Dual DNA and Protein Tagging of Open Chromatin Unveils Dynamics of Epigenomic Landscapes in Leukemia

Dual DNA and protein tagging of open chromatin unveils dynamics of epigenomic landscapes in leukemia

Jonathan D. Lee[1,2,3,4]*, Joao A. Paulo[5], Ryan R. Posey[1,2,3], Vera Mugoni[1,2,3], Nikki R. Kong[6], Giulia Cheloni[1,2,3], Yu-Ru Lee[1,2,3,7], Frank J. Slack[8], Daniel G. Tenen[6,9], John G. Clohessy[1,2,3,10], Steven P. Gygi[5] & Pier Paolo Pandolfi[1,2,3,11,12]*

[1]Cancer Research Institute, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA.
[2]Department of Medicine and Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA.
[3]Ludwig Center at Harvard, Harvard Medical School, Boston, Massachusetts, USA.
[4]Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA.
[5]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts, USA.
[6]Harvard Stem Cell Institute, Boston, Massachusetts, USA.
[7]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.
[8]HMS Initiative for RNA Medicine, Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA.
[9]Cancer Science Institute of Singapore, Singapore, Singapore.
[10]Preclinical Murine Pharmacogenetics Facility, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA.
[11]MBC, Department of Molecular Biotechnology and Health Sciences, University of Torino, Torino, Italy.
[12]Renown Institute for Cancer, Nevada System of Higher Education, Reno, Nevada, USA.

*Correspondence to: jdlee@post.harvard.edu and pierpaolo.pandolfiderinaldis@renown.org

Abstract

The architecture of chromatin specifies eukaryotic cell identity by controlling transcription factor access to sites of gene regulation. Here we describe a dual transposase/peroxidase approach, integrative DNA And Protein Tagging (iDAPT), which detects both DNA (iDAPT-seq) and protein (iDAPT-MS) associated with accessible regions of chromatin. In addition to direct identification of bound transcription factors, iDAPT enables the inference of their gene regulatory networks, protein interactors, and regulation of chromatin accessibility. We applied iDAPT to profile the epigenomic consequences of granulocytic differentiation of acute promyelocytic leukemia, yielding previously undescribed mechanistic insights with potential therapeutic implications. Our findings demonstrate the power of iDAPT as a discovery platform for both the dynamic epigenomic landscapes and their transcription factor components associated with biological phenomena and disease.

Introduction

In the eukaryotic cell, DNA and protein intertwine as chromatin, forming a dynamic landscape comprised of genes, their regulatory sequence elements, and the transcription factor complexes modulating gene expression[1–3]. To perform their regulatory activities, transcription factor components require access to these encoded DNA elements, otherwise impeded by nucleosomal occupancy or higher-order steric hindrance[4,5]. These regions of open chromatin are continuously remodeled to control access of the transcriptional machinery and to modulate gene expression[4,6]. Thus, profiles of accessible genomic regions and their corresponding proteomes would provide a comprehensive framework to understand genome-wide transcriptional regulation, especially as it applies to cellular identity or disease.

While sequence-based profiling methods of open chromatin, such as DNase hypersensitivity[6,7] and the assay for transposase-accessible chromatin using sequencing (ATAC-seq)[8], have expanded our understanding of the interplay between chromatin states and transcription, identification of the transcription factor components associated with these accessible chromatin regions remains inferential from these datasets[9]. Specifically, these bioinformatic “footprinting” approaches are limited to sequence-specific transcription factors with long residence times on chromatin, despite known binding of a number of transcription factors with undetectable footprints[9,10]. On the other hand, mass spectrometry-based methods have emerged to characterize proteins associated with open chromatin directly such as through chromatin fractionation[11–14], yet these approaches neither specify differentially bound genomic loci nor provide insight into their transcriptional regulatory activity. To bridge these two approaches, we developed an integrative DNA And Protein Tagging (iDAPT) platform, combining biochemical enrichment via a bifunctional transposase/peroxidase probe and bioinformatic analysis of both genomic and proteomic profiles of open chromatin, attainable from a single nuclear lysate preparation (**Fig. 1a**).

Results

Tn5 transposase preferentially tags and fragments (tagments) sterically accessible DNA in native chromatin[8]. Because Tn5 transposase remains physically bound to its DNA substrate after insertion of its transposon payload[15], we hypothesized that Tn5 transposase may also serve as an anchor for proximal labeling of proteins associated with open chromatin. The APEX2 peroxidase represents an attractive choice for iDAPT due to its widespread use as a genetic tag for spatially restricted proteomic enrichment, its short labeling timeframe of one minute, and its previously described peroxidase activity as a purified protein[16,17]. For these reasons, we fused APEX2 with Tn5 transposase for sequential transposition and peroxidase-mediated biotin labeling.

We cloned and purified a series of transposase/peroxidase fusion probes consisting of APEX2 peroxidase fused either N- or C-terminal to Tn5 transposase (peroxidase/transposase [PT] and transposase/peroxidase [TP], respectively), adjoined via several linkers (L1-L5) (**Extended Data Fig. 1a-b**). C-terminal peroxidase (TP1-TP5) fusions yielded ATAC-seq library quantifications similar to commercial (Nextera) Tn5 transposase and in-house purified untagged or FLAG-tagged Tn5 transposases (C-terminal FLAG [Tn5-F] and N-terminal FLAG [F-Tn5]), whereas N-terminal peroxidase (PT1-PT5) fusions exhibited decreased transposase activity (**Extended Data Fig. 1c**). DNA fragment size analysis of ATAC-seq libraries generated from all TP fusions yielded distributions corresponding to ~200 base pair-wide nucleosomal periods typically observed with open chromatin enrichment[8] (**Extended Data Fig. 1d**). Furthermore, we observed an expected gel shift of linearized DNA in the presence of transposase domain-containing enzymes but not in the presence of FLAG-tagged APEX2 domain alone (APEX2-F)[15], with corresponding DNA fragmentation profiles dependent on both transposase-DNA association and absence of the divalent cation chelator EDTA[18] (**Extended Data Fig. 1e-f**).

