splinter E, Heath H, Kooren J, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*. 2006;20(17):2349-2354.
22. Fullwood MJ, Liu MH, Pan YF, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*. 2009;462(7269):58-64.
23. Rabinovich A, Jin VX, Rabinovich R, Xu X, Farnham PJ. E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. *Genome Res*. 2008;18(11):1763-1777.
24. Yang A, Zhu Z, Kapranov P, et al. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol Cell*. 2006;24(4):593-602.
25. Tanay A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res*. 2006;16(8):962-972.
26. Rhead B, Karolchik D, Kuhn RM, et al. The UCSC Genome Browser database: update 2010. *Nucleic Acids Res*;38(Database issue):D613-619.
27. Putz G, Rosner A, Nuesslein I, Schmitz N, Buchholz F. AML1 deletion in adult mice causes splenomegaly and lymphomas. *Oncogene*. 2006;25(6):929-939.
28. Chang AN, Cantor AB, Fujiwara Y, et al. GATA-factor dependence of the multitype zinc-finger protein FOG-1 for its essential role in megakaryopoiesis. *Proc Natl Acad Sci U S A*. 2002;99(14):9237-9242.
29. Nichols KE, Crispino JD, Poncz M, et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet.* 2000;24(3):266-270.
30. Vyas P, Ault K, Jackson CW, Orkin SH, Shivdasani RA. Consequences of GATA-1 deficiency in megakaryocytes and platelets. *Blood*. 1999;93(9):2867-2875.
31. Yu M, Riva L, Xie H, et al. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol Cell.* 2009;36(4):682-695.
32. Cheng Y, Wu W, Kumar SA, et al. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res.* 2009;19(12):2172-2184.
33. Fujiwara T, O'Geen H, Keles S, et al. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell*. 2009;36(4):667-681.
34. Rodriguez P, Bonte E, Krijgsveld J, et al. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *Embo J*. 2005;24(13):2354-2366.
35. Komatsu N, Suda T, Moroi M, et al. Growth and differentiation of a human megakaryoblastic cell line, CMK. *Blood*. 1989;74(1):42-48.
36. Tohyama Y, Tohyama K, Tsubokawa M, Asahi M, Yoshida Y, Yamamura H. Outside-In signaling of soluble and solid-phase fibrinogen through integrin alphaIIbbeta3 is different and cooperative with each other in a megakaryoblastic leukemia cell line, CMK. *Blood*. 1998;92(4):1277-1286.

37. Adachi M, Ryo R, Sato T, Yamaguchi N. Platelet factor 4 gene expression in a human megakaryocytic leukemia cell line (CMK) and its differentiated subclone (CMK11-5). *Exp Hematol*. 1991;19(9):923-927.
38. Giese K, Kingsley C, Kirshner JR, Grosschedl R. Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev*. 1995;9(8):995-1008.
39. Arman M, Aguilera-Montilla N, Mas V, et al. The human CD6 gene is transcriptionally regulated by RUNX and Ets transcription factors in T cells. *Mol Immunol*. 2009;46(11-12):2226-2235.
40. Hollenhorst PC, Chandler KJ, Poulsen RL, Johnson WE, Speck NA, Graves BJ. DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet*. 2009;5(12):e1000778.
41. Goetz TL, Gu TL, Speck NA, Graves BJ. Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. *Mol Cell Biol*. 2000;20(1):81-90.
42. Gibbins JM, Briddon S, Shutes A, et al. The p85 subunit of phosphatidylinositol 3-kinase associates with the Fc receptor gamma-chain and linker for activitor of T cells (LAT) in platelets stimulated by collagen and convulxin. *J Biol Chem*. 1998;273(51):34437-34443.
43. Tiwari S, Italiano JE, Jr., Barral DC, et al. A role for Rab27b in NF-E2-dependent pathways of platelet formation. *Blood*. 2003;102(12):3970-3979.
44. Bourette RP, Grasset MF, Mouchiroud G. E2a/Pbx1 oncogene inhibits terminal differentiation but not myeloid potential of pro-T cells. *Oncogene*. 2007;26(2):234-247.
45. Sykes DB, Kamps MP. E2a/Pbx1 induces the rapid proliferation of stem cell factor-dependent murine pro-T cells that cause acute T-lymphoid or myeloid leukemias in mice. *Mol Cell Biol*. 2004;24(3):1256-1269.
46. Soler E, Andrieu-Soler C, de Boer E, et al. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev*;24(3):277-289.
47. Rosson D, O'Brien TG. AP-1 activity affects the levels of induced erythroid and megakaryocytic differentiation of K562 cells. *Arch Biochem Biophys*. 1998;352(2):298-305.
48. Eriksson M, Arminen L, Karjalainen-Lindsberg ML, Leppa S. AP-1 regulates alpha2beta1 integrin expression by ERK-dependent signals during megakaryocytic differentiation of K562 cells. *Exp Cell Res*. 2005;304(1):175-186.

Figure legends

Figure 1. Phenotypic and gene expression response of K562 and K562^{Runx1KD} cells to TPA treatment. (A) qRT-PCR and Western blot analysis of RUNX1 expression in untreated (U.T) and TPA treated (TPA) cells as well as in TPA treated cells stably expressing shRNA-miR (K562^{RUNX1KD}), which knocked-down RUNX1 expression (KD), in comparison to non-silencing shRNA-miR control (Neg). qRT-PCR Data represent the mean ± SD of two independent experiments performed in

triplicates. Western blot of nuclear extract using anti RUNX1 antibodies. Emerin was used as control of protein loading (see Supplementary Methods). (B) Proliferation assay using K562-TPA Neg (red) and KD (blue) cells. 1×10^6 cells were grown in culture and counted every 24 hours. TPA was added 48h after seeding of cells. Data represent the mean ± SD of two independent experiments performed in triplicates. (C) Cell morphology changes. Representative microscopic images of May-Grunwald-Giemsa stained untreated U.T, TPA, TPA Neg and TPA KD K562 cells. (D) Expression of megakaryocytic markers in K562 (Control) and $K562^{RUNX1KD}$ (KD) in response to TPA. Q-RT-PCR analysis of ITGA2B (CD41), ITGB3 (CD61), and CD9 expression in K562 cells. Data represent the mean ±SD of two independent experiments performed in quadruplicates. The decreased expression in TPA treated $K562^{RUNX1KD}$ relative to K562 control was significant [$P<0.01$] for all three markers. Primers used for qRT-PCR assays are listed in supplemental table S2. (E) TPA-induced transcriptional changes. Genes are plotted based on their expression level (log scale) in K562- TPA cells vs. their basal level in untreated cells. Genes showing 2-fold (1 in log scale) increase or decrease in expression levels are indicated in green or red, respectively. Indicated are examples of upregulated genes known to play important role in megakaryocytic differentiation along with the reference TRPV6 gene. (F) RUNX1 knockdown impaired TPA-induced transcriptional activation. Gene expression levels in K562-TPA vs. $K562^{Runx1KD}$-TPA cells, stably expressing a TPA responsive RUNX1 shRNA-miR are shown in Supplementary Table S1. Note the change in RUNX1-dependent expression of genes indicated in (E). (G) Change in RUNX1-dependent transcriptional regulation in TPA treated cells. Genes were divided into groups according to differential expression following TPA treatment. Shown are boxplots representing the expression change distribution in $K562^{Runx1KD}$ cells (blue) and $K562^{CONTROL}$ (red) for each group of genes. In RUNX1 expression KD, transcriptional repression (left section) did not change, whereas transcriptional activation (right section) was almost completely abolished.

