A comprehensive motifs-based interactome of the C/EBPα transcription factor

Ramberger, Evelyn[1], Sapozhnikova, Valeria[1], Kowenz-Leutz, Elisabeth[1], Zimmermann, Karin[1], Nicot, Nathalie[2], Nazarov, Petr V.[2], Perez-Hernandez, Daniel[1,2], Reimer, Ulf[3], Mertins, Philipp[1], Dittmar, Gunnar[1,2], Leutz, Achim[1,4,*]

[1] Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Robert-Rössle-Strasse 10, 13125 Berlin, Germany

[2] Quantitative Biology Unit, Luxembourg Institute of Health, Luxembourg

[3]JPT Peptide Technologies GmbH, Volmerstrasse 5, 12489 Berlin, Germany

[4] Institute of Biology, Humboldt University of Berlin, 10115 Berlin, Germany

*Corresponding author: aleutz@mdc-berlin.de

Keywords: intrinsic disorder, CEBPA interactome, BioID, peptide array, mass-spectrometry

Abstract

The pioneering transcription factor C/EBPα coordinates cell fate and cell differentiation. C/EBPα represents an intrinsically disordered protein with multiple short linear motifs and extensive post-translational side chain modifications (PTM), reflecting its modularity and functional plasticity. Here, we combined arrayed peptide matrix screening (PRISMA) with biotin ligase proximity labeling proteomics (BioID) to generate a linear, isoform specific and PTM-dependent protein interaction map of C/EBPα in myeloid cells**.** The C/EBPα interactome comprises promiscuous and PTM-regulated interactions with protein machineries involved in gene expression, epigenetics, genome organization, DNA replication, RNA processing, and nuclear transport as the basis of functional C/EBPα plasticity. Protein interaction hotspots were identified that coincide with homologous conserved regions of the C/EBP family and revealed interaction motifs that score as molecular recognition features (MoRF). PTMs alter the interaction spectrum of multi-valent C/EBP-motifs to configure a multimodal transcription factor hub that allows interaction with multiple co-regulatory components, including BAF/SWI-SNF or Mediator complexes. Combining PRISMA and BioID acts as a powerful strategy to systematically explore the interactomes of intrinsically disordered proteins and their PTM-regulated, multimodal capacity.

Key points

- Integration of proximity labeling and arrayed peptide screen proteomics refines the interactome of C/EBPα isoforms
- Hotspots of protein interactions in C/EBPα mostly occur in conserved short linear motifs
- Interactions of the BAF/SWI-SNF complex with C/EBPα are modulated by arginine methylation and isoform status
- The integrated experimental strategy suits systematic interactome studies of intrinsically disordered proteins

Introduction

CCAAT enhancer binding protein alpha (C/EBPα) is a lineage instructing pioneering transcription factor involved in cell fate decisions in various cell types, including myeloid cells, adipocytes, hepatocytes, and skin cells. In myelopoiesis, C/EBPα cooperates with other transcription factors and chromatin modifying enzymes to regulate hematopoietic stem cell biology, lineage choice and expression of myeloid genes (Avellino and Delwel, 2017; Zaret and Carroll, 2011). Knockout experiments in mice show that C/EBPα regulates hematopoietic stem cell functions and that its removal blocks progenitor differentiation at the myeloid commitment stage (Bereshchenko et al., 2009; Kirstetter et al., 2008a; Zhang et al., 2004). The intronless *CEBPA* gene is mutated in approximately 10-15% of human acute myeloid leukemia (AML) cases. Mutations located within the N-terminal part of the gene obstruct expression of the full-length isoform p42 C/EBPα and favor expression of the N-terminally truncated isoform p30 C/EBPα (Fasan et al., 2014; Lin et al., 2005; Pabst et al., 2001). Experimental hematology and targeted mouse genetics have demonstrated that p30 *C/EBPα* represents a highly penetrant AML gene (Bereshchenko et al., 2009; Kirstetter et al., 2008b).

Unraveling the protein interaction network of C/EBPα in hematopoietic cells is a prerequisite for the understanding of diverse gene regulatory and epigenetic functions of C/EBPα in normal hematopoiesis and AML. Previous research has shown that the N-terminus of vertebrate C/EBPα harbors short, conserved regions (CRs) that function in a modular and combinatorial fashion to regulate gene expression (Leutz et al., 2011; Tsukada et al., 2011). The N-terminally truncated, leukemogenic p30 C/EBPα isoform contains only one of the transactivation modules (CR1L, TEIII). P30 C/EBPα retains residual gene regulatory and epigenetic capacity to direct myeloid lineage commitment but lacks the cell maturation inducing and

proliferation restricting functions of the long C/EBPα isoform p42 (Bereshchenko et al., 2009; Pedersen et al., 2001). Only the C-terminal part of C/EBPα, comprising approximately one quarter of the protein, may adopt a structured basic leucine zipper (bZIP) domain upon binding to DNA target sites (Seldeen et al., 2008). Otherwise, C/EBPα and other family members represent mosaics of intrinsically disordered regions (IDRs) with alternating small linear motifs (SLiM) and molecular recognition features (MoRF), typical hallmarks of gene regulatory proteins that integrate signal transmission, gene regulation, and epigenetic regulation through promiscuous interactomes (Dunker et al., 2002; Dyson and Wright, 2005; Tompa, 2012; Uversky et al., 2005; van der Lee et al., 2014; Ward et al., 2004; Wright and Dyson, 1999). PTMs, downstream of signaling cascades, frequently alter dynamic protein interaction interfaces of IDR/SLiM/MoRF modules and are of high biological relevance but remain difficult to capture with traditional immune affinity-based methods as a result of their transient nature, weak affinity, and/or sub-stoichiometric participation in biomolecular condensates (Perkins et al., 2010; Sabari et al., 2020). Within the C/EBP family, mutation of PTM sites, including phosphorylation (S, T, Y residues), methylation (K, R residues), and acetylation sites (K residues) or deletion of CRs in their N-termini may radically alter their biological functions (Leutz et al., 2011).

Assessment of the known C/EBPα interactome data showed a surprisingly low overlap (less than 5%) between various published immuno-affinity-mass spectrometry (MS) datasets, hampering the deduction of C/EBPα functions and adjunct molecular mechanisms (Cirilli et al., 2017; Giambruno et al., 2013; Grebien et al., 2016). Here, we combined the protein interaction screen using a peptide matrix (PRISMA) method (Dittmar et al., 2019) with biotin proximity labeling identification (BioID) (Roux et al., 2012) to match high confidence interactomes across the primary sequence and PTM sites with the p42/p30 C/EBPα isoforms in living myeloblastic

cells. This experimental strategy may be adapted to explore other multi-valent intrinsically disordered transcriptional regulators, many of which are involved in development and disease.

Results

Mapping the C/EBPα interactome with PRISMA

We have previously established the PRISMA method to explore the linear interactomes of intrinsically disordered proteins and applied it to C/EBPβ and its PTMs (Dittmar et al., 2019). Here, we slightly modified the PRISMA protocol to explore the linear and PTM-dependent C/EBPα interactome. Briefly, membrane spot-synthesized, tiling peptides that covered the primary sequence of C/EBPα were incubated with nuclear extracts derived from the human promyelocytic leukemia cell line, NB4. Bound proteins of each spot were subsequently analyzed using quantitative mass spectrometry **(Figure 1A)**. The C/EBPα peptide array included 114 peptides with and without PTMs that were designed with a sequence overlap of seven amino acids (**Supplemental Table 1**). PTMs within the bZIP domain were omitted from the PRISMA screen since they may modulate DNA binding in addition to protein interactions. The overall technical reproducibility of the PRISMA screen replicates showed a median Pearson correlation coefficient of 0.85 and discrete patterns of correlation between different C/EBPα regions **(Supplemental Figure 1)**. PRISMA detected a total of 2,274 proteins, of which 785 proteins passed the significance threshold (FDR < 0.1) in at least one peptide spot **(Supplemental Table 2)**. The majority of significant protein interactions were observed in the regions CR2, CR3, CR4 and CR1L, which together correspond to the trans-regulatory regions of C/EBPα (Leutz et al., 2011). Extracted binding profiles, as shown in **Figure 2C,**

highlight distinct interaction profiles of selected protein complexes such as Mediator, or BAF/SWI-SNF components, previously shown to interact with C/EBPα.

The mediator of transcription complex (MED) is an essential transcriptional coactivator in eukaryotes that interacts with RNA Pol II and many transcription factors and co-factors. PRISMA revealed 17 MED proteins with similar binding patterns with dominant peaks in C/EBPα peptides corresponding to CR2,3,4 and CR1L. This finding suggests that the MED-complex interacts with multiple CRs in the N-terminus of p42 C/EBPα and with CR1L/TEIII, which corresponds to the N-terminus of p30 C/EBPα. The acetyltransferases, CBP/p300 (KAT3A/KAT3B), most strongly bound to peptides spanning regions CR3,4 with residual binding in CR2 and CR1L. Interaction with the chromatin remodeling BAF/SWI-SNF complex is essential for the anti-proliferative and differentiation functions of C/EBPα and has been previously found to require CR1L/TEIII in addition to the N-terminus of p42 C/EBPα (Muller et al., 2004; Pedersen et al., 2001). In PRISMA, BAF/SWI-SNF subunits bound to peptides corresponding to CR1L and CR3,4.

C/EBP proteins are extensively decorated with PTMs that may alter their interactome to direct their function (Kowenz-Leutz et al., 2010; Leutz et al., 2011). Of particular interest are a number of C/EBP-family member specific arginine residues, several of which have been shown in C/EBPβ to be targets of mono- and di-methylation, with their mutation profoundly changing C/EBPβ biology (Dittmar et al., 2019; Leutz et al., 2011). We examined C/EBPα for PTMs by mass spectrometry and discovered several known and novel methylation sites, including methylation of R12, R35 and R156 and further identified R142 (CR1L) by a targeted MS-parallel reaction monitoring approach as a novel methylation site **(Supplemental Table 3** and **Supplemental Figure 2)**. The PRISMA data suggested increased binding of the BAF/SWI-SNF subunit SMARCE1 to R142 methylated peptides, as compared to the

unmodified counterpart. Other subunits of BAF/SWI-SNF (SMARCA4, SMARCC2) followed a similar trend but their differential binding to the methylated peptide spanning R142 scored somewhat below the threshold set of statistical significance.

Given the sequence conservation between CEBP proteins, we compared the PRISMA derived interactome of C/EBPα with the previously published C/EBPβ PRISMA interactome (Dittmar et al., 2019). Despite some sequence differences in homologous regions and peptides, several shared interactors were identified. For example, the mediator complex was found to bind to the same homologue CRs in both proteins **(Supplemental Figure 3)**, confirming the functional similarity of these regions within the CEBP family (Jones et al., 2002).

Cross-validation of PRISMA and BioID-C/EBPα interactomes

Next, we compared the interactomes derived from PRISMA and BioID proximity labeling data to generate a myeloid live cell validated linear C/EBPα interactome (**Figure 2 A**). Briefly, NB4 cells were transduced with a Tet-On inducible lentiviral construct encoding a promiscuous biotin ligase (BirA*) fused to the C-terminus of C/EBPα. As controls, we used non-induced NB4 C/EBPα-BirA* cells or cells expressing only the ligase moiety (NB4 BirA*). Proximity labeling identified 397 C/EBPα proximal interactors (FDR < 0.05). Among the most enriched proteins were several transcription factors of the C/EBP and ATF families, representing known heterodimerization partners of C/EBPα (Tsukada et al., 2011) and thus confirming successful BioID labeling (**Figure 2B, Supplemental Table 4**).