Next, we generated ATAC-seq/iDAPT-seq libraries of GM12878 cells using the recently developed OmniATAC protocol, which improves signal-to-noise ratios, decreases mitochondrial read proportions, and increases assay reproducibility as compared to the original ATAC-seq protocol, with Nextera Tn5, in-house purified Tn5-F, and representative fusion probes TP3 and TP5[19]. Here we distinguish iDAPT-seq from ATAC-seq with the use of TP fusion enzymes for tagmentation, allowing for subsequent proteomic labeling and enrichment (**Fig. 1a**). ATAC-seq and iDAPT-seq libraries exhibited similar nucleosomal periodicities in their fragment size distributions, high signal-to-noise ratios, and broad decreases in mitochondrial read proportions relative to published GM12878 ATAC-seq libraries generated via the original ATAC-seq protocol[8,18–20] (**Extended Data Fig. 2a-c**). Furthermore, TP3 and TP5 iDAPT-seq libraries

exhibit high correlations with Tn5 transposase-generated ATAC-seq libraries (**Fig. 1b-c, Extended Data Fig. 2d**). Thus, TP3 and TP5 fusion enzymes yield high quality iDAPT-seq libraries, akin to ATAC-seq libraries generated via Tn5 transposase enzyme lacking a peroxidase domain.

As further assessment of TP localization to open chromatin, we performed ATAC-see, an assay of *in situ* transposase activity and localization[18], with co-immunofluorescence of various markers of chromatin state. TP3 and Tn5-F exhibit similarly positive correlations with histone H3 lysine 27 acetylation (H3K27Ac) and RNA polymerase II serine-2 phosphorylation (RNAPII S2P) immunofluorescence signals, markers of transcriptionally active chromatin, and similarly poor correlations with H3 lysine 9 trimethylation (H3K9me3) immunofluorescence, a marker of transcriptionally inactive chromatin, albeit with slight differences in colocalization patterns between the two probes (**Fig. 1d-e**). These data indicate that our TP fusion probes retain native Tn5 transposase activity and preferentially tag open chromatin.

Having confirmed TP fusion tagging of and localization to open chromatin, we next assessed APEX2 peroxidase functionality when fused with Tn5 transposase. First to confirm this, we added 1 mM hydrogen peroxide to purified proteins alone and detected peroxidase activity from the fusion proteins via resorufin fluorescence after one minute (**Supplementary Fig. 1a-b**). Interestingly, all TP fusions exhibit higher peroxidase activities than APEX2-F alone, possibly due to increased thermal stability or heme binding of APEX2 dimer formation induced by the proximity of the two C-termini of dimeric Tn5 transposase[16,21–23] (**Supplementary Fig. 1c**). Next, in extracted HEK293T nuclei, we observed strong peroxidase-dependent biotin signal in the presence of the TP3 fusion probe and low signal in the presence of the negative control probes Tn5-F and APEX2-F (**Supplementary Fig. 2**). Residual APEX2-F-mediated signal further decreased with additional washing and blocking steps while maintaining strong TP3-mediated biotin signal (**Supplementary Fig. 2**). In line with our hypothesis that Tn5 transposase remains physically bound to native chromatin, Tn5 transposase and TP3 fusion enzyme are found in the nuclear lysate, whereas APEX2 is mostly lost despite equimolar addition of recombinant protein to the tagmentation buffer (**Supplementary Fig. 1a, 2b-c**). Indeed, we found all TP fusion enzymes to promote strong biotin labeling in K562 nuclei, with TP5 and TP3 enzymes exhibiting the highest levels of labeling (**Extended Data Fig. 3a**). Finally, we confirmed that this labeling is dependent on the presence of both hydrogen peroxide and biotin-phenol (**Extended Data Fig. 3b**). Thus, our findings indicate that TP probes label transposase-accessible chromatin in a peroxidase-dependent manner.

With our optimized iDAPT protocol, we performed quantitative mass spectrometry on the iDAPT-enriched proteome (iDAPT-MS) from K562 nuclei[24] (**Fig. 2a, Supplementary Table 1**). As negative control probes enrich for nonspecific background signal, akin to an IgG negative control for an immunoprecipitation assay, we interpreted the substantial proteomic content enriched by TP over negative control probes as *bona fide* proteins proximal to Tn5 transposase localization in isolated nuclei (**Fig. 2b**). By hierarchical clustering and correlation analyses, nuclear lysates labeled via TP3 and TP5 segregate from lysates labeled via single enzymatic domains, with substantial overlap between TP3- and TP5-enriched proteomes (**Extended Data Fig. 3c-d**). We observed a similarly substantial iDAPT-MS enrichment pattern from TP3 versus negative control probes from the NB4 cell line, incorporating an additional wash step to block endogenous peroxidase activity prior to tagmentation and biotin labeling (**Extended Data Fig. 4, Supplementary Table 2**).

To validate highly enriched proteins by iDAPT-MS, we performed CUT&RUN (ERH and WBP11) and analyzed published ENCODE ChIP-seq datasets from the K562 cell line[25,26] (**Supplementary Table 3**). We found substantial enrichment of protein binding at sites of open chromatin (**Fig. 2c, Extended Data Fig. 5**). These results demonstrate the ability of iDAPT-MS to discover proteins associated with open chromatin.

Next, we performed enrichment analyses of our iDAPT-MS datasets. Subcellular enrichment analysis identified nuclear speckles and nucleoplasm in both K562 and NB4 iDAPT-MS datasets[27] (**Extended Data Fig. 6a-b**). Indeed, ATAC-see signal of Tn5-F colocalizes with the nuclear speckle marker SC35 in multiple cell lines, in agreement with recent reports of nuclear speckle localization at active promoters[28,29] (**Fig. 2d, Extended Data Fig. 6c-e**). We further identified significant enrichment of protein complexes such as Mediator, which regulates communication from enhancer- and promoter-bound transcription factors to RNA polymerase II[30], and BAF, which remodels chromatin accessibility[31], in both K562 and NB4 cell lines[32] (**Fig. 2e-f**). Chromatin remodelers and RNA-binding proteins were highly represented (>50% of annotated proteins) among enriched proteins, whereas transcription factors and histone variants were not as well represented (<25% of annotated proteins) (**Extended Data Fig. 6f**). While histone protein H2AX/H2AFX was highly enriched in both NB4 and K562 iDAPT-MS proteomes, other detected histone proteins were weakly enriched over negative control probes or not detected, suggesting that histone proteins as a class are not predominantly enriched by iDAPT-MS (**Fig. 2b, Extended Data Fig. 4c, 6f-g**).