Figure 2. TPA-dependent recruitment of RUNX1 to preprogrammed remote enhancer regions. (A) Validation of anti RUNX1 antibodies efficacy and specificity in immunoprecipitation (IP). IP of RUNX1 from CMK, K562 or K562-TPA whole cell lysates by the anti-RUNX1 antibodies used in the ChIP-seq experiments. 1-

Protein-A agarose beads (Supplementary Methods) were incubated without cell lysate. 2- Western blot using 50ug protein of cell lysate and antiRUNX1 antibodies (α RUNX1). (2*) longer exposure of lane 2 . 3 to 5- Western blot analysis of proteins immunoprecipitated from CMK (2), K562 (3) and K562-TPA cell lysates. 6- Western blot analysis of CMK cell lysate using anti GAPDH antibodies (α GAPDH). 7- Western blot analysis, using anti GAPDH antibodies, of proteins immunoprecipitated from CMK by anti RUNX1 antibodies. (B) A Venn diagram showing genomewide RUNX1 occupancy in K562 and K562-TPA cells. RUNX1 binding peaks from ChIP-seq analysis (covered beyond a threshold and lacking significant coverage in control experiment) before and 24 hours after TPA treatment were compared. (C) Distribution of RUNX1 ChIP-seq peaks relative to the TSS. Shown are RUNX1 peak frequencies relative to the distance from the nearest annotated TSS. While regions proximal to TSSs show increased RUNX1 binding, the majority of RUNX1 peaks (inset) were localized more than 10 kb from the nearest TSS. (D) Enrichment of RUNX1 binding in regions with open chromatin. Distributions of H3K4me1 ChIP-seq readouts at RUNX1 bound sites (light green) and at background regions (gray) demonstrate that RUNX1 binding was largely restricted to H3K4me1 marked remote enhancers. (E) Change in H3K4me1 marked regions following TPA induction. Histograms depict the numbers of ChIP-seq H3K4me1 marked genomic regions that were identified in both pre- and post- TPA treated cells (constitutive, gray), in pre-TPA treated cells (green) and in TPA-induced cells alone (blue). (F) *De-novo* occupied RUNX1 sites are marked by H3K4me1 prior to TPA induction. Histograms depict the distributions of H3K4me1 ChIP-seq readouts before (light green) and after (blue) TPA in de-novo occupied RUNX1 sites. The distribution of H3K4me1 levels at background regions (gray) served as a control. The data demonstrate that *de-novo* RUNX1 occupied regions were marked open by H3K4me1 prior to TPA induction and before RUNX1 binding. **(**G & H) RUNX1 binding is negatively correlated with H3K27me3 ChIP-seq occupancy. A RUNX1 enrichment value is computed by dividing the frequency of RUNX1 peaks over defined part of the genome by the RUNX1 peak frequency over the entire genome. Plotted are RUNX1 enrichment values as a function of H3K27me3 occupancy, before (G) and after (H) TPA induction.

Figure 3. **RUNX1-dependent transcriptional activation is linked to remote *de-novo* RUNX1 occupancy sites.** (A) Enrichment of *de novo* RUNX1 binding in proximity of activated genes. The density (peaks/20kb, y axis) of de-novo RUNX1 bound sites is plotted relative to TSSs for 147 *bone fide* RUNX1-regulated genes following TPA induction (blue) and for genes not activated by TPA (gray). (B) De novo binding of RUNX1 at long-range enhancer regions significantly correlates with its *de novo* binding at gene promoters. Log of maximal difference in RUNX1 ChIP-seq coverage before and after TPA within 3kb (promoter region) and 200kb (enhancer region) around the TSS was computed. Boxplots represent the distribution of differential RUNX1 occupancy at enhancers (Y axis) for groups of promoters with similar differential RUNX1 occupancy (X axis). The outliers represent values >90th percentile. (C) Remote constitutive and *de-novo* RUNX1 occupied sites and H3K4me1/H3K27me3 profile at key TPA-responsive RUNX1-regulated megakaryocytic gene loci. RUNX1, H3K4me1, and H3K27me3 ChIP-seq readouts, before (green) and after (blue) TPA treatment, in several genomic loci encompassing megakaryocytic important genes that were activated following TPA in a RUNX1 dependent manner. (D) Quantitative evaluation of RUNX1 ChIP-seq results in K562, K562-TPA cells and murine primary megakaryocytes. Q-PCR analysis of RUNX1 binding to regions spanning *ITGB3*, *VEGFA* and *CTNNB* indicated in (C). Data represent the mean ± SD of two independent ChIP-qPCR experiments performed in quadruplicates using K562 cells (green), K562-TPA cells (blue) and murine primary megakaryocytes (gray). Red asterisks mark regions containing RUNX motifs. Primers used for qPCR assays are listed in supplemental table S2 and details described in Supplementary Methods.

Figure 4. RUNX1 and GATA1 are over-represented at constitutive RUNX1 occupied regions. (A) RUNX motif binding energy is correlated with RUNX1 ChIP-seq occupancy. Indicated RUNX motif was inferred directly from RUNX1 occupancy peaks (M & M). Boxplots depict the distributions of motif binding energies (Y-axis) in groups of regions with increasing RUNX1 ChIP-seq coverage (X-axis). Only regions marked with H3K4me1 were considered. The regions that passed the threshold and were thus defined as RUNX1 bound peaks are shown in green boxplots.

Data is plotted separately for promoters (<3kb from TSS, upper panel) and enhancers (>3kb from TSS, lower). (B) GATA binding motif is correlated with RUNX1 ChIP-seq occupancy. A GATA motif was inferred directly from RUNX1 constitutively bound peaks. The correspondence between RUNX1 occupancy and GATA motif binding energies is presented as in (A) above, (but using a GATA motif model instead of a RUNX motif model). Data reveal a correlation between RUNX1 ChIP-seq occupancy and the intensity of GATA motifs. Boxplots depict the distributions of motif binding energies (Y-axis) in groups of H3K4me1-marked regions with increasing RUNX1 ChIP-seq readout coverage (X-axis). (C) A Venn diagram summarizing the overlap between RUNX1 and GATA1 bound sites in K562. (D-F) Co-occurrence analysis of RUNX1 and GATA1. Shown are distributions of distances from RUNX1 bound sites to the nearest GATA motif (D), to the nearest GATA1 ChIP-seq peak in K562 cells (E) and to the nearest GATA1 ChIP-seq peak in K562-TPA cells (F). Distances distributions were computed separately for constitutively occupied RUNX1 sites (light green), *de novo* occupied RUNX1 sites (blue), or RUNX1 motifs in H3K4me1 marked regions without significant RUNX1 occupancy (gray).