In total, 88 proteins overlapped between the PRISMA and BioID derived C/EBPα interactomes of which 21 were previously identified interactors, including members of chromatin remodeling and histone acetylation / deacetylation complexes. In addition,

49 proteins significant in PRISMA and 26 proteins significant in BioID have been previously described to interact with C/EBPα (Chatr-Aryamontri et al., 2017; Grebien et al., 2016; Szklarczyk et al., 2015). Taken together, 137 proteins represent the subset of myeloid C/EBPα interactors with the highest confidence that can be depicted across the linear C/EBPα sequence +/- PTM sites. These interactors show high connectivity according to experimentally validated interactions listed in the STRING database **(Figure 2C)**. Using the spatial information provided by PRISMA, we could also reveal distinct functional roles of individual CRs including chromatin remodeling (CR3,4; CR1L; bZIP), transcriptional regulation (CR2; CR3,4; CR1L; CR6; bZIP) and hematopoietic progenitor cell differentiation (CR3,4) (**Supplemental Figure 4)**. CR3,4 is of particular interest since it contains most of the significantly enriched GO terms, confirming the importance of the core transactivating region of p42 C/EBPα. Although several proteins interacted with CR7, no GO terms were enriched with this region, pointing towards functional heterogeneity.

Arginine methylation-dependent interaction of BAF/SWI-SNF complex subunits

Differential interactions with post-translationally modified peptides in the primary C/EBPα sequence were also detected by PRISMA. Among them was the BAF/SWI-SNF complex that has been previously described to interact with C/EBPα (Pedersen et al., 2001). PRISMA revealed that the interaction with the SMARCE1 component is modulated by the newly discovered arginine methylation site R142 within CR1L/TEIII **(Figure 3A)**. Accordingly, we performed BioID with a R142/149/156 to L142/149/156 mutant (triple R to L mutation; tR>L) of C/EBPα to examine the PTM-dependent BAF/SWI-SNF interaction in more detail. In accordance with the PRISMA results, SMARCE1 and three additional BAF/SWI-SNF subunits (ARID1A, ARID1B, ARID2) were significantly enriched in the p30 C/EBPα-tR>L mutant, as compared to the WT

p30 C/EBPα-BioID **(Figure 3B)**. BioID with p30 C/EBPα-tR>L also verified the methylation-dependent interaction with the E3 ubiquitin ligase TRIM33 and the NuRD complex component GATAD2A, as also detected by PRISMA. In addition, several Myb-Muvb/DREAM complex members (LIN9, LIN37, MYBL2) were identified as p30 C/EBPα-tR>L-specific interactors but were not detected in PRISMA.

C/EBPα isoform specific interactions detected with BioID

The PRISMA data predicted that p30 C/EBPα isoform can still function to recruit major components of the transcriptional and epigenetic machinery, albeit with lower efficiency compared to p42 C/EBPα. To further examine the differences between the two C/EBPα isoforms, we expressed p42 and p30 C/EBPα as BioID fusions in NB4 cells to compare their BioID interactomes. The isoform-specific interactomes **(Figure 4A, Supplemental Table 4)** confirmed the binding profiles observed in PRISMA: p42 specific interactors bound more strongly to PRISMA peptides corresponding to the N-terminal, p42 unique part of C/EBPα, while p30 specific interactors predominantly bound to PRISMA peptides corresponding to C-terminal C/EBPα regions **(Figure 4B)**. In accordance with the PRISMA interactions, most interactors identified by in vivo proximity labelling were found to interact with both C/EBPα isoforms (**Figure 4C**). These data suggest that multi-valent interactions with different affinities may occur with distinct C/EBPα CRs, including CR1L as part of p30 C/EBPα. Overall, p42 C/EBPα-BioID and p30 C/EBPα-BioID pulldowns revealed 71 and 9 interactors that preferentially interacted with p42 and p30 C/EBPα, respectively.

Proteins differentially interacting with the p30 C/EBPα isoform and its PTMs may pose a selective vulnerability for AML cells expressing p30 C/EBPα. We therefore extracted CRISPR Cas9 knockout study derived dependency scores of the nine p30

C/EBPα specific interactors in 18 different AML cell lines using the DepMap Portal **(Supplemental Figure 5)** (Tsherniak et al., 2017). Among the AML cell lines inspected, 9 out of 18 scored as sensitive to TFAP4 knockout (threshold < – 0.5). Two and one cell line tested were sensitive to GATA1 and BCL11A or BLM knockout, respectively. The p30 C/EBPα specific binding of the BAF/SWI-SNF complex member BCL11A further highlights context-specific regulation of BAF/SWI-SNF complex interaction to C/EBPα isoforms.

To analyze connections between the C/EBPα isoform specific interactome with the transcriptome we also performed RNA expression analysis of NB4 cell lines ectopically expressing p42 or p30 C/EBPα. We found that despite the largely overlapping interactome, the two isoforms induced differential gene expression changes (**Figure 4D**) that could be attributed to differential interactions with other transcription factors. Gene expression profiles were analyzed using gene set enrichment analysis (GSEA) employing immunologic and transcription factor target databases from the molecular signature database (Subramanian et al., 2005) (**Figure 4E**). The erythroid transcription factor GATA1 was found to interact differentially with p30 C/EBPα; GSEA analysis detected enrichment of a GATA1 signature specifically in p30 C/EBPα expressing cells. Likewise, PPARG specifically interacts with p42 C/EBPα and p42 expressing cells also displayed significant enrichment of a previously published macrophage PPARG gene signature (Roszer et al., 2011). Data from BioID experiments indicated that EGR1 specifically interacted with p42 C/EBPα. Interestingly, we found that a gene signature based on the presence of EGR1 motifs was enriched in p30 but not p42 C/EBPα expressing cells. RNA expression levels of EGR1 were upregulated by both p42 and p30 expression. Indirect interaction of EGR1 with DNA not at EGR1 motifs, but through other CEBPs, has been described previously (Jakobsen et al., 2013) and may contribute to the differential EGR1 signature observed. Taken together, our data suggest that the specific interactions of

C/EBPα isoforms with lineage defining transcription factors are implicated in co-regulation of target genes in the hematopoietic system.

Discussion

In the present study, we combined peptide array screening and BioID to fine map the interactome of the intrinsically disordered transcription factor C/EBPα. The integration of both methods overcame limitations of classical immuno-affinity based interactomes of intrinsically disordered proteins. Contribution of individual PTMs and protein regions to the interactome remain difficult to deconvolute in classical pull-down MS approaches, while PRISMA reveals both SLiM- and PTM-dependent interactions (Dittmar et al., 2019; Meyer and Selbach, 2020). Moreover, transient interactions mediated by IDRs/SLiMs are easily lost during affinity-purification. In contrast, BioID covalently biotinylates proteins that engage with C/EBPα in the living cell and thus enables the detection of transient and dynamic interactions (Roux et al., 2012). Altogether, PRISMA and BioID revealed 785 and 397 interacting proteins, respectively, and the overlay of C/EBPα BioID and PRISMA data sets with public resources disclosed a linear, high confidence C/EBPα core-interactome of 137 proteins. The C/EBPα core-interactome can be depicted as an interaction landscape across the C/EBPα sequence and PTM sites and its components were found to be highly interconnected. In addition to known C/EBPα interactors that operationally represent positive controls, we also discovered novel associations of C/EBPα with proteins involved in gene expression and epigenetic regulation, chromosome organization, RNA processing, and DNA replication, extending the repertoire of myeloid C/EBPα targets and associated functions. The majority of C/EBPα interactions were nevertheless detected in only one of the two datasets. Some of the unique C/EBPα interactors detected by BioID may require more complex,

simultaneously occurring, multi-site interactions or multiple induced fit processes and were missed in PRISMA. Discrepancies may also relate to restriction of BioID to proximal interactions, the absence of suitable biotinylation sites, or stability during the labelling period. PRISMA, on the other hand, may readily detect distal, secondary interactors, such as in large protein complexes.

PRISMA revealed major hotspots of SLiM-based protein interactions in C/EBPα that strongly correlated with data derived from the related transcription factor C/EBPβ (Dittmar et al., 2019; Leutz et al., 2011). The interactors shared between C/EBPα and C/EBPβ may help to explain the partially redundant functions of both proteins (Chen et al., 2000; Hirai et al., 2006; Jones et al., 2002). Many of the co-regulators identified also displayed multiple interactions with several SLiMs/MoRFs/CRs, predominantly in CR2,3,4 and CR1L. These regions correspond to previously defined trans-regulatory regions found within the C/EBP family (Leutz et al., 2011). The same conserved regions scored highly as molecular recognition features (MoRF; **Figure 1B**) that may undergo induced folding and transient disorder-to-order transition during contact with partner proteins (Oldfield et al., 2005). For example, the region CR3,4 of C/EBPε was previously shown to fold into two short orthogonal amphipathic helical regions on interaction with the TAZ2 domain of CBP (Bhaumik et al., 2014). Likewise, CBP/p300 was found to interact with PRISMA peptides covering CR3,4 of C/EBPα, in concordance with homologous regions in C/EBPβ (Dittmar et al., 2019). Interestingly, CR3,4 in C/EBPα and C/EBPε are separated by only few amino acids, whereas species-specific IDRs separate both CRs in C/EBPβ, suggesting structural spatial flexibility and independent, yet combinatorial functions of CR3,4.

Many regulatory proteins involved in signaling, gene expression, and epigenetics harbor IDRs and PTMs that determine their function and connectivity (Dyson and Wright, 2005, 2016). Although protein interactions directed by SLiMs are in general of

low affinity, they may nevertheless exhibit high specificity. This raises the question of how specificity is achieved. Specificity may be forged by the combined action of several SLiMs and PTMs to regulate selective interactions of a subset of SLiMs at a time. Multivalent, promiscuous interactions of transcription factors with co-regulatory proteins have been observed before and may relate to context dependent contacts during dynamic gene regulatory processes (Brzovic et al., 2011; Clark et al., 2018; Vojnic et al., 2011). IDRs within and between SLiMs permit structural flexibility and dynamic interactions, which can also drive the formation of specific protein condensates with transiently favorable interactions (Hahn, 2018). This concept is in accordance with the promiscuous nature of SLiMs and may reflect dynamic multi-modal regulation with the rapid exchange of alternative interaction partners (Dyson and Wright, 2016; Ivarsson and Jemth, 2019). Modular transactivation domains in conjunction with multi-valency are thought to participate in or to initiate biomolecular condensates that involve dynamic, “fuzzy” interactions with multiple co-regulators in interaction hubs (Boija et al., 2018; Brzovic et al., 2011; Chong et al., 2018; Hahn, 2018; Martin and Holehouse, 2020; Tuttle et al., 2018). Remarkably, the C/EBPα primary sequence shows a very high degree of “fuzziness” with strong predictions as initiator regions of liquid-liquid phase separation in the N-terminus **(Figure 1B, red line),** while the PRISMA interaction hotspots largely coincide within regions that show maxima of MoRF predictions **(Figure 1B, blue line)** (Horvath et al., 2020; Malhis et al., 2016).

Our data show that PTMs in C/EBPα may orchestrate multi-modal functions of C/EBPα by modulating interactions with the components of transcriptional and epigenetic machinery. The comparison of C/EBPα WT and the R142L methylation mimicry mutant by proximity labeling confirmed the differential interaction with BAF/SWI-SNF components, as detected by PRISMA. These data are consistent with previous findings showing that BAF/SWI-SNF interacts with CR1L (Muller et al.,

2004; Pedersen et al., 2001). The newly discovered R142 methylation-dependent differential interaction of BAF/SWI-SNF will help to conceive hypothesis driven approaches to explore arginine methyltransferase-dependent epigenetic downstream events.