Despite low background peroxidase signal, APEX2-F yields some proteomic enrichment over Tn5-F, although not as strongly as signal generated by TP3/TP5 (**Supplementary Fig. 3a-**

f). To assess whether APEX2-F has a different labeling propensity over TP3/TP5 fusion probes in K562 nuclei, we used quantile normalization as a proxy for normalizing APEX2-F peroxidase activity with TP3 and TP5 activities (**Supplementary Fig. 3g**). We found this quantile normalization scheme to yield similar subcellular enrichment patterns, albeit with increased mitochondrial enrichment, as with our primary streptavidin/trypsin peptide normalization scheme (**Extended Data Fig. 6a, Supplementary Fig. 3h**). Taken together, these data suggest that TP fusion proteins exhibit different labeling patterns from diffusely nuclear APEX2.

Next, we compared iDAPT-MS enrichment relative to other techniques used to assess protein abundance on chromatin. First, we collated sets of detected proteins from K562 RNA-seq (protein-coding transcripts)[25], whole cell proteome[33], and nuclear proteome[34] datasets and then assessed the proportions of proteins detected across subcellular compartments in each of these datasets to normalize for proteome complexity. While we observed mild subcellular enrichment differences between RNA-seq and whole cell proteome datasets, we found increased enrichment of nucleoli, nucleoplasm, and nucleus localization terms from iDAPT-MS and nuclear proteome datasets (**Supplementary Fig. 4a-b**). The K562 iDAPT-MS-enriched proteome exhibits increased enrichment of nuclear speckles, nucleoplasm, and nuclear body localization terms and decreased cytosolic, plasma membrane, and Golgi apparatus localization terms over the nuclear proteome (**Supplementary Fig. 4b**). Second, we assessed how iDAPT-MS enrichment compares with incremental salt extractions from K562 nuclei, partitioning euchromatic and heterochromatic proteins via disrupting electrostatic protein-protein and protein-DNA interactions[34] (**Supplementary Fig. 4c-d**). After converting protein sets to subcellular enrichment scores and performing principal component analysis, we found that K562 iDAPT-MS coincides with proteins identified by both isotonic and 250 mM salt extractions along the first principal component, largely representing euchromatic proteins. Third, we compared iDAPT-MS enrichment with additional published salt extraction- and micrococcal nuclease (MNase) fragmentation-based chromatin proteomic datasets in a similar manner[12–14] (**Supplementary Fig. 4e-f**). Indeed, iDAPT-MS enrichment corresponds with chromatin proteomes enriched by light MNase digestion and salt extraction along the first principal component. Together, these findings demonstrate that iDAPT-MS enriches for the open chromatin proteome.

A critical advantage of iDAPT-MS over ATAC-seq/iDAPT-seq or chromatin immunoprecipitation (ChIP)-based approaches is its ability to capture numerous transcription co-factors associated with open chromatin in a single assay, which regulate their associated sequence-specific transcription factors. As proof of principle, we found the MAX protein

interaction network to be significantly enriched on open chromatin by K562 iDAPT-MS[35] (**Fig. 2g**). To validate this finding, by ChIP-seq analysis, protein interactors of MAX colocalize more tightly with MAX across the open chromatin landscape than do non-interacting proteins (**Fig. 2h, Supplementary Table 3**). Therefore, iDAPT-MS together with protein interaction annotations facilitates the identification of active transcription factor protein complexes on open chromatin, expanding the inference of *cis*-regulatory transcription factor networks.

Transcription factors regulate gene expression by binding to DNA in a sequence-specific manner and recruiting transcriptional activators and/or repressors to their target genes. Most transcription factors are found within regions of open chromatin, a pattern we also observed in our iDAPT-MS data[3,6,36] (**Fig. 3a, Extended Data Fig. 7a**). As iDAPT enables profiling of both genomic and proteomic content of the open chromatin landscape, we sought to compare transcription factor enrichment profiles obtained from iDAPT-MS and iDAPT-seq approaches. To assess the enrichment of transcription factors obtained via iDAPT-seq, we profiled both nuclei and "naked" genomic DNA from both K562 and NB4 cell lines. iDAPT-seq analysis confirms loss of both nucleosomal enrichment and promoter insertion preference in naked DNA; furthermore, insertion profiles segregate along the first principal component and exhibit skewed statistical significance towards chromatinized peaks in both datasets (**Extended Data Fig. 7b-h**).

With these iDAPT-seq profiles, we performed footprinting analysis to infer transcription factor activities at their cognate motifs. By a genome-wide bivariate footprinting approach, accounting for both transcription factor footprint depth (FPD) and flanking chromatin accessibility (FA) near the transcription factor motif, we observed significant enrichment of most CisBP transcription factor motifs in iDAPT-seq profiles from native chromatin[10,36] (**Fig. 3b-c, Extended Data Fig. 8a-c**). We categorized motifs emerging from our footprint analysis into three classes: strong footprinting (class A), weak footprinting (class B), and no or negative footprinting (class C) (**Extended Data Fig. 8d**). In line with previous reports, transcription factors with longer residence times on chromatin exhibit stronger footprints: for instance, CTCF, an insulator protein with a long retention time on DNA, exhibits a strong footprint (class A) and is detected by both iDAPT-MS and ChIP-seq[9,37] (**Fig. 3d**). RELA/NF-κB complexes (class B) have short DNA residence times and substantially weaker footprinting potential, despite being detected by both iDAPT-MS and ChIP-seq[38] (**Fig. 3e**). While class C motifs such as IKZF1 exhibit nonsignificant or even significantly negative footprinting activity, several of these transcription factors are nonetheless found on open chromatin by both iDAPT-MS and ChIP-seq (**Fig. 3f-h**). Broadly, we observed no clear relationship between inferred transcription factor footprint activity by iDAPT-seq and magnitude of transcription factor abundance by iDAPT-MS

(**Fig. 3g, Extended Data Fig. 8e**). Indeed, ChIP-seq and iDAPT-MS both directly identify transcription factors spanning all three classes of footprint activities (**Fig. 3h, Supplementary Table 3**), yet neither assay alone can inform how transcription factor binding might affect chromatin accessibility. Conversely, footprinting analysis of iDAPT-seq is able to detect changes to chromatin accessibility, but these changes may be independent of whether a transcription factor is bound or not. Thus, we posit that, for the analysis of transcription factors with annotated motifs, iDAPT-seq and iDAPT-MS together identify transcription factors bound to open chromatin and reveal their activity on chromatin accessibility as a consequence of their abundance, providing greater insight into transcription factor mechanisms than either assay alone.