Figure 5. Co-occurrence of RUNX1 and AP-1 at *de-novo* RUNX1 bound sites and RUNX1-mediated increase of *AP-1* in response to TPA. (A) RUNX1 and AP1 motifs were inferred directly from de novo RUNX1 bound sites. Distributions of differential (post vs. pre TPA treatment) RUNX1 ChIP-seq readouts are plotted for groups of regions with increasing binding energy of RUNX (left) or AP-1 (right) motifs. Unlike RUNX motifs, AP-1 motifs are significantly correlated with differential (±TPA) RUNX1 occupancy. (B) Correlation between differential occupancy of RUNX1 and AP-1 occupancy. Shown are the distributions of increased RUNX-1 occupancy (Y-axis; post Vs. pre TPA treatment) in groups of loci with range of AP-1 occupancy level post-TPA treatment (X axis). AP-1 occupancy is a good predictor for TPA induced RUNX1 recruitment. (C) Co-occurrence of AP-1 and RUNX1 chip-seq peaks at *de-novo* RUNX1 occupied sites. Shown are the distributions of distances between RUNX1 peaks and most proximal AP-1 peaks at constitutively (light green) or *de-novo* (+TPA blue) RUNX1 sites. For a reference, the distribution of distances for the nearest AP-1 sites from random H3K4me1 enhancers lacking RUNX1 binding is presented (gray). (D) Combinatorial analysis of RUNX1

and AP-1 sites and motifs. The relative fold enrichment of RUNX (orange) and AP-1 (green) motifs relative to the genomic background are shown for groups (I to IX) of loci with various combinatorial ChIP-seq readouts of the two factors. Constitutive sites are those observed in K562 cells both pre- and post-TPA treatment. *De-novo* sites are those observed only at post-TPA treatment. The pattern depicted in groups VI and VIII suggest that RUNX1 occupancy enables AP-1 recruitment to regions lacking AP-1 motif, and vice versa, AP-1 binding enables recruitment of RUNX1 to sites lacking RUNX motifs. (E) TPA dependent binding of RUNX1 to three *AP-1* genes. RUNX1 ChIP-seq tracks at loci encompassing *FOS, FOSB* and *JUN* before (light green) and after (blue) TPA, in K562 cells. *de-novo* RUNX1 bound sites in remote enhancers are noted. (F) qRT-PCR analysis of *FOS, FOS-B* and *JUN* expression in K562 cells. Data represent the mean ± SD of three independent experiments performed in triplicates. The increased expression of *FOS, FOS-B* and *JUN* in K562-TPA relative to K562 and in K562-TPA with RUNX1 knockdown control (TPA neg) relative to K562 KD was significant [K562±TPA P = 0.01, 0.005,0.005, K562±KD P = 0.01, 0.001, 0.01].

Figure 6. **Proximity of RUNX1 occupied sites to ETS motifs in CMK cells and schematic illustration of regulatory interplay between RUNX1 and cooperating TFs.** (A) Venn diagram showing the relationships btween RUNX1 ChIP-seq occupancy profiles in K562, K562-TPA and CMK cells. A substantial number (~7000) of RUNX1 bound sites are CMK specific. (B) RUNX1 occupancy profile at several CMK specific loci. Shown are RUNX1 occupancy profiles in K562 cells before (light green) and after (blue) TPA, and in CMK cells (orange), at loci spanning the *ETS1, FLI1, PIK3R6 and RAB27b* genes. (C) A multi-modal RUNX1 occupancy landscape of distinct TF motif combinations characterizing the megakaryopoitic gene expression program. RUNX1 bound sites in K562, K562-TPA and CMK were grouped according to their co-occurrence with motifs of RUNX, GATA, AP-1 and ETS. Histograms show percentage of occurrence of motifs with binding energies in the top 5% of background. ETS motifs are highly prevalent in CMK cells while AP-1 motif is clearly biased to *de novo* occupied TPA-induced sites. (D) A schematic model summarizing our hypothesis about stage specific RUNX1-mediated regulation. RUNX1 (orange crescent) is preferentially bound to remote enhancers and cooperates

with GATA1 (blue cluster) to regulate early myeloid genes (orange rectangles). Upon induction, RUNX1 recruits co-activators (purple crescent) to activate the *AP1* genes. Thereafter, AP1 TFs (green clusters) facilitate the binding of RUNX1 to early megakaryocytic genes thereby launching and driving the differentiation program. At subsequent differentiation stage (CMK) RUNX1 cooperates with ETS family TFs (yellow ellipses) to activates a different set of megakaryocytic genes. This scenario underscores the notion that RUNX1 functions in a context dependent manner to regulate the transcriptional program in differentiating megakaryocytic cell lines.

Figure 7. Functional cooperation between RUNX1 and its partner TFs GATA1, AP-1 and ETS. Reporter assays in megakaryocytic cell lines demonstrate cooperation between RUNX1-GATA (A), RUNX1-AP-1 (B) and RUNX1-ETS (C) in regulation of gene expression of *HEMGN*, *ITGB3* and *ITPR1* genes. (A) Upper left: RUNX1 and GATA1 ChIP-seq tracks in K562 and K562-TPA cells proximal to the *HEMGN* locus. The regulatory region cloned in vectors and used in transfection assays is indicated by the red rectangle, as are the evolutionary conserved RUNX and GATA binding sites (lower left) located within this region. Upper right: ChIP-qPCR validation of RUNX1 and GATA1 binding to the indicated ChIP-seq region. Shown are independent ChIP assays followed by qPCR using K562 and K562-TPA cells. Lower right: Dual luciferase reporter assays in transfected K562 cells using PGL4.73 vector alone or with intact/mutated *HEMGN* regulatory region (see Supplementary Methods for details). (B) Upper left: FOS and RUNX1 ChIP-seq tracks proximal to the ITGB3 locus in K562 and K562-TPA cells. The regulatory region cloned in vectors and used in transfection assays is indicated by the red rectangle, as are the evolutionary conserved RUNX and AP-1 binding sites (lower left) located within this region. Upper right: ChIP-qPCR validation of RUNX1 and AP-1 binding to the indicated ChIP-seq region. Shown are independent ChIP assays followed by qPCR using K562 and K562-TPA cells. Lower right: Dual luciferase reporter assays in transfected K562 cells using PGL4.73 vector alone or with intact/mutated *ITGB3* regulatory region (see M & M for details). (C) Upper left: RUNX1 ChIP-seq tracks proximal to the ITRP1 locus in K562, K562-TPA and CMK cells. The regulatory region cloned in vectors and used in transfection assays is indicated by the red rectangle, as are the evolutionary conserved RUNX and ETS binding sites (lower left) located within this region. Upper right: ChIP-qPCR validation of RUNX1 binding to

the indicated ChIP-seq region. Shown are independent ChIP assays followed by qPCR using K562, K562-TPA and CMK cells. Lower right: Dual luciferase reporter assays in transfected K562 cells using PGL4.73 vector alone or with intact/mutated *ITPR1* regulatory region (see M & M for details). The data of qPCR and dual reporter assays shown in (A), (B) and (C) represent means ± SE of at least two biological repeats performed in triplicates.