Data presented here further demonstrate that most of the C/EBPα interactors bind to both C/EBPα isoforms, whereas a subset of 71 and 9 proteins preferentially interacted with the p42 and p30 C/EBPα isoform, respectively. Many of the shared C/EBPα interactors showed diminished signal strength in p30 C/EBPα BioID, supporting the concept of p30 as a “weak” C/EBPα variant with relatively few exclusive binding partners. This view is supported by the fact that defective myelopoiesis in the absence of C/EBPα was largely rescued by p30 C/EBPα, although p30, in the absence of p42 C/EBPα, eventually elicits AML (Bereshchenko et al., 2009; Kirstetter et al., 2008). Interestingly, several of the p30 C/EBPα-specific binders (TFAP4, GATA1, BCL11A) affect the survival of AML cell lines. The p30 C/EBPα specific interaction with GATA1 may further hint at a role at the branch point of erythroid / myeloid commitment in replicative multi-potential progenitors (Drissen et al., 2019), whereas p42 C/EBPα specific interactors, such as EGR1 or PPARγ, are important regulators of myeloid differentiation (Chinetti et al., 1998; Krishnaraju et al., 2001; Lefterova et al., 2010; Lefterova et al., 2008; Mildner et al., 2017; Nguyen et al., 1993; Roszer et al., 2011; Tontonoz et al., 1994).

In summary, the cross-validated myeloid C/EBPα interaction map presented in this study may serve as a resource for further exploration of the biological importance of individual and combinatorial SLiM and PTM functions of C/EBPα. Beyond the C/EBPα,β interactomes, the integration of PRISMA and BioID approaches may help to explore the linear and PTM dependent interactomes of many other important intrinsically disordered proteins involved in cell signaling and cell fate determination.

Materials and Methods

Cell culture

NB4 cells were acquired from Leibniz Institute DSMZ- German Collection of Microorganisms and Cell Culture, Germany (DSMZ no.: ACC 207). Cells were cultivated in a humidified incubator at 37°C, 5% CO2 in RPMI1640. For metabolic labeling, NB4 cells were grown in SILAC RPMI1640 supplemented with 10% dialyzed FCS, 100 mg/ml penicillin-streptomycin, 25mM HEPES, 28 µg/ml L-arginine and 48.67 µg/ml L-lysine (light) 13C615N2 (heavy lysine) or L-lysine D4 (medium heavy lysine).

Plasmids and Generation of Cell Lines

The rat C/EBPα p30 3L mutant (R140; 147; 154 mutated to leucine) was generated using the QuickChange Site Directed Mutagenesis Kit according to the manufacturer's protocol (Agilent #200519). The BirA Ligase containing plasmid was purchased (Addgene #64395). The BirA Ligase was cloned in frame to the C/EBPα C-terminus using BamHI and XhoI restriction sites. PCR primer: 5'-CGCGGATCCAGCGGTGGAAGTGGTGGCCTGAAGGACAACACCGTG and 3'–TGCTCTAGACTCGAGTTATTTATCGTC. The C/EBPα p42, p30 or p30 R3L-BirA Ligase fragments were cloned into the pENTRY2B vector (Clontech #3064) using BamHI and XhoI restriction sites and subsequently introduced into the pInducer21 GFP lentiviral vector (Clontech #3044) using the Gateway LR Clonase TM II cloning kit (ThermoFisher Scientific #11791-020). Viral supernatants were obtained from Lenti-X 293T cells (Clontech #632180) transfected with either pInducer21 GFP lentiviral vector or pInducer21 GFP constructs containing either p42, p30 WT-BirA Ligase, p30 R3L-BirA Ligase, or pInducer21 GFP-BirA ligase only using Lenti-X

Packing Single Shots (Clontech #631275) according to the manufacturer's protocol. NB4 cells were centrifuged (1 h, 900g) with infectious supernatant collected 72 h after transfection and 8 µg/mL hexadimethrine bromide and left for recovery overnight. GFP positive NB4 cells were sorted twice using an Aria II sorter (Becton Dickinson) three days after infection.

NB4 induction

NB4 cells were briefly treated with differentiation inducing agents prior to harvesting. SILAC labeled NB4 cells were seeded during exponential growth in SILAC media supplemented with 2µM all-trans-retinoic acid (ATRA; heavy), 50nM Tetradecanoylphorbol-acetat (TPA; medium heavy) or solvent only (light). Cells were harvested after 12h (ATRA and solvent only treated cells) or 6h (TPA treated cells) and nuclear extracts were prepared as described.

Nuclear extract preparation

Nuclear extracts from NB4 cells were prepared as described previously (Dignam et al., 1983) with slight modifications. NB4 cells were harvested by centrifugation at 1000g, 4 min at 4°C and washed twice with ice cold PBS. Packed cell volume (pcv) was estimated and cells were resuspended in 5xpcv of ice-cold hypotonic buffer (10mM HEPES pH 7.5, 10mM NaCl, 3mM MgCl2) supplemented with protease inhibitors. Cells were incubated on ice for 5 min, followed by addition of dodecyl-β-D-maltosid (DDM) to a final concentration of 0.05%. The sample was vortexed briefly and immediately centrifuged for 5 min at 600g, 4°C. The cytosolic fraction was removed and the nuclei were washed with 20xpcv hypotonic buffer (5 min, 600g, 4°C). The supernatant was removed and the nuclei were washed with 20xpcv PBS (5 min, 600g, 4°C). The nuclei were extracted with 2/3xpcv of high salt buffer (20mM

HEPES pH 7.5, 400mM NaCl, 1mM EDTA ph 8, 1mM EGTA pH 8, 20% glycerol, 1mM DTT) supplemented with protease inhibitors while shaking on a tubeshaker at 4°C for 20 min at 750 rpm. Nuclear extracts were cleared by centrifugation at 18000g for 20 min at 4°C and the buffer was exchanged to membrane binding buffer (20mM HEPES pH 7.5, 400mM NaCl, 1mM EDTA ph 8, 1mM EGTA pH 8, 25% glycerol, 1mM DTT) by gel filtration with PD MidiTrap G10 columns (GE healthcare) according to instructions of manufacturers.

Protein Interaction Screen on a peptide Matrix

Custom PepSpot cellulose membranes were ordered from JPT and PRISMA was performed as described before (Dittmar et al., 2019) with slight adaptations. All washing and incubation steps were performed at 4°C on a rocking platform set to 700 rpm. Membranes were wetted in membrane binding buffer for 15 min, followed by a blocking step with 1 mg/ml yeast tRNA in membrane binding buffer for 10 min. Membranes were washed 5 x for 5 min with membrane binding buffer and incubated with nuclear extracts on ice for 30 min. The protein extract was removed and the membranes were washed 3 x 5 min with membrane binding buffer. The individual peptide spots were punched out with a 2mm biopsy puncher and placed into single wells of a 96 well plate containing 20µl denaturation buffer (6M urea, 2M thiourea, 10mM HEPES pH 8). The samples were digested in solution on a PAL robot system. In brief, proteins were reduced with 1mM TCEP for 30 min followed by alkylation with 5mM CAA for 20 min. To each sample 0.5 µg of sequencing grade lysyl endopeptidase (LysC) was added. Samples were digested for 2h before being diluted with four volumes of 50mM ammonium-bi-carbonate and continuation of the digestion over night at room temperature. Digested samples were acidified with TFA and desalted with C18 stage tips as described before (Rappsilber et al., 2003).

434 **LC MS/MS**

435 Desalted and dried peptides were resuspended in MS sample buffer (3% ACN/ 0.1%
436 FA) and separated online with a Easy-nLC™ 1200 coupled to a Q-exactive+ or a Q
437 exactive HF-X mass spectrometer equipped with an orbitrap electrospray ion source.
438 Samples were separated on line on a 20 cm reverse-phase column (inner diameter
439 75µm) packed in house with 3 µm C18-Reprosil beads with a linear gradient ramping
440 from 3% to 76% acetonitrile. PRISMA samples were separated with a 1h gradient
441 and MS data was acquired on a Q-exactive+ in data dependent acquisition mode
442 with a top10 method. Full scan MS spectra were acquired at a resolution of 70 000 in
443 the scan range from 300 to 1700 m/z, automated gain control (AGC) target value of
444 $1e^6$ and maximum injection time of 120ms. MS/MS spectra were acquired at a
445 resolution of 17500, AGC target of $1e^5$ and maximum IT of 60 ms. Ions were isolated
446 with a 2 m/z isolation window and normalized collision energy was set to 26.
447 Unassigned charge states and single charged precursors were excluded from
448 fragmentation and dynamic exclusion was set to 20s. BioID pulldowns were
449 separated on with a 2h gradient and MS data was acquired on a Q-exactive HF-X in
450 data dependent acquisition mode with a top20 method. Full scan MS spectra were
451 acquired at a resolution of 60000 in the scan range from 350 to 1700 m/z, automated
452 gain control (AGC) target value was set to $3e^6$ and maximum injection time to 10ms.
453 MS/MS spectra were acquired at a resolution of 30000, AGC target of $1e^5$ and
454 maximum IT of 86 ms. Ions were isolated with a 1.6 m/z isolation window and
455 normalized collision energy was set to 26. Unassigned charge states and ions with a
456 chare state of one, seven or higher were excluded from fragmentation and dynamic
457 exclusion was set to 30s. The mass spectrometry proteomics data and search results
458 have been deposited to the ProteomeXchange Consortium via the PRIDE partner
459 repository (Perez-Riverol et al., 2019) with the dataset identifier PXD022903.

460

461 **BioID experiments**

462 NB4 stable cell lines were grown in RPMI supplemented with tetracycline free FCS.
463 Cells were seeded in exponential growth phase in media supplemented with 1mM
464 biotin and 1µg/ml doxycycline. Cells were harvested after 24h by centrifugation and
465 washed twice with ice cold PBS. Cell pellets were resuspended in modified RIPA
466 buffer (lysis buffer: 50 mM tris-HCl (pH 7.2), 150 mM, NaCl, 1% NP-40, 1 mM EDTA,
467 1 mM EGTA, 0.1% SDS, 1% sodium deoxycholate, freshly added protease inhibitors)
468 and incubated on ice for 20 min. Samples were sonicated with a probe sonicator for 5
469 pulses and centrifuged for 20 min at 4°C, 20 000g. For each pulldown (1x 15 cm
470 dish), 80µl neutravidin-agarose bead slurry (Thermo Fisher Scientific) was used.
471 Beads were washed twice in lysis buffer and added to the protein extracts. The
472 samples incubated rotating at 4°C for 2.5 h. Beads were washed 3x with lysis buffer,
473 1x with 1M KCl, 1x 2M Urea in 50mM Tris pH8 and 3x with 50mM Tris pH8. Washing
474 buffers were kept on ice and each washing step was performed with 1ml, inverting
475 the tube 5 times and then centrifuging for 1 min at 2000g to pellet the beads. Washed
476 beads were resuspended in 80µl urea/trypsin buffer (2M urea, 50mM Tris pH7.5,
477 1mM DTT, 5µg/ml trypsin) and incubated 1h at RT on a tubeshaker at 1000rpm. The
478 supernatant was transferred into a fresh tube, the beads washed twice with 60µl 2M
479 urea/50mM Tris pH7.5 and the supernatant combined with the supernatant from the
480 previous step. Residual beads were removed by an additional centrifugation (1 min,
481 5000g) and sample transfer step. Eluted proteins were reduced with 4mM DTT for 30
482 min on a tubeshaker (RT, 1000 rpm). Proteins were alkylated with 10mM IAA in the
483 dark for 45 min (RT, 1000 rpm) and digested over night with 0.5µg trypsin at RT on a
484 thermoshaker at 700 rpm. For an AspN digest, 0.5 µg of trypsin and 0.5 µg AspN
485 were added to the sample. Following overnight digestion, samples were acidified by
486 adding TFA and desalted with stage tips.