Next, we assessed how transcription factor abundances and chromatin accessibility states correlate upon granulocytic differentiation of the NB4 acute promyelocytic leukemia (APL) cell line. Differentiation of NB4 cells via all-*trans* retinoic acid (ATRA) leads to degradation of the PML-RARA oncogenic fusion protein, decreased proliferation, and granulocytic differentiation of the leukemia[39] (**Fig. 4a-b, Extended Data Fig. 9a-c**). iDAPT-MS reveals a dramatic shift in the open chromatin proteome, with profiles clustering by treatment (**Extended Data Fig. 4b, d**). In line with previous reports, we observed negative enrichment of RARA, degraded upon ATRA treatment[40,41], and positive enrichment of PU.1/SPI1, CEBPB, and CEBPE, upregulated in response to ATRA[42–44] (**Extended Data Fig. 9d**). Pathway enrichment analysis reveals positive associations with MAPK signaling, neutrophil differentiation, and the innate immune response (**Extended Data Fig. 9e**). On the other hand, loss of histone deacetylase enrichment, the most significantly negative pathway, may explain the previously described decrease in histone acetylation states and sensitivity to histone deacetylase inhibitors in APL[45,46]. These observations validate the ability of iDAPT-MS to capture both specific proteins and proteomic signatures as they dynamically shift upon changes in cell identity.

Given the different transcription factor classes captured by iDAPT at steady state, we explored how transcription factor activities and abundances change on open chromatin upon ATRA-mediated cellular differentiation. By iDAPT-seq, we observed both increased and decreased regions of open chromatin and motif footprinting activity upon ATRA treatment, with footprinting parameters FPD and FA correlating strongly with composite footprinting scores (**Supplementary Fig. 5**). Intriguingly, both concordant and discordant enrichment patterns between iDAPT-seq and iDAPT-MS transcription factor enrichment profiles were observed (**Fig. 4c**). Furthermore, some transcription factors exhibit only one of either differential footprinting or protein abundance, discrepancies that have been observed previously between chromatin

accessibility and chromatin immunoprecipitation-based assays[9,10] (**Fig. 4c**). To corroborate our findings, we replaced our iDAPT-seq footprinting and iDAPT-MS analyses with either motif enrichment analysis via ChromVAR or RNA-seq analysis, which correlates well with our iDAPT-MS protein analysis, both yielding similar transcription factor patterns[47–49] (**Supplementary Fig. 6-7**). Hence, iDAPT reveals nine distinct classes (classes I-IX) arising as a consequence of integrating both iDAPT-seq, a readout of transcription factor activity, and iDAPT-MS, a readout of transcription factor protein abundance at open chromatin (**Fig. 4c, Extended Data Fig. 10a**). Furthermore, we interpreted concordance (classes III, VII) as chromatin activating activity by the transcription factor of interest and discordance (classes I, IX) as chromatin repression (**Fig. 4c, Extended Data Fig. 10a**). In support of this functional classification scheme, among transcription factors decreasing in abundance upon ATRA treatment, those classified as activating (class VII), which should be easier to tag by TP fusion proteins in the vehicle-treated setting, are generally more enriched by TP3 over negative control probes than repressive transcription factors (class I) (**Extended Data Fig. 10b**). Thus, iDAPT-MS and iDAPT-seq together uncover functional relationships between transcription factor binding dynamics and chromatin accessibility, which neither assay can elucidate alone.

As iDAPT-MS reveals abundance changes of proteins beyond transcription factors, we assessed how proteins interacting with transcription factors may cooperate to regulate chromatin accessibility states. For a given transcription factor, we superimposed iDAPT-MS protein abundance changes onto its first-order protein interaction network from BioGrid[35]. Of these putative transcription factor complex profiles, we found the PU.1/SPI1 protein interaction network to be the most significantly decreased complex upon ATRA treatment (**Fig. 4d**). Intriguingly, while many of its protein interactors such as the transcriptional corepressor SIN3A decrease in abundance, PU.1/SPI1 itself increases in abundance to promote chromatin accessibility at its cognate motif (class III)[42,50] (**Fig. 4d-e**). Furthermore, the decrease in RARA protein abundance, also an interactor of PU.1/SPI1, leads to increased chromatin accessibility at its binding motif due to its ATRA-mediated degradation, implicating its transcriptional repressive activity (class I)[51] (**Supplementary Fig. 8a**). Thus, in the APL setting, transcriptional repressors bind to PU.1/SPI1 to repress chromatin accessibility at PU.1/SPI1 motifs; this repressive binding is relieved upon ATRA treatment, enabling PU.1/SPI1 to activate transcription at its motifs. This analysis may be extended to other transcription factors and their protein complexes: BCL11A, together with many of its annotated protein interactors, decreases in abundance while increasing chromatin accessibility upon ATRA treatment (class I), suggestive of a coordinated downregulation of this repressive transcription factor and its protein

complex components[52] (**Fig. 4f-g**). While JUNB[53–55], CEBPB[56], and CEBPE[57] have both activating and repressive behaviors reported, we observed class VII activating behavior from the JUNB transcription factor and class IX repressive behavior from the CEBPB and CEBPE transcription factors upon ATRA treatment, with their dynamic protein complex components providing potential context-specific insights into their regulatory activities on chromatin state (**Supplementary Fig. 8b-c**). In this manner, integrating protein interaction information with iDAPT-MS and iDAPT-seq profiles reveals the interplay between transcription factors, their activities on chromatin accessibility, and their putative protein complexes as these components change during ATRA treatment of NB4 cells.

Given the numerous transcription factors and associated components differentially bound at open chromatin upon ATRA treatment, some of these newly identified proteins may have functional roles in APL differentiation. We superimposed our iDAPT-MS results with NB4 genetic dependencies and identified both PML and RARA, corroborating our analysis[58] (**Fig. 4h**). After filtering out essential genes across hematopoietic cell lines, we identified a number of candidate transcription factor effectors, including CEBPA, EBF3, and ZEB2, which may act downstream or independently of PML-RARA (**Fig. 4h, Supplementary Fig. 9**). In agreement with previous reports, our transcription factor classification scheme assigns ZEB2 as repressive[59] (class I) and EBF3[60–62] and CEBPA[63] as activating (class VII) (**Fig. 4c, Supplementary Fig. 9c-d**). This analysis reifies the power of combining forward genetic screens with iDAPT-MS to identify critical transcription factors and their regulators for a given biological phenotype.