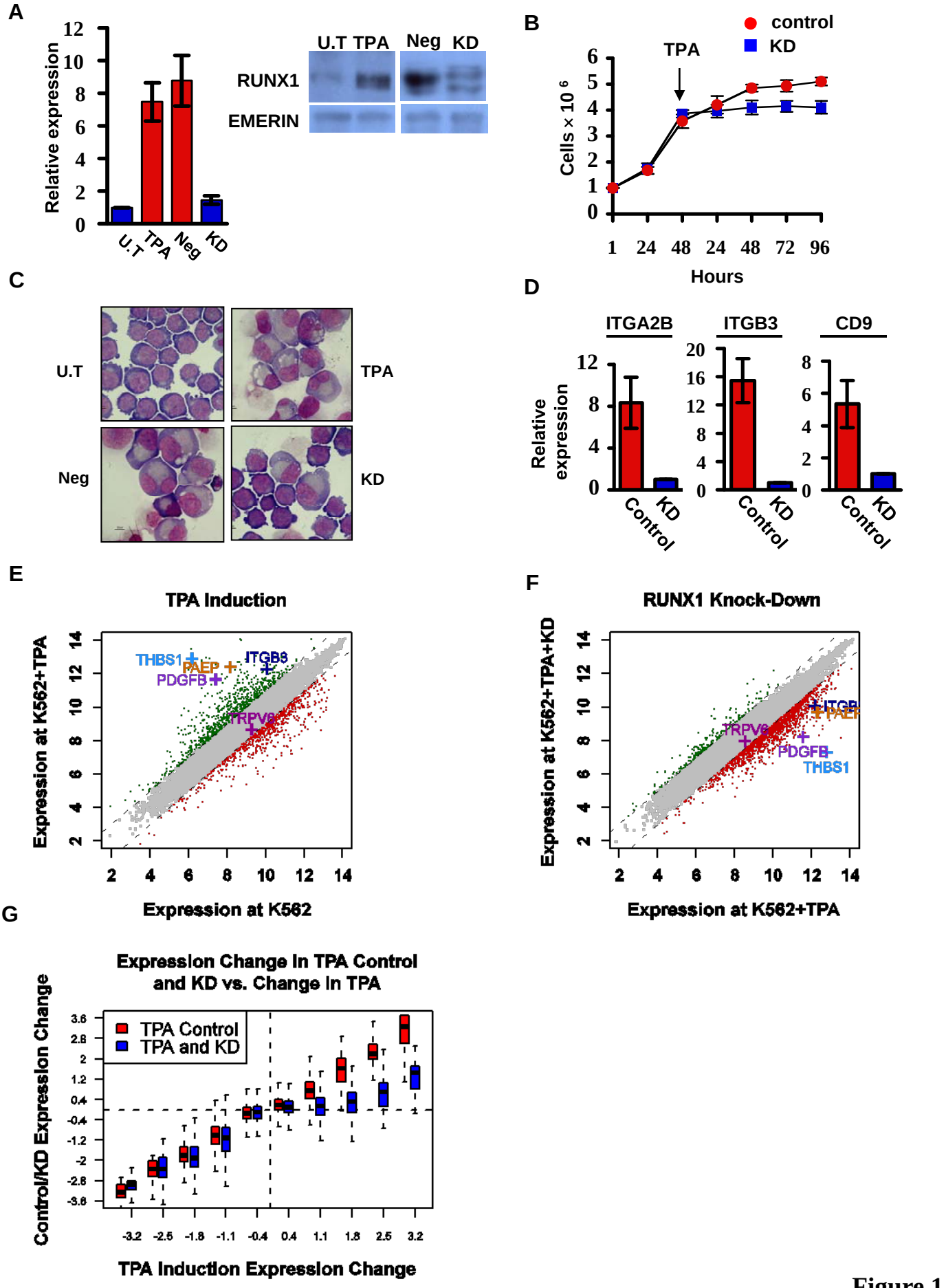

A
Relative expression
U.T
TPA
Neg
KD
U.T TPA
Neg KD
RUNX1
EMERIN
B
control
KD
TPA
Cells × 10 6
Hours
C
U.T
TPA
Neg
KD
D
ITGA2B
ITGB3
CD9
Relative expression
Control
KD
E
TPA Induction
Expression at K562+TPA
Expression at K562
THBS1
PAEP
PDGFB
ITGB3
TRPV6
F
RUNX1 Knock-Down
Expression at K562+TPA+KD
Expression at K562+TPA
G
Expression Change In TPA Control and KD vs. Change In TPA
Control/KD Expression Change
TPA Control
TPA and KD
TPA Induction Expression Change