Targeted MS analysis of R142 methylation

Synthetic heavy peptide standards with the sequence DGRLEPLEYER and DGR[me]LEPLEYER were custom synthesized by JPT (spiketides L, labeled at the C-terminus with heavy arginine (Arg10)). CEBPA BioID pulldowns were digested with AspN and trypsin and desalted as described. Samples were resuspended in MS sample buffer containing 100 fmol/µl of synthetic heavy peptides. Parallel reaction monitoring (PRM) measurements were acquired on a Q exactive HF-X mass spectrometer coupled to an Easy-nLC^{TM} 1200 HPLC system. Peptides were separated on a 60 min gradient ramping from 2% to 76% acetonitrile. Data acquisition mode cycled between a Top1 MS/data dependent MS2 and data independent measurement of an inclusion list that included the m/z of the synthetic heavy peptides as well as their light counterparts coming from the sample. Resolution of MS2 for data independent acquisition was 60 000 with an AGC target of 1e6, 200 ms maximum IT, 0.7 m/z isolation window and NCE of 27. PRM data was analyzed with Skyline (Pino et al., 2020). Identity of peptides was verified by comparison with the elution profile and fragmentation spectrum of heavy peptide standards.

Mass spec raw data processing with MaxQuant

Mass spectrometry raw files were processed with MaxQuant (version 1.5.2.8) (Cox and Mann, 2008) searching against a human protein database containing isoforms downloaded from uniprot (June 2017) and a database including common contaminants. Fixed modifications were set to carbamidomethylation of cysteines and variable modifications were set to methionine oxidation and N-terminal acetylation. For BioID experiments, lysine biotinylation was added as an additional

variable modification. Depending on digestion mode (trypsin or LysC only), enzyme specificity was selected with a maximum of 2 missed cleavages per peptide. The initial allowed mass deviation of the precursor ion was up to 6 ppm and 20 ppm for fragments. False-discovery rate was set to 1% on protein and peptide level. For SILAC measurements, the requantify option was enabled and minimum ratio count was set to 2. For LFQ analysis, the match between run and fast LFQ option was used with default settings.

Data analysis of mass spec data

Statistical analysis of the dataset was performed using R-statistical software package (version 3.4.1). The protein groups output file from MaxQuant was filtered for contaminants, reverse hits and proteins only identified by site. Minimum peptide count for SILAC and LFQ data was at least 2 peptides per protein group. For LFQ datasets, proteins were filtered by detection in at least 3 replicates of the same sample and missing values were imputed from a distribution at the detection limit of the mass spectrometer. For this purpose a normal distribution was created for each run, with a mean 1.8 standard deviations away from the observed mean and a standard deviation of 0.3 x the observed standard deviation. LFQ data was analysed with a two sample moderated t-test (Limma package) and p-values were corrected for multiple testing with Benjamini-Hochberg procedure. The significance cutoff of CEBPA interactors was enrichment against both controls (BirA* only and cells not treated with doxycycline) with an FDR < 0.05. For isoform and tR>L mutant specific interactors, threshold was enrichment against both controls with an FDR < 0.05 and FDR < 0.1 in the pairwise comparison of isoforms and mutants. Interactors overlapping between PRISMA and BioID or BioID and literature (STRING and BioGrid database, Grebien et al., Giambruno et al.) were visualized with the STRING app built into cytoscape (v3.71) (Shannon et al., 2003). Known interaction

(experimentally validated or deposited in databases with a score > 0.5) were visualized as edges. Interactors not connected by any edges were removed from figure 2C (12 proteins). GO term analysis of interactors was performed with DAVID functional annotation tool (version 6.8) (Huang et al., 2009). PRISMA data was analysed with a two sample moderated t-test (Limma package), creating a specific control group for each peptide that contained all other peptides excluding peptides with a sequence overlap ≥ 50%. P-values were corrected for multiple testing with Benjamini-Hochberg procedure, significance cut off was < 0.1 FDR. To obtain binding profiles of interactors, LFQ intensities of significant proteins were normalized between 0 and 1 across all PRISMA peptides.

Microarray

Cells were harvested at exponential growth phase and seeded at 0.5 x 10^6 cells/ml in RPMI supplemented with or without 1µg/µl doxycycline as independent biological triplicates. After 24h cells were harvested via centrifugation (1200 g, 5 min, 4°C) and washed once with 1x ice cold PBS. Total RNA was extracted with the RNeasy Mini Kit (Qiagen) following manufacturer's instructions. DNA was removed with the DNase Max Kit (Qiagen) following manufacturer's instructions. RNA integrity was assessed using a Fragment Analyzer system and a standard sensitivity RNA kit. All RQN scores were above 9.1. RNA expression analysis was performed with Affymetrix Human ClariomS® microarray using the WT Plus Reagent kit (ThermoFisher Scientific). As starting material, 100ng of total RNA was used and RNA was prepared for hybridization with the GeneChip® Whole Transcript (WT) PLUS Reagent Kit (ThermoFisher Scientific) following manufactures instructions. CEL files were processed using the standard Transcriptome Analysis Console (TAC 4.0) software. Expression values were automatically normalized and summarized using SST-RMA method. Only mRNAs with log2 expression above 6 in at least one sample were

considered for further analysis. Statistical analysis was performed using LIMMA package of R/Bioconductor and p-values were adjusted using Benjamini-Hochberg's FDR. Microarray data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-9947.

GSEA of microarray data

Gene expression changes induced by CEBPA isoforms were calculated by comparing microarray data from NB4 cells expressing CEBPA (P42 or P30) against BirA* expressing cells. Averaged log2 fold changes were used as input for the single sample gene set enrichment analysis (ssGSEA) analysis tool (Barbie et al., 2009) implemented in R (https://github.com/broadinstitute/ssGSEA2.0) using standard parameters. Data was analyzed employing immunologic and transcription factor target gene sets from the molecular signature database (Subramanian et al., 2005). The running score displayed in Fig.4 was calculated with Kolmogorov-Smirnov running sum statistics.

Author contributions

Conceptualization, A.L., G.D., E.R., D.P.; Methodology, G.D., A.L., E.R., E.K.L., V.S.; Investigation and Validation, E.R., V.S., E.K.L.; Resources, G.D., A.L., P.M., U.R., N.N., P.N.; Data curation, E.R., K.Z., A.L.; Writing original draft, E.R.; Writing, review and editing, E.R., A.L., E.K.L., P.M. G.D.; Visualization, E.R., A.L.; Supervision, project administration, and funding acquisition, A.L. (ORCID: 0000-0001-8259-927X), G.D. (ORCID: 0000-0003-3647-8623), P.M. (ORCID: 0000-0002-2245-528X).

Acknowledgement

We thank Tommaso Mari and Marie-Luise Kirchner for scientific discussions and JPT peptide technologies for providing peptide membranes at cost. E.R. was supported by a fellowship of the Berlin School of Integrated Oncology (BSIO).

Conflict of Interest

The authors declare no conflicts of interest.

References

Avellino, R., and Delwel, R. (2017). Expression and regulation of C/EBPalpha in normal myelopoiesis and in malignant transformation. Blood *129*, 2083-2091.

Bereshchenko, O., Mancini, E., Moore, S., Bilbao, D., Månsson, R., Luc, S., Grover, A., Jacobsen, S.E.W., Bryder, D., and Nerlov, C. (2009). Hematopoietic stem cell expansion precedes the generation of committed myeloid leukemia-initiating cells in C/EBPα mutant AML. Cancer cell *16*, 390-400.

Bhaumik, P., Davis, J., Tropea, J.E., Cherry, S., Johnson, P.F., and Miller, M. (2014). Structural insights into interactions of C/EBP transcriptional activators with the Taz2 domain of p300. Acta Crystallogr D Biol Crystallogr *70*, 1914-1921.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M.*, et al.* (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855 e1816.

Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R.E.*, et al.* (2011). The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. Mol Cell *44*, 942-953.

Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A.*, et al.* (2017). The BioGRID interaction database: 2017 update. Nucleic Acids Res *45*, D369-D379.

Chen, S.S., Chen, J.F., Johnson, P.F., Muppala, V., and Lee, Y.H. (2000). C/EBPbeta, when expressed from the C/ebpalpha gene locus, can functionally replace C/EBPalpha in liver but not in adipose tissue. Mol Cell Biol *20*, 7292-7299.

Chinetti, G., Griglio, S., Antonucci, M., Torra, I.P., Delerive, P., Majd, Z., Fruchart, J.C., Chapman, J., Najib, J., and Staels, B. (1998). Activation of proliferator-activated receptors alpha and gamma induces apoptosis of human monocyte-derived macrophages. J Biol Chem *273*, 25573-25580.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X.*, et al.* (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science *361*, eaar2555.

631 Cirilli, M., Bereshchenko, O., Ermakova, O., and Nerlov, C. (2017). Insights into
632 specificity, redundancy and new cellular functions of C/EBPa and C/EBPb
633 transcription factors through interactome network analysis. Biochim Biophys Acta
634 Gen Subj *1861*, 467-476.

635 Clark, S., Myers, J.B., King, A., Fiala, R., Novacek, J., Pearce, G., Heierhorst, J.,
636 Reichow, S.L., and Barbar, E.J. (2018). Multivalency regulates activity in an
637 intrinsically disordered transcription factor. Elife *7*, 1-28.

638 Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates,
639 individualized p.p.b.-range mass accuracies and proteome-wide protein
640 quantification. Nat Biotechnol *26*, 1367-1372.

641 Dittmar, G., Hernandez, D.P., Kowenz-Leutz, E., Kirchner, M., Kahlert, G.,
642 Wesolowski, R., Baum, K., Knoblich, M., Hofstatter, M., Muller, A.*, et al.* (2019).
643 PRISMA: Protein Interaction Screen on Peptide Matrix Reveals Interaction Footprints
644 and Modifications- Dependent Interactome of Intrinsically Disordered C/EBPbeta.
645 iScience *13*, 351-370.

646 Drissen, R., Thongjuea, S., Theilgaard-Monch, K., and Nerlov, C. (2019).
647 Identification of two distinct pathways of human myelopoiesis. Sci Immunol *4*.

648 Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., and Obradovic, Z.
649 (2002). Intrinsic disorder and protein function. Biochemistry *41*, 6573-6582.

650 Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their
651 functions. Nat Rev Mol Cell Biol *6*, 197-208.

652 Dyson, H.J., and Wright, P.E. (2016). Role of Intrinsic Protein Disorder in the
653 Function and Interactions of the Transcriptional Coactivators CREB-binding Protein
654 (CBP) and p300. J Biol Chem *291*, 6714-6722.

655 Fasan, A., Haferlach, C., Alpermann, T., Jeromin, S., Grossmann, V., Eder, C.,
656 Weissmann, S., Dicker, F., Kohlmann, A., Schindela, S.*, et al.* (2014). The role of
657 different genetic subtypes of CEBPA mutated AML. Leukemia *28*, 794-803.

658 Giambruno, R., Grebien, F., Stukalov, A., Knoll, C., Planyavsky, M., Rudashevskaya,
659 E.L., Colinge, J., Superti-Furga, G., and Bennett, K.L. (2013). Affinity purification
660 strategies for proteomic analysis of transcription factor complexes. J Proteome Res
661 *12*, 4018-4027.

662 Grebien, F., Vedadi, M., Getlik, M., Giambruno, R., Avellino, R., Skucha, A., Vittori,
663 S., Kuznetsova, E., Barsyte-lovejoy, D., Li, F.*, et al.* (2016). Pharmacological
664 targeting of the Wdr5-MLL interaction in C / EBPα N-terminal leukemia. Nature
665 Chemical Biology *11*, 571-578.

666 Hahn, S. (2018). Phase Separation, Protein Disorder, and Enhancer Function. Cell
667 *175*, 1723-1725.

668 Hirai, H., Zhang, P., Dayaram, T., Hetherington, C.J., Mizuno, S., Imanishi, J.,
669 Akashi, K., and Tenen, D.G. (2006). C/EBPbeta is required for 'emergency'
670 granulopoiesis. Nat Immunol *7*, 732-739.