Finally, we assessed how our interpretations of transcription factor dynamics would change between iDAPT-MS, measuring protein abundances directly, and RNA-seq profiles. While we observed a positive correlation between iDAPT-MS and RNA-seq profiles upon ATRA treatment, several discordant cases emerged, including JUNB/JUND and RARA, with their RNA-seq effect sizes opposite in magnitude of their corresponding iDAPT-MS effects (**Fig. 4c, Supplementary Fig. 7b-c**). Indeed, ATRA binds to RARA, and prolonged ligand binding and transcriptional activity leads to RARA protein degradation[40] (**Supplementary Fig. 8a**). Furthermore, as transcript levels of RARA and several other protein interactors of PU.1/SPI1 do not fully match iDAPT-MS enrichment trends, the significantly negative enrichment of the PU.1/SPI1 protein complex observed upon ATRA treatment by iDAPT-MS is lost by RNA-seq (**Supplementary Fig. 10**). Thus, among open chromatin-associated proteins, bulk RNA-seq may broadly provide similar patterns as iDAPT-MS, but discrepancies between the two limit the ability of RNA-seq to replace proteomic analysis.

341
342

Discussion

In summary, we have developed iDAPT to capture both the genomic and proteomic contents of open chromatin, realized via a recombinant transposase/peroxidase probe. Integrative analysis of iDAPT-seq and iDAPT-MS profiles reveals nine transcription factor classes based on both changes in protein abundance on open chromatin (decreased, unchanged, or increased) and transcription factor activity (closed, unchanged, open) (**Fig. 4c, Extended Data Fig. 10a**). Furthermore, iDAPT-MS together with protein interaction annotations implicates changes in transcription factor complex compositions that may explain the corresponding changes in chromatin accessibility (**Fig. 4d-g, Supplementary Figs. 8-9**). Identification of such relationships between transcription factors, their protein complex components, and their functional outputs on chromatin accessibility may be informative for mechanistic and therapeutic study, especially in conjunction with genetic screening approaches. Indeed, in the context of APL, our analyses suggest targets for which approved therapies already exist, such as histone deacetylases, and those which may warrant further investigation, such as EBF3 and ZEB2.

From our transcription factor classification scheme, we are able to assign activating or repressive activities to sequence-specific transcription factors based on their concordance or discordance between iDAPT-MS and iDAPT-seq profiles. At the heart of this finding is the question, if repressive factors close chromatin at their cognate binding sites, how are they still detected by iDAPT-MS? Due to chromatin “breathing” or stochastic transposition, Tn5 transposase may insert proximal to repressive transcription factors on chromatin, albeit at a decreased frequency as compared to activating transcription factors, enabling the tagging of such repressive factors for mass spectrometry detection. In support of this explanation, as in **Fig. 4g** and **Supplementary Figs. 8a** and **c**, repressive transcription factors (classes I and IX) exhibit detectable transposase activity proximal to their cognate binding motifs above background in both ATRA- and control-treated cells. On the other hand, the inference of transcription factor activity via genome-wide footprinting from iDAPT-seq/ATAC-seq datasets may be partially artifactual, leading to misleading classifications of transcription factor activity. First, footprinting analysis relies on the quality of curated DNA binding motifs, whereas actual transcription factor localization to open chromatin may not be restricted to such motif-containing chromatin regions. Second, genome-wide footprinting analysis in bulk may mask locus-specific or cell-specific transcription factor activities, a consequence of broadly enriching for transposase-accessible chromatin, only one of many regulatory features of gene expression. Thus, the combination of iDAPT-MS and iDAPT-seq provides a powerful opportunity to identify

such key relationships between transcription factor abundance and genome-wide regulation of chromatin accessibility.

In addition to chromatin accessibility state, additional factors such as histone and DNA modifications may modulate transcription factor activity at a given genetic locus[25]. To explore these relationships further, complementary methods to identify the transcription factors and associated proteins at these specific chromatin states include ChIP-based enrichment[64] and proximity labeling via chromatin reader domains[65]. At a finer genetic resolution are locus-specific enrichment methods, including recently developed CRISPR/Cas9-based proximity labeling approaches[11,66]. Integrating these methods with assays of the accessible genome such as ATAC-seq may reveal context-specific transcription factor activities and protein complex compositions that iDAPT would not reveal. On the other hand, classification of global transcription factor activities via iDAPT may better inform their regulation of cellular phenotypes, encompassing mechanistic information across all of its binding sites. Furthermore, as iDAPT does not require genetic manipulation of biological samples of interest as with traditional APEX2 or biotin ligase genetic tagging[16,17,66], our approach may be readily applied to numerous biological systems to uncover novel chromatin-level molecular correlates and mechanistic insights. Thus, our findings substantiate the unprecedented capability of iDAPT to unravel epigenomic landscapes as they change during development and disease.

Acknowledgments

We thank J. Boehm, P. Cheung, J. Harper, J. Heo, P. Kharchenko, and all members of the Pandolfi laboratory for their input. We are grateful to the Harvard Medical School Biopolymers Facility, Harvard Medical School Research Computing, and BIDMC Confocal Imaging Core for their assistance and support. This work was supported in part by the Ludwig Center at Harvard, the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research (STaR) Investigator Award, the National Research Foundation Singapore and the Singapore Ministry of Education under its Research Centres of Excellence initiative, a Harvard Medical School Innovation Grant Program grant awarded to J.D.L., and National Institutes of Health (NIH) grants R01 GM132129 to J.A.P., R35 CA232105 to F.J.S., R35 CA197697 and P01 HL131477 to D.G.T., GM67945 to S.P.G., and R35 CA197529 to P.P.P.

Author Contributions

J.D.L. conceived the project, supervised the study, designed and performed experiments, carried out computational analyses, and wrote the manuscript. J.A.P. performed mass spectrometry analyses. R.R.P. designed and performed experiments and performed image analyses. V.M. designed and performed experiments. N.R.K. designed and performed experiments. G.C. designed and performed experiments. Y.-R.L. designed and performed experiments. F.J.S. supervised the study. D.G.T. supervised the study. J.G.C. supervised the study. S.P.G. supervised the study. P.P.P. conceived the project, supervised the study, and wrote the manuscript.

Ethics Declaration

J.D.L., J.G.C., and P.P.P. have filed a patent describing iDAPT. All other authors declare no competing interests.