Figure 1

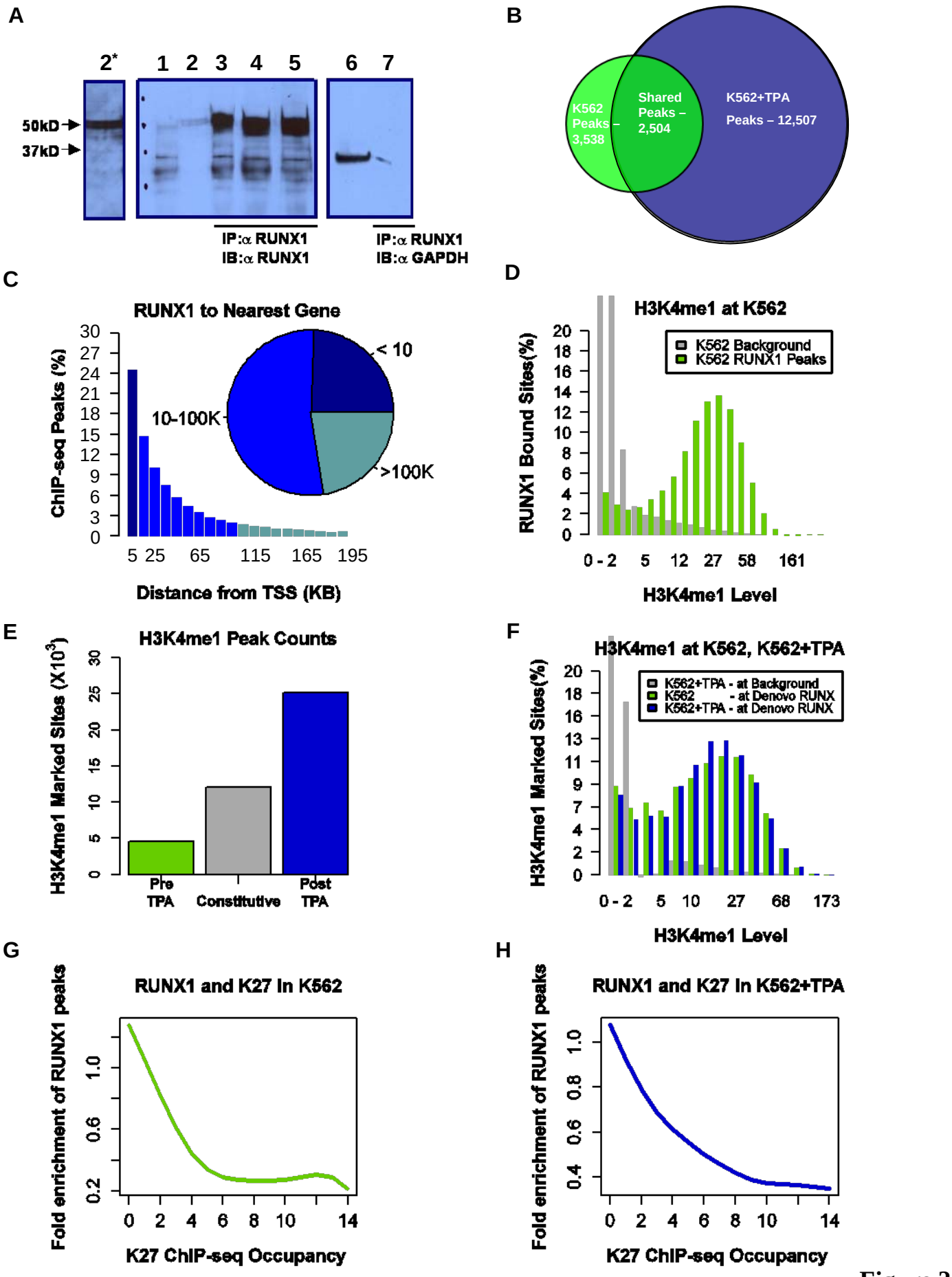

A
2*
1 2 3 4 5
6 7
50kD
37kD
IP:α RUNX1
IB:α RUNX1
IP:α RUNX1
IB:α GAPDH
B
K562 Peaks – 3,538
Shared Peaks – 2,504
K562+TPA Peaks – 12,507
C
RUNX1 to Nearest Gene
ChIP-seq Peaks (%)
< 10
10-100K
>100K
5 25 65 115 165 195
Distance from TSS (KB)
D
H3K4me1 at K562
K562 Background
K562 RUNX1 Peaks
RUNX1 Bound Sites(%)
0 - 2 5 12 27 58 161
H3K4me1 Level
E
H3K4me1 Peak Counts
H3K4me1 Marked Sites (X10^3)
Pre TPA
Constitutive
Post TPA
F
H3K4me1 at K562, K562+TPA
K562+TPA - at Background
K562 - at Denovo RUNX
K562+TPA - at Denovo RUNX
H3K4me1 Marked Sites(%)
0 - 2 5 10 27 68 173
H3K4me1 Level
G
RUNX1 and K27 In K562
Fold enrichment of RUNX1 peaks
K27 ChIP-seq Occupancy
H
RUNX1 and K27 In K562+TPA
Fold enrichment of RUNX1 peaks
K27 ChIP-seq Occupancy

Figure 2

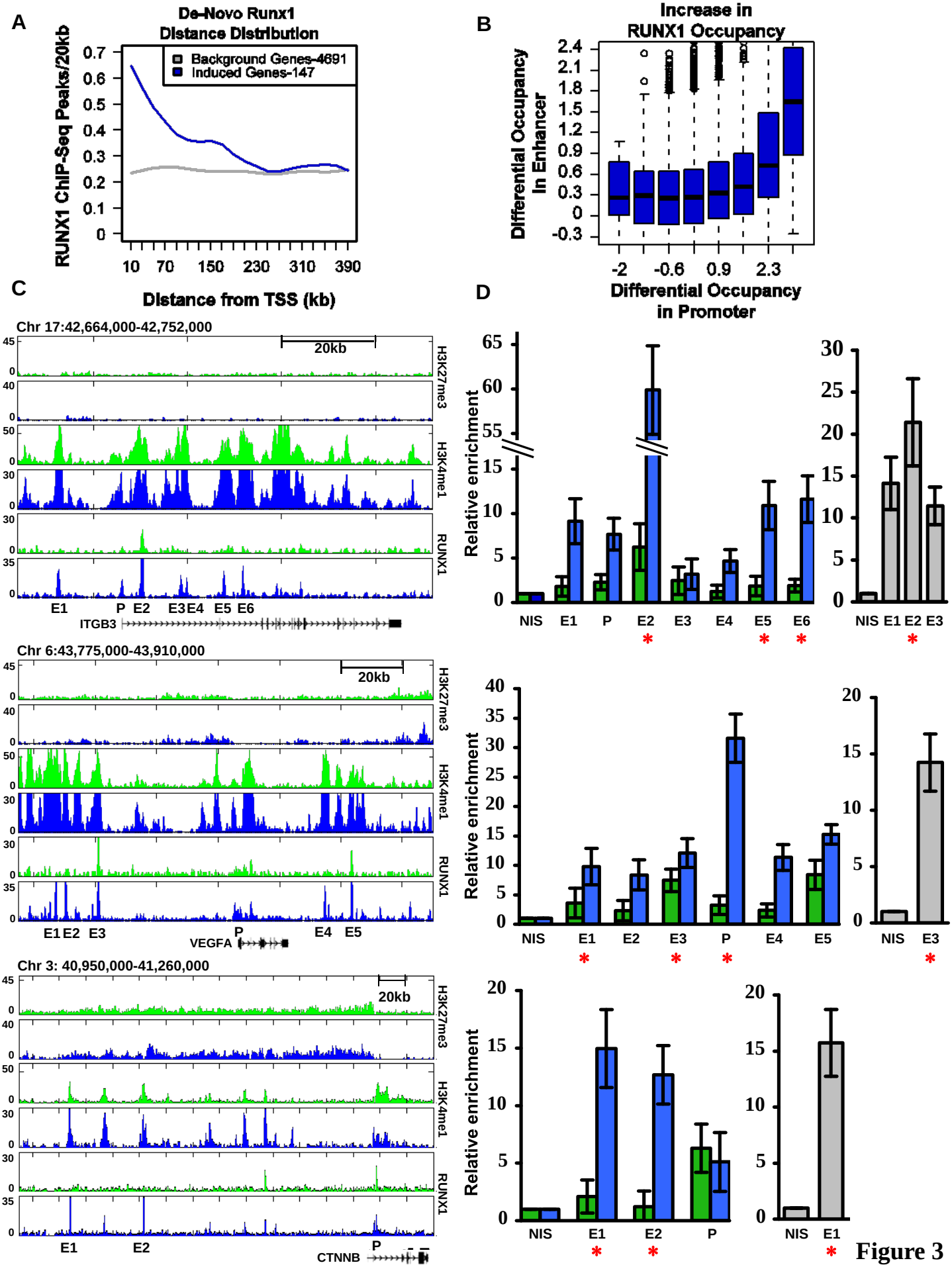
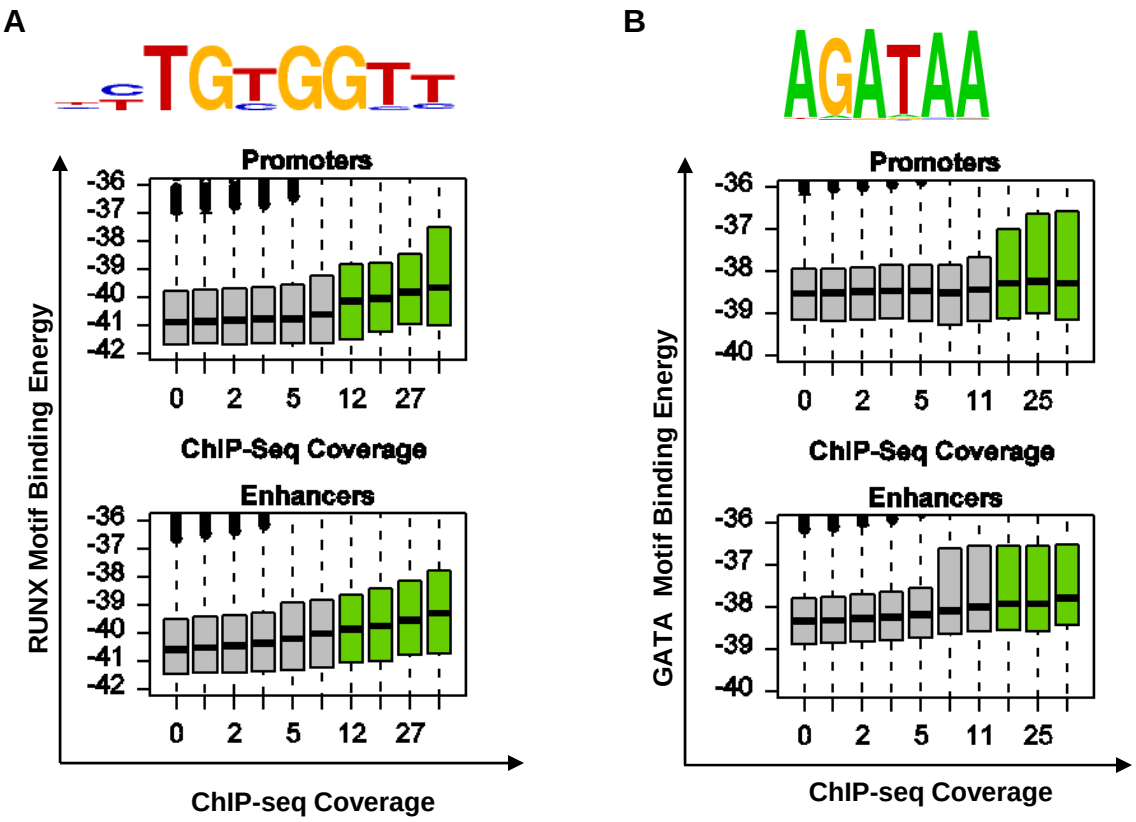
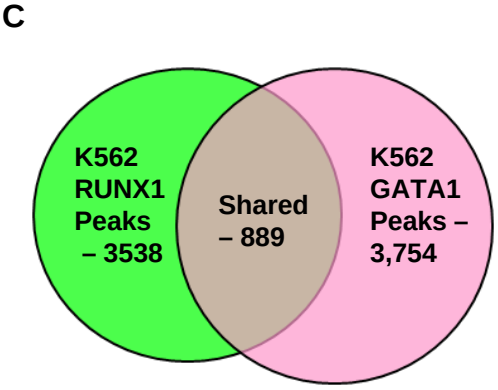
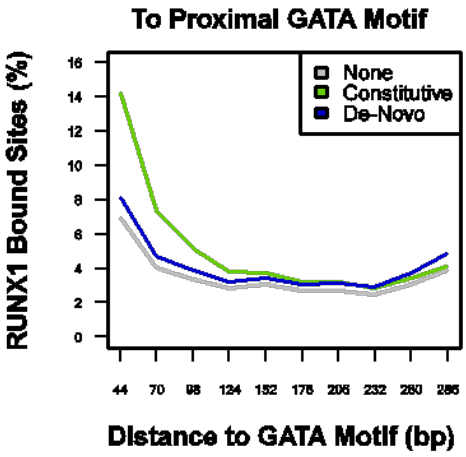
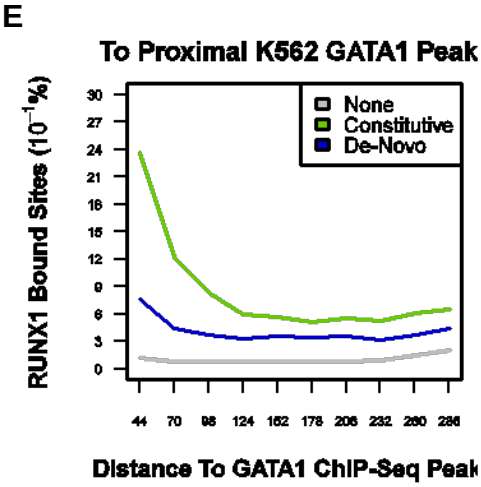