671 Horvath, A., Miskei, M., Ambrus, V., Vendruscolo, M., and Fuxreiter, M. (2020).
672 Sequence-based prediction of protein binding mode landscapes. PLoS Comput Biol
673 *16*, e1007864.

674 Ivarsson, Y., and Jemth, P. (2019). Affinity and specificity of motif-based protein-
675 protein interactions. Curr Opin Struct Biol *54*, 26-33.

676 Jakobsen, J.S., Waage, J., Rapin, N., Bisgaard, H.C., Larsen, F.S., and Porse, B.T.
677 (2013). Temporal mapping of CEBPA and CEBPB binding during liver regeneration

678 reveals dynamic occupancy and specific regulatory codes for homeostatic and cell
679 cycle gene batteries. Genome Res *23*, 592-603.

680 Jones, L.C., Lin, M.L., Chen, S.S., Krug, U., Hofmann, W.K., Lee, S., Lee, Y.H., and
681 Koeffler, H.P. (2002). Expression of C/EBPbeta from the C/ebpalpha gene locus is
682 sufficient for normal hematopoiesis in vivo. Blood *99*, 2032-2036.

683 Kirstetter, P., Schuster, M.B., Bereshchenko, O., Moore, S., Dvinge, H., Kurz, E.,
684 Theilgaard-Mönch, K., Månsson, R., Pedersen, T.Å., and Pabst, T. (2008a).
685 Modeling of C/EBPα mutant acute myeloid leukemia reveals a common expression
686 signature of committed myeloid leukemia-initiating cells. Cancer cell *13*, 299-310.

687 Kirstetter, P., Schuster, M.B., Bereshchenko, O., Moore, S., Dvinge, H., Kurz, E.,
688 Theilgaard-Monch, K., Mansson, R., Pedersen, T.A., Pabst, T.*, et al.* (2008b).
689 Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common
690 expression signature of committed myeloid leukemia-initiating cells. Cancer Cell *13*,
691 299-310.

692 Kowenz-Leutz, E., Pless, O., Dittmar, G., Knoblich, M., and Leutz, A. (2010).
693 Crosstalk between C/EBPbeta phosphorylation, arginine methylation, and
694 SWI/SNF/Mediator implies an indexing transcription factor code. EMBO J *29*, 1105-
695 1115.

696 Krishnaraju, K., Hoffman, B., and Liebermann, D.A. (2001). Early growth response
697 gene 1 stimulates development of hematopoietic progenitor cells along the
698 macrophage lineage at the expense of the granulocyte and erythroid lineages. Blood
699 *97*, 1298-1305.

700 Lefterova, M.I., Steger, D.J., Zhuo, D., Qatanani, M., Mullican, S.E., Tuteja, G.,
701 Manduchi, E., Grant, G.R., and Lazar, M.A. (2010). Cell-specific determinants of
702 peroxisome proliferator-activated receptor gamma function in adipocytes and
703 macrophages. Mol Cell Biol *30*, 2078-2089.

704 Lefterova, M.I., Zhang, Y., Steger, D.J., Schupp, M., Schug, J., Cristancho, A., Feng,
705 D., Zhuo, D., Stoeckert, C.J., Jr., Liu, X.S.*, et al.* (2008). PPARgamma and C/EBP
706 factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale.
707 Genes Dev *22*, 2941-2952.

708 Leutz, A., Pless, O., Lappe, M., Dittmar, G., and Kowenz-Leutz, E. (2011). Crosstalk
709 between phosphorylation and multi-site arginine/lysine methylation in C/EBPs.
710 Transcription *2*, 3-8.

711 Lin, L.I., Chen, C.Y., Lin, D.T., Tsay, W., Tang, J.L., Yeh, Y.C., Shen, H.L., Su, F.H.,
712 Yao, M., Huang, S.Y.*, et al.* (2005). Characterization of CEBPA mutations in acute
713 myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and
714 show a distinct immunophenotype of the leukemic cells. Clin Cancer Res *11*, 1372-
715 1379.

716 Malhis, N., Jacobson, M., and Gsponer, J. (2016). MoRFchibi SYSTEM: software
717 tools for the identification of MoRFs in protein sequences. Nucleic Acids Res *44*,
718 W488-493.

719 Martin, E.W., and Holehouse, A.S. (2020). Intrinsically disordered protein regions and
720 phase separation: sequence determinants of assembly or lack thereof. Emerg Top
721 Life Sci *4*, 307-329.

722 Meyer, K., and Selbach, M. (2020). Peptide-based Interaction Proteomics. Mol Cell
723 Proteomics *19*, 1070-1075.

724 Mildner, A., Schonheit, J., Giladi, A., David, E., Lara-Astiaso, D., Lorenzo-Vivas, E.,
725 Paul, F., Chappell-Maor, L., Priller, J., Leutz, A.*, et al.* (2017). Genomic

726 Characterization of Murine Monocytes Reveals C/EBPbeta Transcription Factor
727 Dependence of Ly6C(-) Cells. Immunity *46*, 849-862 e847.

728 Muller, C., Calkhoven, C.F., Sha, X., and Leutz, A. (2004). The CCAAT enhancer-
729 binding protein alpha (C/EBPalpha) requires a SWI/SNF complex for proliferation
730 arrest. J Biol Chem *279*, 7353-7358.

731 Nguyen, H.Q., Hoffmanliebermann, B., and Liebermann, D.A. (1993). The Zinc
732 Finger Transcription Factor Egr-1 Is Essential for and Restricts Differentiation Along
733 the Macrophage Lineage. Cell *72*, 197-209.

734 Oldfield, C.J., Cheng, Y., Cortese, M.S., Romero, P., Uversky, V.N., and Dunker,
735 A.K. (2005). Coupled folding and binding with alpha-helix-forming molecular
736 recognition elements. Biochemistry *44*, 12454-12470.

737 Pabst, T., Mueller, B.U., Zhang, P., Radomska, H.S., Narravula, S., Schnittger, S.,
738 Behre, G., Hiddemann, W., and Tenen, D.G. (2001). Dominant-negative mutations of
739 CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute
740 myeloid leukemia. Nat Genet *27*, 263-270.

741 Pedersen, T.Å., Kowenz-leutz, E., Leutz, A., and Nerlov, C. (2001). Cooperation
742 between C / EBPalpha TBP / TFIIB and SWI / SNF recruiting domains is required for
743 adipocyte differentiation. Genes and Development *15*, 3208-3216.

744 Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S.,
745 Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M.*, et al.* (2019). The
746 PRIDE database and related tools and resources in 2019: improving support for
747 quantification data. Nucleic Acids Res *47*, D442-D450.

748 Perkins, J.R., Diboun, I., Dessailly, B.H., Lees, J.G., and Orengo, C. (2010).
749 Transient protein-protein interactions: structural, functional, and network properties.
750 Structure *18*, 1233-1243.

751 Pino, L.K., Searle, B.C., Bollinger, J.G., Nunn, B., MacLean, B., and MacCoss, M.J.
752 (2020). The Skyline ecosystem: Informatics for quantitative mass spectrometry
753 proteomics. Mass Spectrom Rev *39*, 229-244.

754 Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for
755 matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample
756 pretreatment in proteomics. Anal Chem *75*, 663-670.

757 Roszer, T., Menendez-Gutierrez, M.P., Lefterova, M.I., Alameda, D., Nunez, V.,
758 Lazar, M.A., Fischer, T., and Ricote, M. (2011). Autoimmune kidney disease and
759 impaired engulfment of apoptotic cells in mice with macrophage peroxisome
760 proliferator-activated receptor gamma or retinoid X receptor alpha deficiency. J
761 Immunol *186*, 621-631.

762 Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase
763 fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell
764 Biol *196*, 801-810.

765 Sabari, B.R., Dall'Agnese, A., and Young, R.A. (2020). Biomolecular Condensates in
766 the Nucleus. Trends Biochem Sci *45*, 961-977.

767 Seldeen, K.L., McDonald, C.B., Deegan, B.J., and Farooq, A. (2008). Coupling of
768 folding and DNA-binding in the bZIP domains of Jun-Fos heterodimeric transcription
769 factor. Arch Biochem Biophys *473*, 48-60.

770 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N.,
771 Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for
772 integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

773 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette,
774 M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set
775 enrichment analysis: a knowledge-based approach for interpreting genome-wide
776 expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

777 Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas,
778 J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., *et al.* (2015). STRING v10:
779 Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids
780 Research *43*, D447-D452.

781 Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. Trends Biochem
782 Sci *37*, 509-516.

783 Tontonoz, P., Hu, E., Graves, R.A., Budavari, A.I., and Spiegelman, B.M. (1994).
784 mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer. Genes Dev *8*,
785 1224-1234.

786 Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley,
787 G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., *et al.* (2017). Defining a
788 Cancer Dependency Map. Cell *170*, 564-576 e516.

789 Tsukada, J., Yoshida, Y., Kominato, Y., and Auron, P.E. (2011). The
790 CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors
791 is a multifaceted highly-regulated system for gene regulation. Cytokine *54*, 6-19.

792 Tuttle, L.M., Pacheco, D., Warfield, L., Luo, J., Ranish, J., Hahn, S., and Klevit, R.E.
793 (2018). Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy
794 Protein-Protein Complex. Cell Rep *22*, 3251-3264.

795 Uversky, V.N., Oldfield, C.J., and Dunker, A.K. (2005). Showing your ID: intrinsic
796 disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit *18*,
797 343-384.

798 van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker,
799 A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., *et al.* (2014). Classification
800 of intrinsically disordered regions and proteins. Chem Rev *114*, 6589-6631.

801 Vojnic, E., Mourao, A., Seizl, M., Simon, B., Wenzeck, L., Lariviere, L., Baumli, S.,
802 Baumgart, K., Meisterernst, M., Sattler, M., *et al.* (2011). Structure and VP16 binding
803 of the Mediator Med25 activator interaction domain. Nat Struct Mol Biol *18*, 404-409.

804 Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004).
805 Prediction and functional analysis of native disorder in proteins from the three
806 kingdoms of life. J Mol Biol *337*, 635-645.

807 Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing
808 the protein structure-function paradigm. J Mol Biol *293*, 321-331.

809 Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors : establishing
810 competence for gene expression Parameters affecting transcription factor access to
811 target sites in chromatin Initiating events in chromatin : pioneer factors bind first.
812 Genes and Development, 2227-2241.

813 Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M.L., Dayaram, T., Owens, B.M.,
814 Shigematsu, H., Levantini, E., Huettner, C.S., Lekstrom-Himes, J.A., *et al.* (2004).
815 Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in
816 the absence of the transcription factor C/EBP alpha. Immunity *21*, 853-863.
817

818

Figure Legends

Figure 1: PRISMA delineates the linear interactome of CEBPA across conserved regions and PTM sites. A: Schematic scaled depiction of CEBPA conserved regions (CRs) and post-translational modification (PTM) sites. Spot synthesized, immobilized peptides with and without PTMs covering the complete amino acid sequence of CEBPA were screened for protein interactions with PRISMA. B: Line plots show liquid-liquid phase separation (LLPS; http://protdyn-fuzpred.org) and molecular recognition feature prediction (MoRF; https://morf.msl.ubc.ca/index.xhtml) across the CEBPA sequence. Heatmap shows the binding profile of all significant proteins (y-axis) across CEBPA PRISMA peptides ordered from N- to C-terminus (x-axis). Bar plot on top corresponds to the summed normalized LFQ intensities of all significant proteins in each peptide spot. C: Extracted CEBPA PRISMA profiles of Mediator and P300 complex (top) and BAF/SWI-SNF complex (bottom) subunits.