References

1. Kornberg, R. and Lorch, Y. Chromatin Structure and Transcription. *Annu. Rev. Cell Dev. Biol.* **8**, 563–587 (1992).
2. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
3. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
4. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
5. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).
6. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
7. Boyle, A. P. *et al.* High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* **132**, 311–322 (2008).
8. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–8 (2013).
9. Sung, M.-H., Baek, S. & Hager, G. L. Genome-wide footprinting: ready for prime time? *Nat. Methods* **13**, 222–228 (2016).
10. Baek, S., Goldstein, I. & Hager, G. L. Bivariate Genomic Footprinting Detects Changes in Transcription Factor Activity. *Cell Rep.* **19**, 1710–1722 (2017).
11. Wierer, M. & Mann, M. Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. *Hum. Mol. Genet.* **25**, R106–R114 (2016).
12. Torrente, M. P. *et al.* Proteomic Interrogation of Human Chromatin. *PLoS One* **6**, e24747 (2011).
13. Alajem, A. *et al.* Differential association of chromatin proteins identifies BAF60a/SMARCD1 as a regulator of embryonic stem cell differentiation. *Cell Rep.* **10**, 2019–2031 (2015).
14. Kulej, K. *et al.* Time-resolved Global and Chromatin Proteomics during Herpes Simplex Virus Type 1 (HSV-1) Infection. *Mol. Cell. Proteomics* **16**, S92–S107 (2017).
15. Reznikoff, W. S. Transposon Tn5. *Annu. Rev. Genet.* **42**, 269–286 (2008).
16. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2014).

17. Paek, J. *et al.* Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell* **169**, 338-349.e11 (2017).
18. Chen, X. *et al.* ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
19. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
20. Johnston, A. D., Simões-Pires, C. A., Thompson, T. V., Suzuki, M. & Greally, J. M. Functional genetic variants can mediate their regulatory effects through alteration of transcription factor binding. *Nat. Commun.* **10**, 3472 (2019).
21. Martell, J. D., Deerinck, T. J., Lam, S. S., Ellisman, M. H. & Ting, A. Y. Electron microscopy using the genetically encoded APEX2 tag in cultured mammalian cells. *Nat. Protoc.* **12**, 1792–1816 (2017).
22. Mandelman, D., Li, H., Poulos, T. L. & Schwarz, F. P. The role of quaternary interactions on the stability and activity of ascorbate peroxidase. *Protein Sci.* **7**, 2089–2098 (1998).
23. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S. & Rayment, I. Three-Dimensional Structure of the Tn5 Synaptic Complex Transposition Intermediate. *Science (80-.).* **289**, 77–85 (2000).
24. Navarrete-Perea, J., Yu, Q., Gygi, S. P. & Paulo, J. A. Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3. *J. Proteome Res.* **17**, 2226–2236 (2018).
25. Encode Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
26. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, (2017).
27. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science (80-.).* **356**, eaal3321 (2017).
28. Xiao, R. *et al.* Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell* **178**, 107-121.e18 (2019).
29. Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548 (2019).
30. Allen, B. L. & Taatjes, D. J. The Mediator complex: A central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 155–166 (2015).
31. Kadoch, C. & Crabtree, G. R. Mammalian SWI/SNF chromatin remodeling complexes

and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* **1**, 1–18 (2015).
32. Ruepp, A. *et al.* CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.* **36**, D646-50 (2008).
33. Nusinow, D. P. *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387-402.e16 (2020).
34. Federation, A. J. *et al.* Highly Parallel Quantification and Compartment Localization of Transcription Factors and Nuclear Proteins. *Cell Rep.* **30**, 2463-2471.e5 (2020).
35. Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47**, D529–D541 (2019).
36. Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
37. Nakahashi, H. *et al.* A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.* **3**, 1678–1689 (2013).
38. Bosisio, D. *et al.* A hyper-dynamic equilibrium between promoter-bound and nucleoplasmic dimers controls NF-kappaB-dependent gene activity. *EMBO J.* **25**, 798–810 (2006).
39. Lanotte, M. *et al.* NB4, a maturation inducible cell line with t(15;17) marker isolated from a human acute promyelocytic leukemia (M3). *Blood* **77**, 1080–6 (1991).
40. Zhu, J. *et al.* Retinoic acid induces proteasome-dependent degradation of retinoic acid receptor alpha (RARalpha) and oncogenic RARalpha fusion proteins. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14807–12 (1999).
41. de Thé, H. *et al.* The PML-RAR alpha fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR. *Cell* **66**, 675–84 (1991).
42. Mueller, B. U. *et al.* ATRA resolves the differentiation block in t(15;17) acute myeloid leukemia by restoring PU.1 expression. *Blood* **107**, 3330–8 (2006).
43. Chih, D. Y., Chumakov, A. M., Park, D. J., Silla, A. G. & Koeffler, H. P. Modulation of mRNA expression of a novel human myeloid-selective CCAAT/enhancer binding protein gene (C/EBP epsilon). *Blood* **90**, 2987–94 (1997).
44. Duprez, E., Wagner, K., Koch, H. & Tenen, D. G. C/EBPbeta: a major PML-RARA-responsive gene in retinoic acid-induced differentiation of APL cells. *EMBO J.* **22**, 5806–16 (2003).
45. Martens, J. H. A. *et al.* PML-RARalpha/RXR Alters the Epigenetic Landscape in Acute Promyelocytic Leukemia. *Cancer Cell* **17**, 173–85 (2010).