A
De-Novo Runx1
Distance Distribution
Background Genes-4691
Induced Genes-147
RUNX1 ChIP-Seq Peaks/20kb
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10 70 150 230 310 390
B
Increase in
RUNX1 Occupancy
Differential Occupancy
In Enhancer
2.4
2.1
1.8
1.5
1.2
0.9
0.6
0.3
0
-0.3
-2 -0.6 0.9 2.3
Differential Occupancy
in Promoter
C
Distance from TSS (kb)
Chr 17:42,664,000-42,752,000
20kb
H3K27me3
H3K4me1
RUNX1
E1 P E2 E3E4 E5 E6
ITGB3
Chr 6:43,775,000-43,910,000
20kb
H3K27me3
H3K4me1
RUNX1
E1E2 E3
P
E4 E5
VEGFA
Chr 3: 40,950,000-41,260,000
20kb
H3K27me3
H3K4me1
RUNX1
E1
E2
P
CTNNB
D
Relative enrichment
NIS E1 P E2 E3 E4 E5 E6
NIS E1 E2 E3
NIS E1 E2 E3 P E4 E5
NIS E1 E3
NIS E1 E2 P
NIS E1

Figure 3


A
B
Promoters
Enhancers
RUNX Motif Binding Energy
GATA Motif Binding Energy
ChIP-Seq Coverage
ChIP-seq Coverage
-36
-37
-38
-39
-40
-41
-42
0 2 5 12 27
0 2 5 11 25


C
K562 RUNX1 Peaks – 3538
Shared – 889
K562 GATA1 Peaks – 3,754


D
To Proximal GATA Motif
RUNX1 Bound Sites (%)
None
Constitutive
De-Novo
Distance to GATA Motif (bp)


E
To Proximal K562 GATA1 Peak
RUNX1 Bound Sites (10⁻¹%)
None
Constitutive
De-Novo
Distance To GATA1 ChIP-Seq Peak

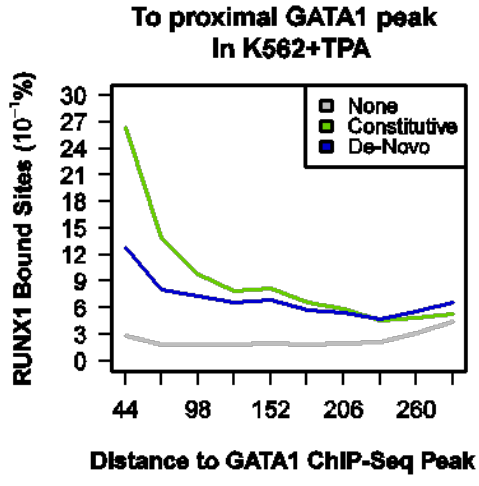

F
To proximal GATA1 peak In K562+TPA
RUNX1 Bound Sites (10⁻¹%)
None
Constitutive
De-Novo
Distance to GATA1 ChIP-Seq Peak