Figure 2: Integration of BioID and PRISMA data revealed a spatial CEBPA interactome with high interconnectivity. A: BioID proximity labelling data obtained in living cells was integrated with biochemical data obtained with PRISMA. B: The CEBPA BioID interactome. Log fold changes (CEBPA/control) of proteins are plotted against their p-values. C: Integrated protein interaction network of CEBPA. Color of nodes corresponds to detection in datasets as depicted in the legend. Edges represent validated protein interactions retrieved from the STRING database. Interactors not connected by any edges were removed from the plot (12 proteins).

Figure 3: BAF/SWI-SNF complex subunit SMARCE1 preferentially interacts with R142 methylated CEBPA. A: PRISMA interaction profile of proteins differentially binding to CEBPA peptides centered at R142. Peptide sequence is shown on top and methylation status of R142 is shown at the bottom. Stars denote significance (< 0.1

FDR) in R142 dimethylated peptide B: BioID experiments with wildtype CEBPA and trR>L CEBPA. Proteins passing the significance threshold in wildtype to mutant comparison and significant compared to controls are marked in color. Significant BAF/SWI-SNF subunits are indicated by bold letters.

Figure 4: BioID with CEBPA p42/p30 isoforms disclose isoform-specific interactors in conjunction with PRISMA profiles. A: BioID interaction proteomics and microarray RNA expression profiling were performed with CEBPA isoform (p42 or p30) expressing cells. B: PRISMA profiles of interactors showing preference for p42 (log_2(p42/p30) > 1) or p30 (log_2(p42/p30) < - 1) in BioID. LFQ intensities of proteins were summed across PRISMA peptides corresponding to exclusively p42 or both isoforms. The resulting numbers were divided by the total LFQ intensity of each protein across all PRISMA peptides. PTM-modified peptides were omitted from the calculations. C: Relative enrichment of CEBPA interactors against the control, p42 is plotted against p30. Proteins marked in color passed 0.1 FDR threshold in a direct comparison. D: Overlap of up- and downregulated genes as detected by microarray gene expression profiling (comparison to BirA* cells, FDR < 0.05, absolute fold change(FC) > 1). E: Induced gene expression changes were subjected to ssGSEA analysis. Normalized enrichment score (NES) and FDR of informative gene sets are displayed next to running score line plots. The running score was calculated with Kolmogorov-Smirnov (K-S) running sum statistics.

Supplemental Figures Legends

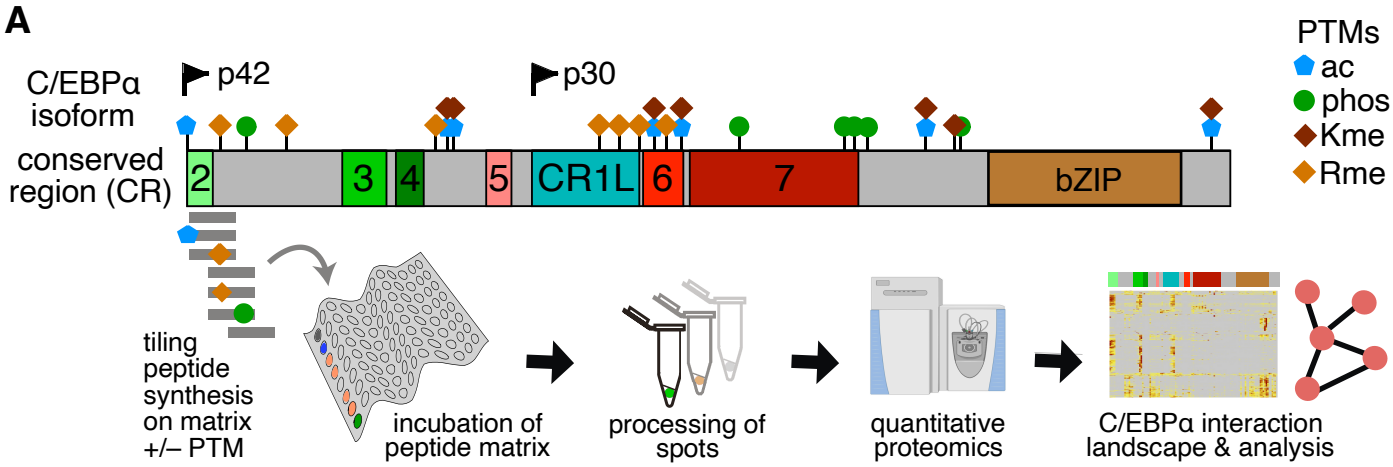
Supplemental Figure 1: Pearson correlation matrix of PRISMA replicates.
Pearson correlation of log2 LFQ values in each PRISMA spot (114 spots). Replicate 1 is plotted against replicate 2. Annotation bars indicates conserved regions of CEBPA.

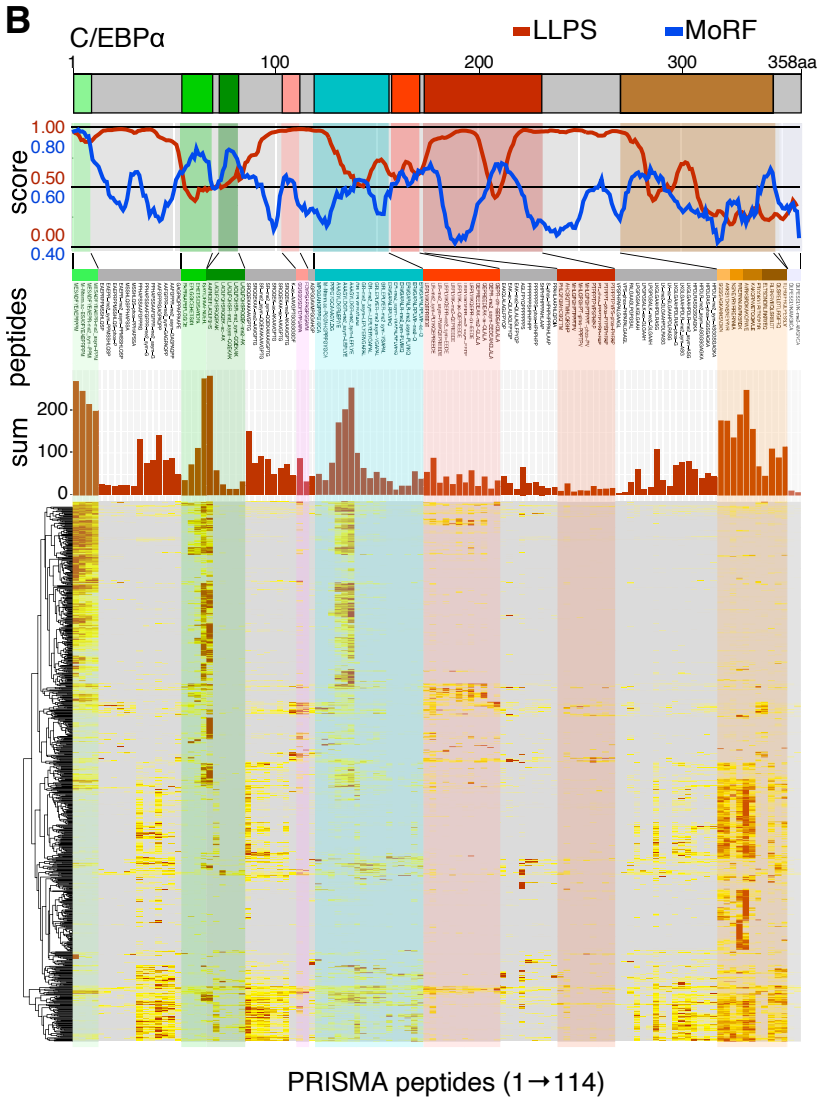
Supplemental Figure 2: Parallel reaction monitoring detects CEBPA methylation at R142.
Parallel reaction monitoring (PRM) of unmodified (top) and methylated (bottom) peptides spanning R142. A heavy peptide standard isotopically labeled at the C-terminal arginine was used to confirm the identity of the peptide. Left: MS2 spectrum of heavy peptide standard; right: elution profile of MS2 fragments of light and heavy peptides. PRM data was analyzed with Skyline.

Supplemental Figure 3: CEBPA and CEBPB share interactors in homologous regions. A: Shared homology of conserved regions in CEBPA and CEBPB. B: Number of high-confidence CEBPA interactors per conserved region in CEBPA as detected by PRISMA and BioID (black bars). Grey bars represent interactors that were also identified as interactors in homologues regions in CEBPB (Dittmar et al., 2019). C: Extracted PRISMA binding profiles of Mediator complex subunits binding to CEBPA and CEBPB (right). Annotation bar on top indicates conserved regions.

Supplemental Figure 4: GO term enrichment of mapped CEBPA interactors. PRISMA CEBPA interactors confirmed by BioID were subjected to GO term analysis using the DAVID tool. Informative significant GO terms (p value < 0.05) are displayed. Grey indicates no significant enrichment. No GO terms were significantly enriched in CR7 binders.

Supplemental Figure 5: **CEBPA P30 specific interactors may represent therapeutic targets in AML**. Dependency scores from CRISPR knockout screens in AML cell lines extracted from the DepMap portal. Scores of P30 specific interactors are displayed. Known tumor-suppressor P53 and oncogene MYB are plotted on top as a reference.





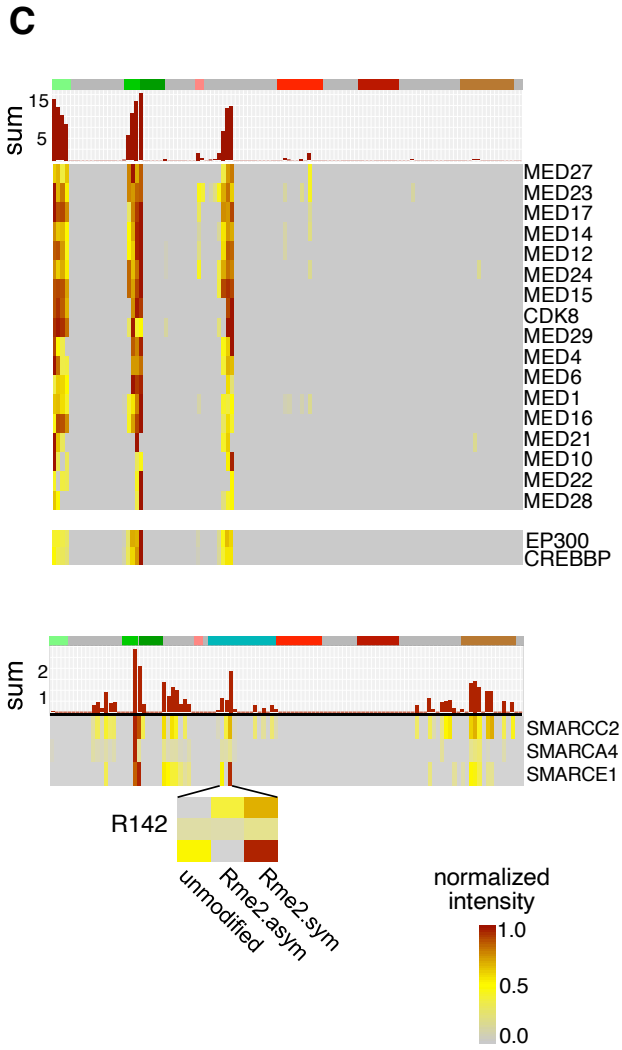


Figure 1, Ramberger et al.