524 46. Warrell, R. P., He, L. Z., Richon, V., Calleja, E. & Pandolfi, P. P. Therapeutic targeting of
525 transcription in acute promyelocytic leukemia by use of an inhibitor of histone deacetylase.
526 *J. Natl. Cancer Inst.* **90**, 1621–5 (1998).
527 47. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring
528 transcription-factor-associated accessibility from single-cell epigenomic data. *Nat.*
529 *Methods* **14**, 975–978 (2017).
530 48. Witzel, M. *et al.* Chromatin-remodeling factor SMARCD2 regulates transcriptional
531 networks controlling differentiation of neutrophil granulocytes. *Nat. Genet.* **49**, 742–752
532 (2017).
533 49. Orfali, N. *et al.* All-trans retinoic acid (ATRA)-induced TFEB expression is required for
534 myeloid differentiation in acute promyelocytic leukemia (APL). *Eur. J. Haematol.* **104**,
535 236–250 (2020).
536 50. Hu, Z. *et al.* RUNX1 regulates corepressor interactions of PU.1. *Blood* **117**, 6498–508
537 (2011).
538 51. Wang, K. *et al.* PML/RARalpha targets promoter regions containing PU.1 consensus and
539 RARE half sites in acute promyelocytic leukemia. *Cancer Cell* **17**, 186–97 (2010).
540 52. Liu, N. *et al.* Direct Promoter Repression by BCL11A Controls the Fetal to Adult
541 Hemoglobin Switch. *Cell* **173**, 430-442.e17 (2018).
542 53. Li, B., Tournier, C., Davis, R. J. & Flavell, R. A. Regulation of IL-4 expression by the
543 transcription factor JunB during T helper cell differentiation. *EMBO J.* **18**, 420–432 (1999).
544 54. Schütte, J. *et al.* jun-B inhibits and c-fos stimulates the transforming and trans-activating
545 activities of c-jun. *Cell* **59**, 987–997 (1989).
546 55. Chiu, R., Angel, P. & Karin, M. Jun-B differs in its biological properties from, and is a
547 negative regulator of, c-Jun. *Cell* **59**, 979–986 (1989).
548 56. Descombes, P. & Schibler, U. A liver-enriched transcriptional activator protein, LAP, and
549 a transcriptional inhibitory protein, LIP, are translated from the sam mRNA. *Cell* **67**, 569–
550 579 (1991).
551 57. Bedi, R., Du, J., Sharma, A. K., Gomes, I. & Ackerman, S. J. Human C/EBP-ϵ activator
552 and repressor isoforms differentially reprogram myeloid lineage commitment and
553 differentiation. *Blood* **113**, 317–327 (2009).
554 58. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity
555 of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
556 59. Postigo, A. A. & Dean, D. C. Differential expression and function of members of the zfh-1
557 family of zinc finger/homeodomain repressors. *Proc. Natl. Acad. Sci.* **97**, 6391–6396

(2000).
60. Sleven, H. *et al.* De Novo Mutations in EBF3 Cause a Neurodevelopmental Syndrome. *Am. J. Hum. Genet.* **100**, 138–150 (2017).
61. Chao, H.-T. *et al.* A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. *Am. J. Hum. Genet.* **100**, 128–137 (2017).
62. Harms, F. L. *et al.* Mutations in EBF3 Disturb Transcriptional Profiles and Cause Intellectual Disability, Ataxia, and Facial Dysmorphism. *Am. J. Hum. Genet.* **100**, 117–127 (2017).
63. Pabst, T. *et al.* Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-α (C/EBPα), in acute myeloid leukemia. *Nat. Genet.* **27**, 263–270 (2001).
64. Ji, X. *et al.* Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3841–6 (2015).
65. Villaseñor, R. *et al.* ChromID identifies the protein interactome at chromatin marks. *Nat. Biotechnol.* (2020). doi:10.1038/s41587-020-0434-2
66. Gauchier, M., van Mierlo, G., Vermeulen, M. & Déjardin, J. Purification and enrichment of specific chromatin loci. *Nat. Methods* **17**, 380–389 (2020).

Figure Legends

Fig. 1. Transposase/peroxidase fusion probes tag DNA at regions of open chromatin. (a) Schematic of integrative DNA And Protein Tagging (iDAPT). TP, transposase/peroxidase fusion protein. (b) Integrative Genomics Viewer (IGV) genome track view of ATAC-seq (Nextera Tn5, Tn5-F) and iDAPT-seq (TP3, TP5) libraries at a ubiquitously accessible control region. Libraries were generated from the GM12878 cell line. (c) Scatterplots comparing genome-wide transposon insertion frequencies of Nextera Tn5 (ATAC-seq) with either in-house Tn5-F (ATAC-seq) or the transposase/peroxidase fusion TP3 (iDAPT-seq) in the GM12878 cell line. Pearson correlation coefficients are displayed inline. (d) Distribution of Pearson correlation coefficients between TP3 or Tn5-F ATAC-see and co-immunostaining of markers of active chromatin (RNA Pol II S2P, H3K27Ac) or repressive chromatin (H3K9me3) per nucleus in the HT1080 cell line. Numbers of nuclei assessed per marker are displayed inline, with images obtained from a single experiment. Center line, median value; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. p-values, two-sided Wilcoxon rank-sum test with Bonferroni correction. (e) Representative images of co-immunofluorescence staining of chromatin state markers with TP3 ATAC-see in the HT1080 cell line. Similar results were visually confirmed for more than ten nuclei for each chromatin marker and are quantified in (d). Scale bars, 5 μm.

Fig. 2. iDAPT-MS reveals the open chromatin-associated proteome. (a) Schematic of iDAPT-MS experimental design and SL-TMT sample labeling for K562 profiling. (b) Volcano plot of proteins enriched by fusion (TP3 and TP5) versus negative control (Tn5-F and APEX2-F) probes in K562 nuclei. Blue points, log2 fold change > 0 and false discovery rate (FDR) < 5%; red points, CisBP sequence-specific transcription factors; black points, points with corresponding gene symbol labels. (c) IGV genome track view of iDAPT-seq (TP3) libraries generated from either intact nuclei or genomic DNA from K562 cells and CUT&RUN libraries from K562 nuclei using ERH, WBP11, or normal rabbit IgG antibodies. (d) Representative images of co-immunofluorescence staining of the SC35 nuclear speckle marker with Tn5-F ATAC-see in the HT1080 cell line. Similar results were visually confirmed for more than ten nuclei for each chromatin marker and are quantified in **Extended Data Fig. 6c**. Scale bars, 5 μm. (e and f) Mediator (e) and BAF (f) CORUM complex enrichment by iDAPT-MS with fusion probes in both K562 and NB4 cell lines. NES (normalized enrichment score) and p-value, gene set enrichment analysis. Legend, individual protein-level iDAPT-MS enrichment. (g) MAX BioGrid first-order protein interaction network enrichment by iDAPT-MS with fusion probes in the

K562 cell line. NES (normalized enrichment score) and p-value, gene set enrichment analysis. Legend, individual protein-level iDAPT-MS enrichment. (h) Distribution of Jaccard indices between MAX ChIP-seq peaks and ChIP-seq peaks of first-order protein interactors within regions of open chromatin in the K562 cell line. MAX ChIP 1, ENCFF618VMC. MAX ChIP 2, ENCFF900NVQ. BG, background ChIP-seq epitopes, collated from ENCODE K562 ChIP-seq datasets of proteins not annotated to interact with MAX by BioGrid. Center line, median value; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; black points, outliers. Red point, replicate MAX ChIP-seq epitope. p-values, two-sided Wilcoxon rank-sum test. *n,* number of represented ChIP-seq epitopes.