Figure 4

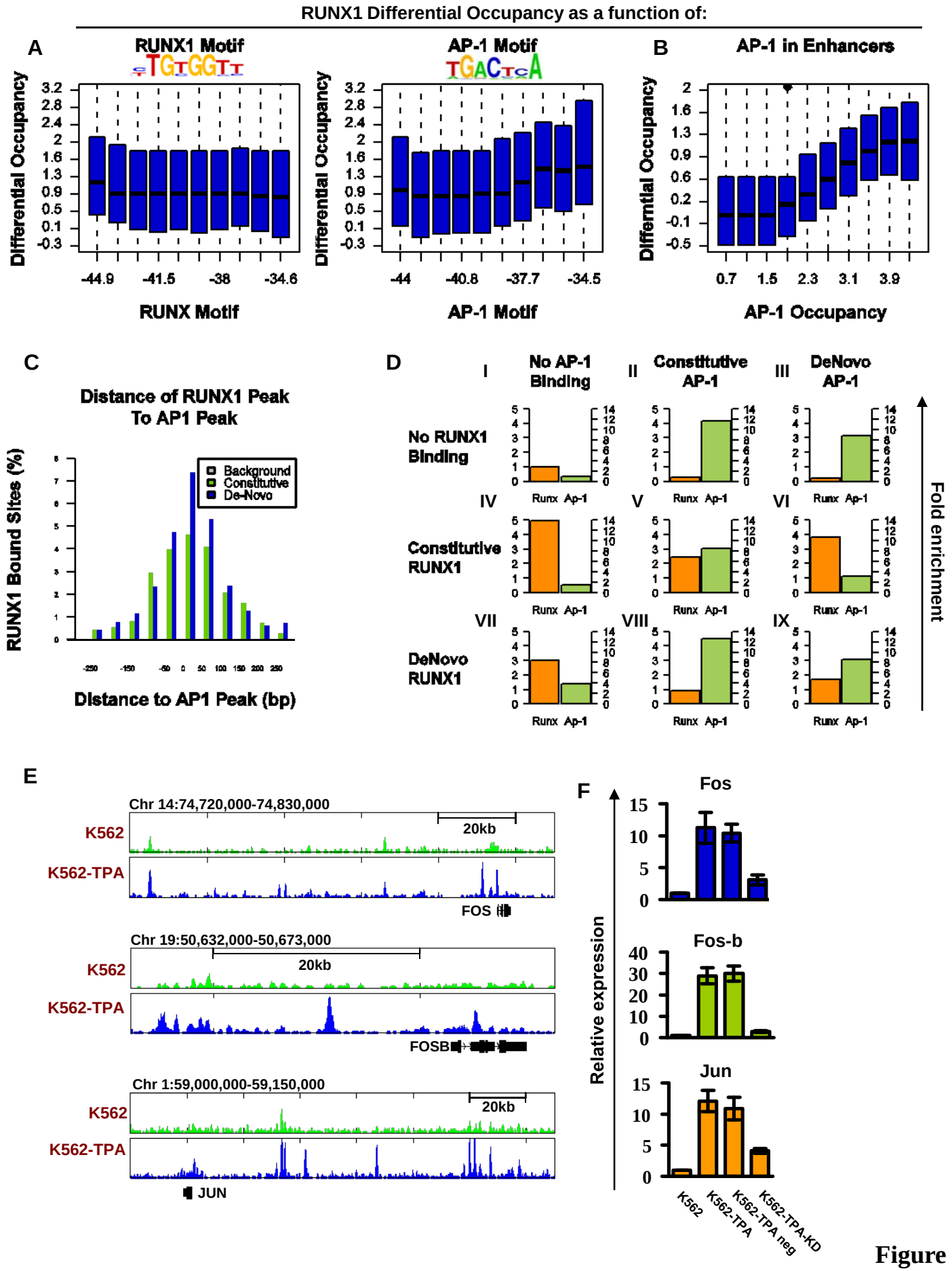

RUNX1 Differential Occupancy as a function of:
A
RUNX1 Motif
AP-1 Motif
Differential Occupancy
RUNX Motif
AP-1 Motif
B
AP-1 in Enhancers
Differntial Occupancy
AP-1 Occupancy
C
Distance of RUNX1 Peak To AP1 Peak
RUNX1 Bound Sites (%)
Background
Constitutive
De-Novo
Distance to AP1 Peak (bp)
D
I No AP-1 Binding
II Constitutive AP-1
III DeNovo AP-1
No RUNX1 Binding
IV
V
VI
Constitutive RUNX1
VII
VIII
IX
DeNovo RUNX1
Runx Ap-1
Fold enrichment
E
Chr 14:74,720,000-74,830,000
K562
K562-TPA
20kb
FOS
Chr 19:50,632,000-50,673,000
FOSB
Chr 1:59,000,000-59,150,000
JUN
F
Fos
Fos-b
Jun
Relative expression
K562
K562-TPA
K562-TPA neg
K562-TPA-KD