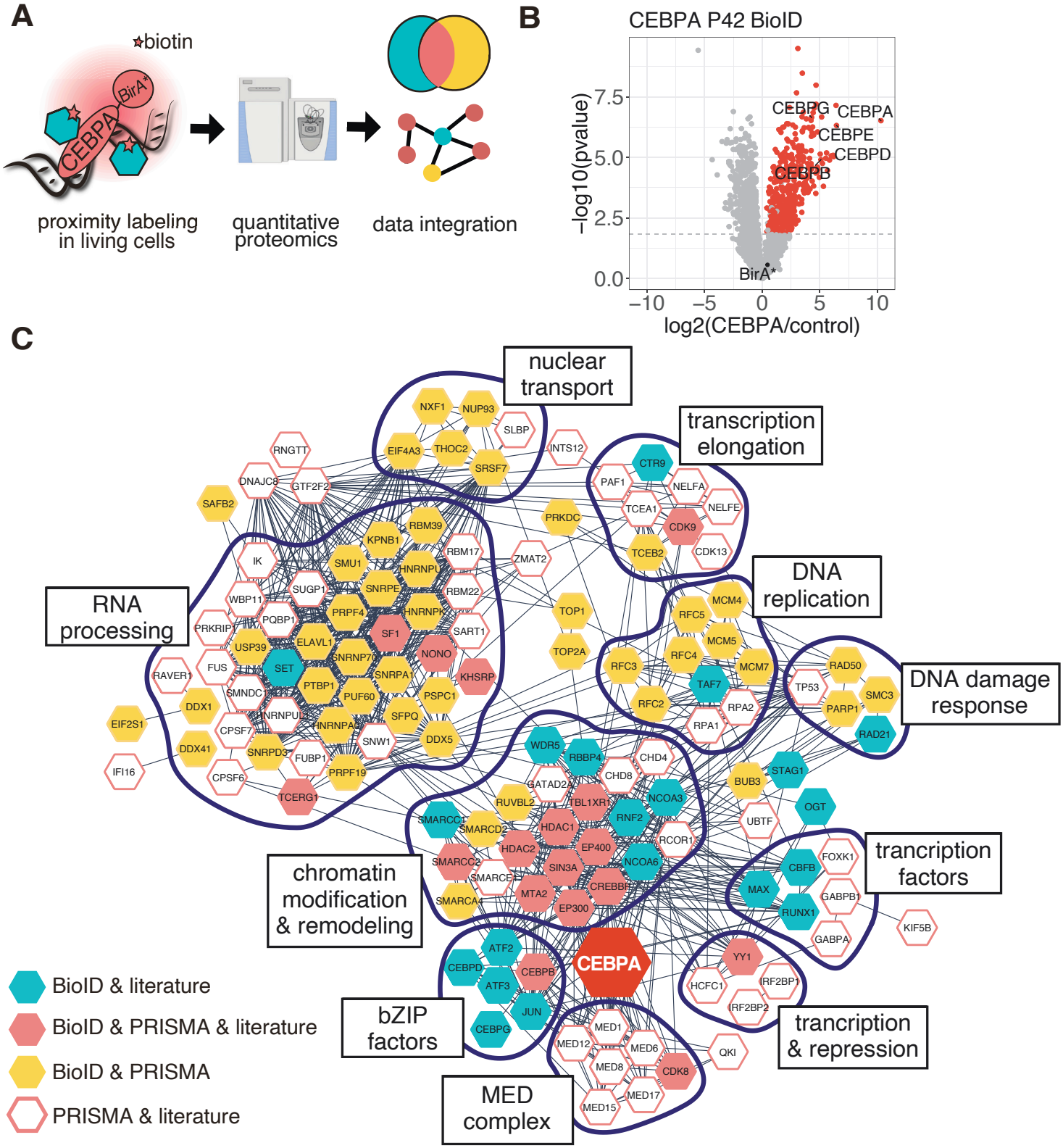


Figure 2, Ramberger et al.

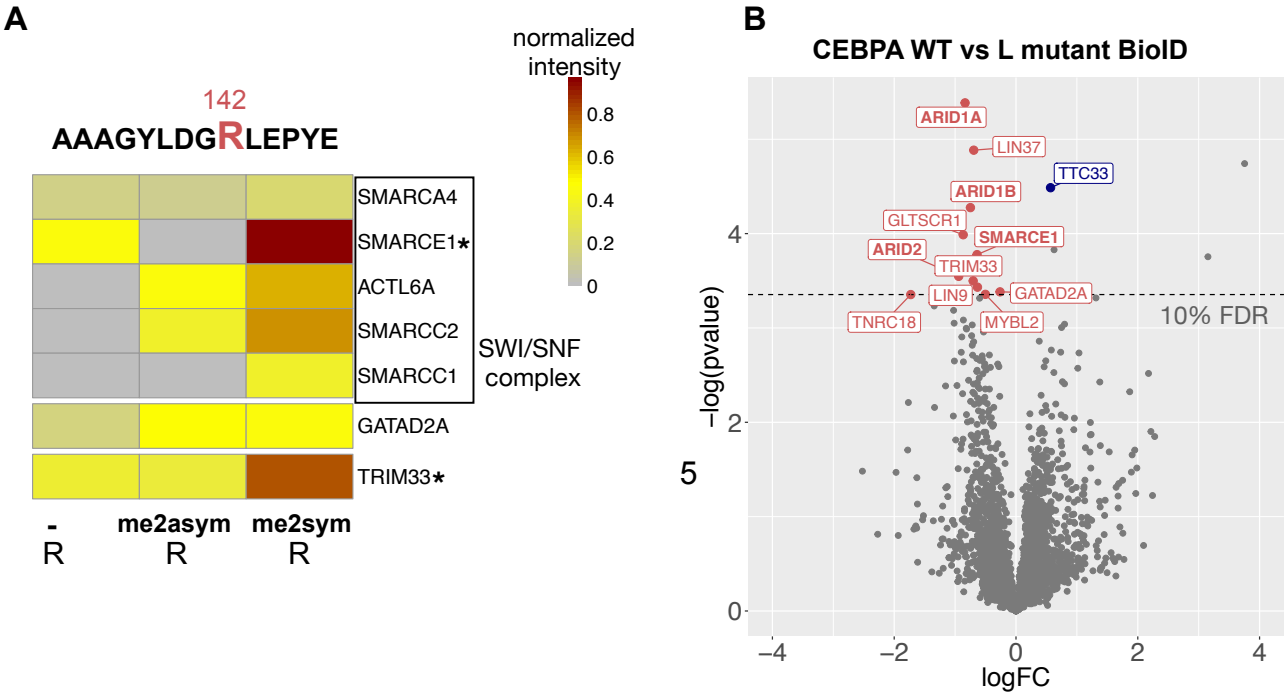


Figure 3, Ramberger et al.

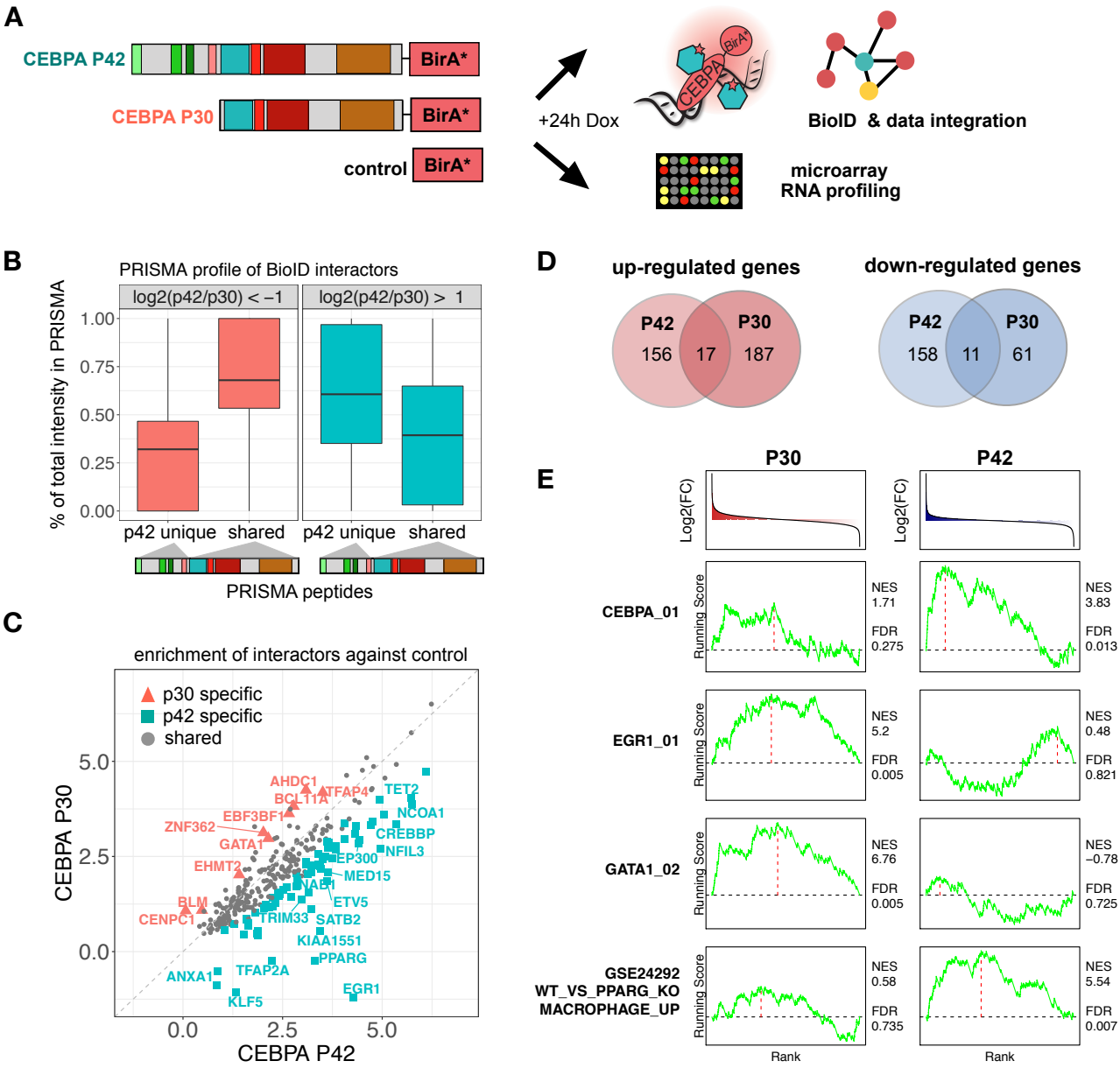
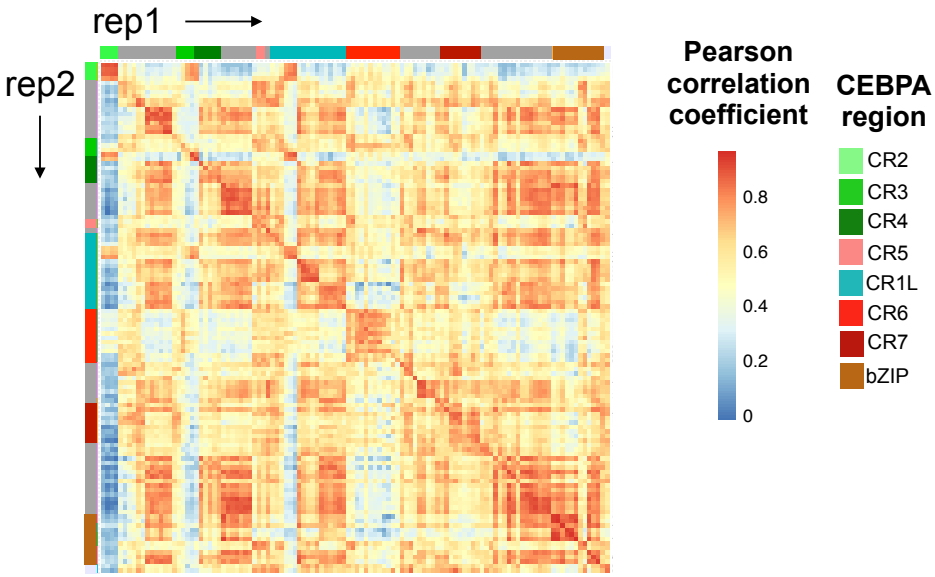
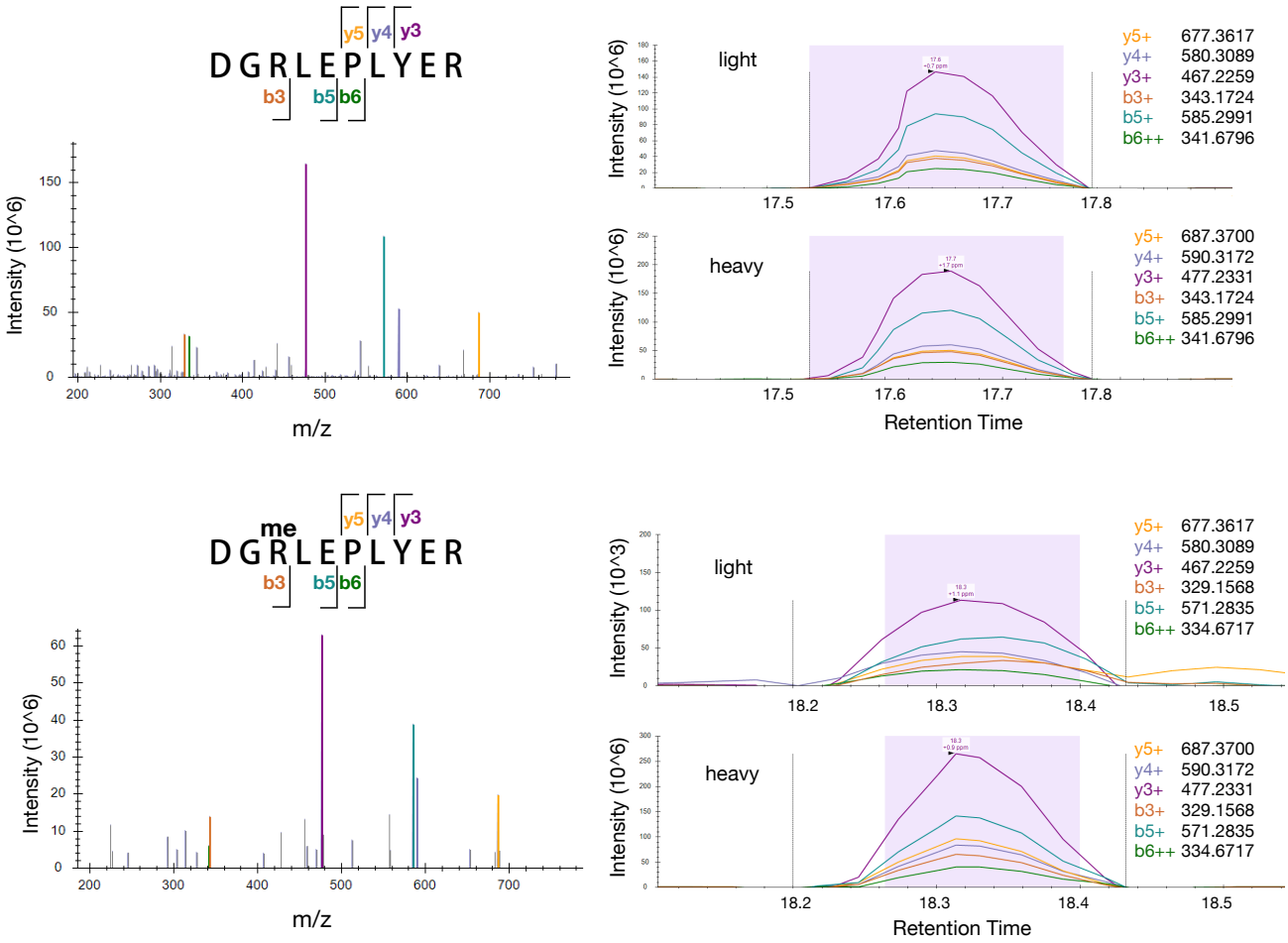


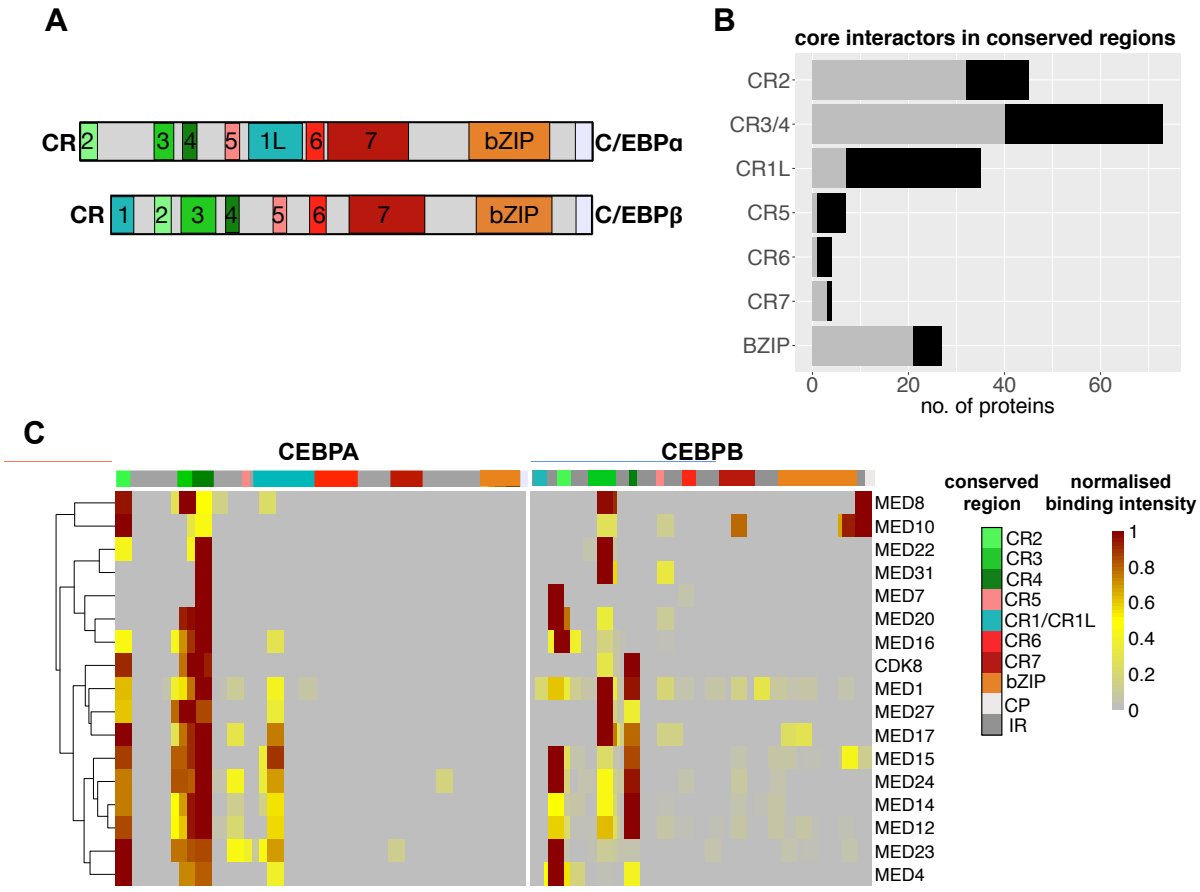
Figure 4, Ramberger et al.



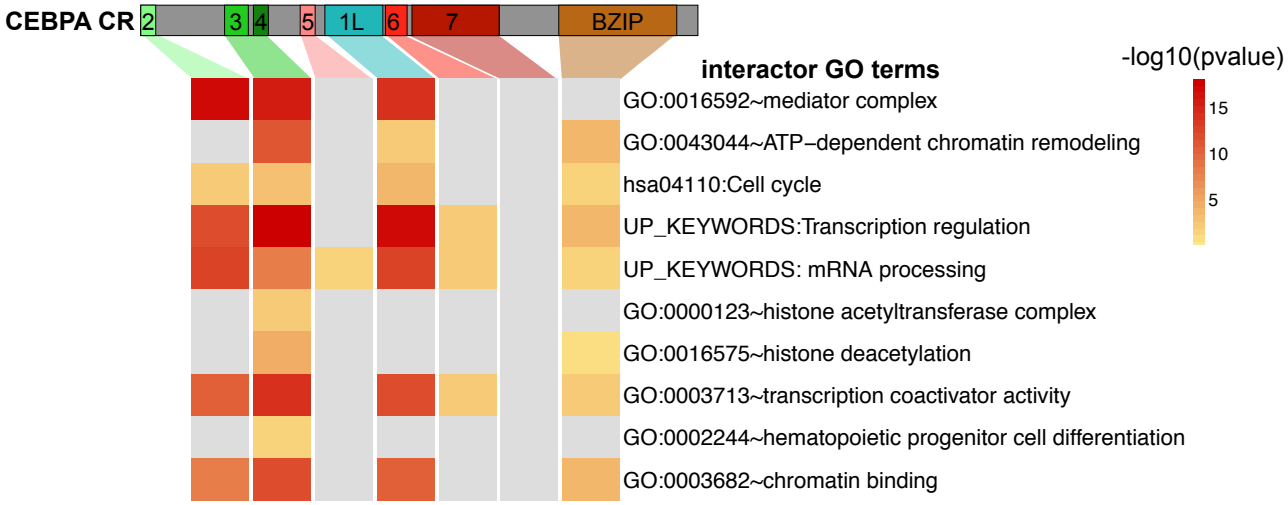
Supplemental Figure 1, Ramberger et al.



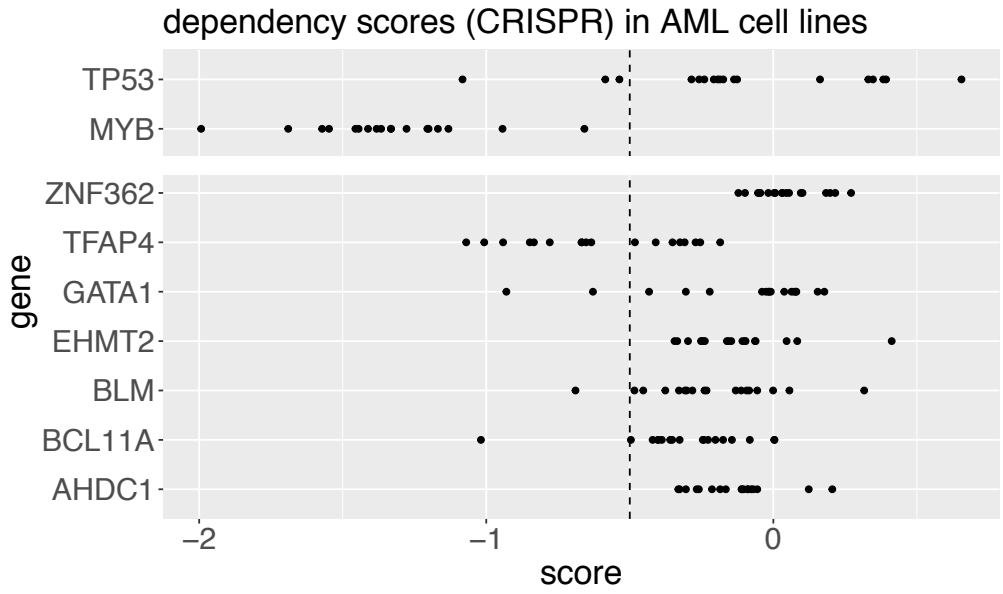
Supplemental Figure 2, Ramberger et al.



Supplemental Figure 3, Ramberger et al.



Supplemental Figure 4, Ramberger et al.



Supplemental Figure 5, Ramberger et al.