Figure 5

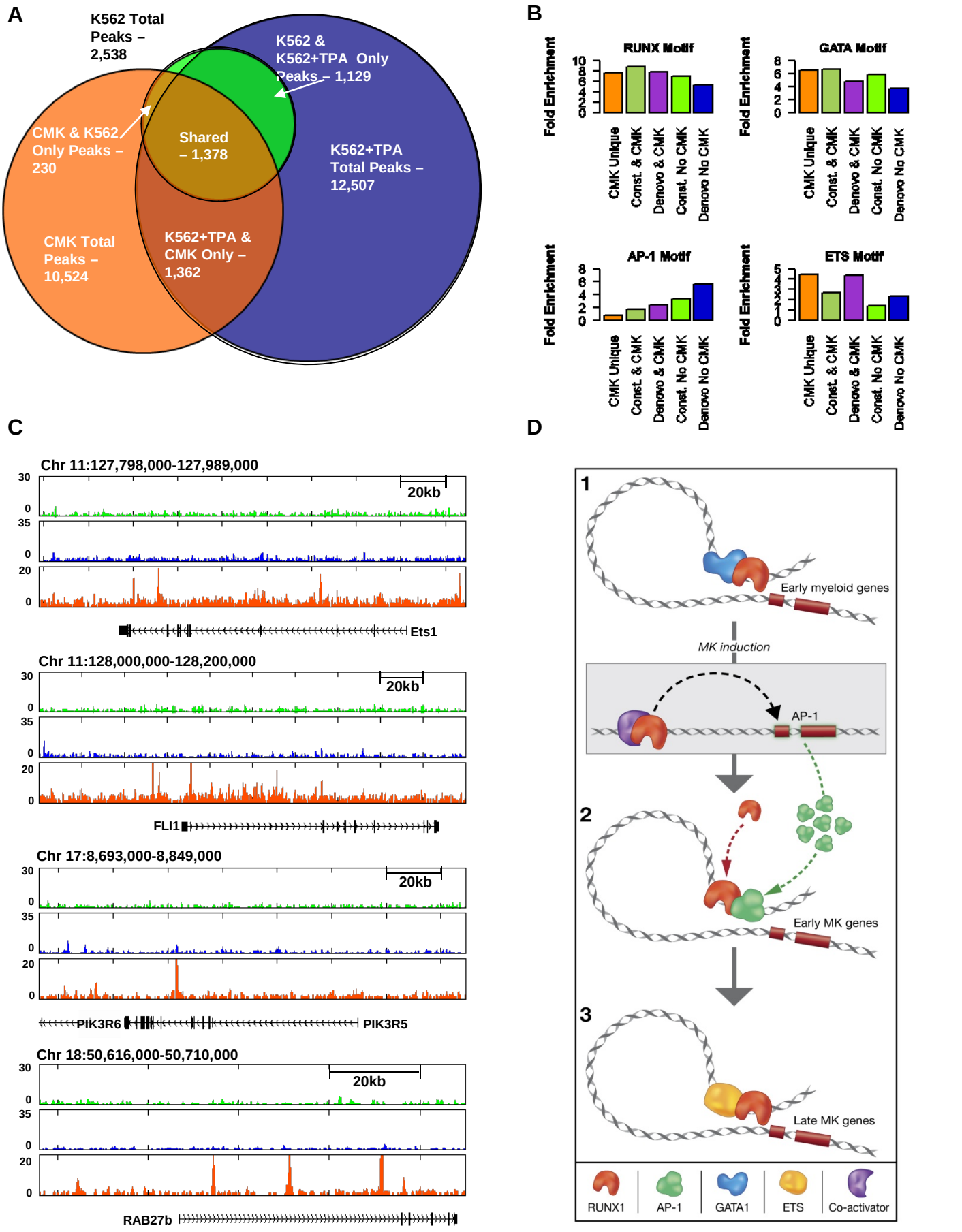

A
K562 Total Peaks – 2,538
K562 & K562+TPA Only Peaks – 1,129
CMK & K562 Only Peaks – 230
Shared – 1,378
K562+TPA Total Peaks – 12,507
CMK Total Peaks – 10,524
K562+TPA & CMK Only – 1,362
B
RUNX Motif
GATA Motif
AP-1 Motif
ETS Motif
Fold Enrichment
CMK Unique
Const. & CMK
Denovo & CMK
Const. No CMK
Denovo No CMK
C
Chr 11:127,798,000-127,989,000
20kb
Ets1
Chr 11:128,000,000-128,200,000
FLI1
Chr 17:8,693,000-8,849,000
PIK3R6
PIK3R5
Chr 18:50,616,000-50,710,000
RAB27b
D
Early myeloid genes
MK induction
AP-1
Early MK genes
Late MK genes
RUNX1
AP-1
GATA1
ETS
Co-activator

Figure 6

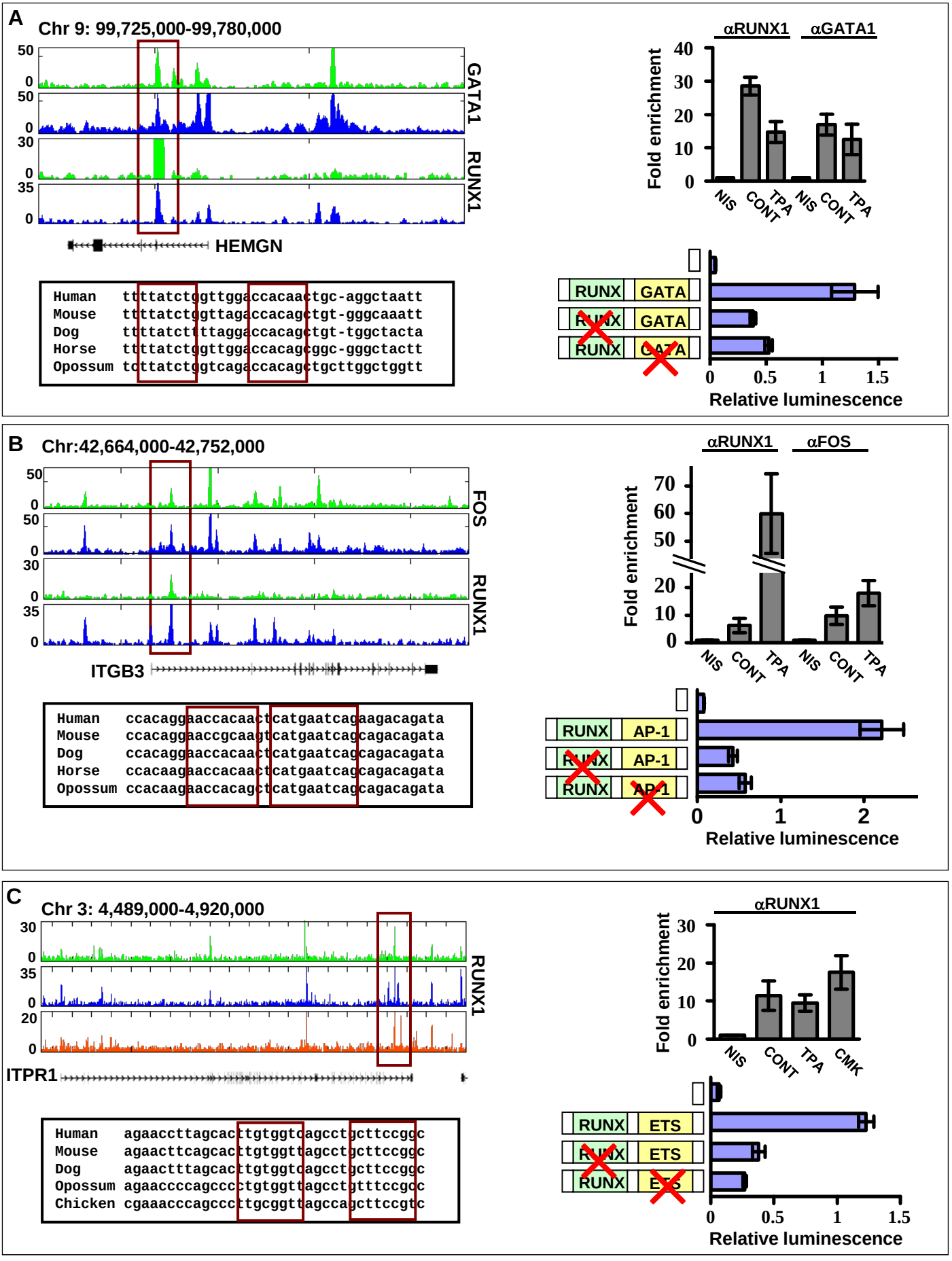

A
Chr 9: 99,725,000-99,780,000
GATA1
RUNX1
HEMGN
Human ttttatctggttggaccacaactgc-aggctaatt
Mouse ttttatctggttagaccacagctgt-gggcaaatt
Dog ttttatctttaggaccacagctgt-tggctacta
Horse ttttatctggttggaccacagcggc-gggctactt
Opossum tcttatctggtcagaccacagctgcttggctggtt
αRUNX1
αGATA1
Fold enrichment
NIS
CONT
TPA
RUNX
GATA
Relative luminescence
B
Chr:42,664,000-42,752,000
FOS
RUNX1
ITGB3
Human ccacaggaaccacaactcatgaatcagaagacagata
Mouse ccacaggaaccgcaagtcatgaatcagcagacagata
Dog ccacaggaaccacaactcatgaatcagcagacagata
Horse ccacaagaaccacaactcatgaatcagcagacagata
Opossum ccacaagaaccacagctcatgaatcagcagacagata
αRUNX1
αFOS
Fold enrichment
AP-1
Relative luminescence
C
Chr 3: 4,489,000-4,920,000
RUNX1
ITPR1
Human agaaccttagcacttgtggtcagcctgcttccggc
Mouse agaacttcagcacttgtggttagcctgcttccggc
Dog agaactttagcacttgtggtcagcctgcttccggc
Opossum agaaccccagccccttgtggttagcctgtttccgcc
Chicken cgaaacccagcccttgcggttagccagcttccgtc
αRUNX1
Fold enrichment
NIS
CONT
TPA
CMK
RUNX
ETS
Relative luminescence

Figure 7