Fig. 3. Integrative analysis of iDAPT-MS and iDAPT-seq classifies transcription factor activities on open chromatin at steady state. (a) Enrichment of CisBP sequence-specific transcription factors via K562 iDAPT-MS. Normalized enrichment score (NES) and p-value, gene set enrichment analysis. (b) Schematic of bivariate footprinting analysis of iDAPT-seq data. FPD, footprint depth. FA, flanking accessibility. (c) Bivariate footprinting analysis of native chromatin versus naked genomic DNA from the K562 cell line. Red, class A transcription factors; blue, class B transcription factors; gray, class C transcription factors. (d-f) K562 genome-wide footprint of CTCF (d, class A), RELA/p65 (e, class B), and IKZF1 (f, class C) from native chromatin (red) and naked DNA (black). Corresponding iDAPT-MS and ENCODE ChIP-seq enrichment metrics are listed below. iDAPT-MS LFC, log2 fold change; FDR, limma false discovery rate. ChIP-seq NES, normalized enrichment score; p, gene set enrichment analysis p-value. (g) Comparison of CisBP sequence-specific transcription factors enriched by iDAPT-MS versus iDAPT-seq footprinting analysis in the K562 cell line. (h) Number of significant CisBP transcription factors in each footprinting class as determined by iDAPT-MS or ENCODE ChIP-seq, with corresponding numbers of associated transcription factor motifs per class as determined by iDAPT-seq.

Fig. 4. iDAPT profiling of the NB4 acute promyelocytic leukemia cell line upon all-*trans* retinoic acid (ATRA) treatment reveals dynamics of transcription factor activity. (a) Schematic of the consequences of PML-RARA fusion oncogene on hematopoiesis and relief of its differentiation blockade by ATRA treatment. (b) Representative flow cytometry plots of NB4 cells treated with or without ATRA after 48 hrs. (c) Comparison of CisBP sequence-specific transcription factor enrichment by TP3 iDAPT-MS (log2 fold change) versus iDAPT-seq footprinting analysis (composite footprinting score) in the NB4 cell line upon treatment with

either ATRA or DMSO. Roman numerals, transcription factor classification as described in **Extended Data Fig. 10a**. (d-g) PU.1/SPI1 and BCL11A BioGrid first-order protein interaction networks (d and f) and corresponding genome-wide motif footprints (e and g) upon treatment with either ATRA (red) or DMSO (black) in the NB4 cell line. NES (normalized enrichment score) and p-value, gene set enrichment analysis. Legend, individual protein-level iDAPT-MS enrichment. (h) Assessment of NB4 cell line-specific genetic dependencies versus NB4 iDAPT-MS negative enrichment upon ATRA treatment. Dependency scores are as reported from the CRISPR (Avana) 19Q3 dataset.

Online Methods

Additional information may be found in the Life Sciences Reporting Summary.

Cell lines and culture conditions. GM12878 cells (Coriell) were cultured in RPMI-1640 supplemented with L-glutamine (Gibco), 15% heat-inactivated fetal bovine serum (FBS) (Gibco), and 1% penicillin/streptomycin (Thermo Fisher Scientific). HT1080 (American Type Culture Collection, ATCC) were cultured in EMEM (ATCC) supplemented with 10% FBS and 1% penicillin/streptomycin. MDA-MB-231 (ATCC) and HEK293T (ATCC) cells were maintained in DMEM (Gibco) supplemented with 10% FBS, 1% L-glutamine, and 1% penicillin/streptomycin. DU145 (ATCC) and K562 (ATCC) cells were cultured in RPMI-1640 supplemented with 10% FBS and 1% penicillin/streptomycin. NB4 cells (DSMZ) were cultured in RPMI-1640 supplemented with 10% charcoal-stripped FBS (Gibco) and 1% penicillin/streptomycin. All-*trans* retinoic acid (ATRA, Sigma) was dissolved in DMSO at a concentration of 10 mM. Cells were incubated at 37 °C and 5% CO_2. Genomic DNA was extracted from K562 and NB4 cells using the Quick-DNA MiniPrep kit (Zymo).

Cloning and purification of recombinant proteins. Expression plasmids were acquired (pTXB1-Tn5, Addgene #60240) or cloned (APEX2 ORF from pTRC-APEX2, Addgene #72558) into the pTXB1 vector (NEB). Fusion constructs with different peptide linkers[67] were generated by site-directed mutagenesis (NEB). Plasmids containing C-terminally tagged gene constructs as described in this study are deposited to Addgene (#160081, #160083-160088). All enzymes were expressed and purified similarly as previously described[68]. In brief, plasmids were transformed into the Rosetta2 E. coli strain (EMD Millipore) and streaked out on an LB agar plate containing ampicillin and chloramphenicol. A single bacterial colony was inoculated into 10 mL LB with antibiotics and incubated overnight; this culture was then inoculated into 500 mL LB medium. Cultures were incubated at 37 °C until the optical density at 600 nm (OD600) reached ~0.9. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 250 μM, cultures were incubated for 2 h at 30 °C, and bacteria were pelleted and frozen at -80 °C.

Bacterial pellets were resuspended in 40 mL HEGX lysis buffer (20 mM HEPES-KOH pH 7.2, 1 M NaCl, 1 mM EDTA, 10% glycerol, 0.2% Triton X-100, 20 μM PMSF) and sonicated with a Sonic Dismembrator 100 (Fisher Scientific) at setting 7, with 5 pulses of 30 s on/off on ice. Lysate was spun at 15,000 x g in a Beckman centrifuge (JA-10 rotor) for 30 min at 4 °C. 1 mL 10% PEI was then added to the supernatant with agitation and clarified by centrifugation (15,000 x g, 15 min, 4 °C). Supernatant was then applied to 5 mL chitin resin (NEB) prewashed

with HEGX buffer and incubated for 1 h at 4 °C with agitation. Chitin slurry was applied to an Econo-Pak column (Bio-Rad) to remove unbound protein, washed with 20 column volumes of HEGX buffer and 1 column volume of HEGX with 50 mM DTT, and then incubated with 1 column volume of HEGX with 50 mM DTT for 48 h at 4 °C. After elution, the column was washed with 1 column volume of 2x dialysis buffer (2xDB: 100 mM HEPES-KOH pH 7.2, 0.2 M NaCl, 0.2 mM EDTA, 20% glycerol, 0.2% Triton X-100, 2 mM DTT). Eluates were combined, concentrated with a 10 kDa MWCO centrifugal filter (EMD Millipore), and subjected to buffer exchange with 2xDB using PD-10 desalting columns (GE Healthcare). Proteins were quantified via detergent-compatible Bradford assay (Thermo Fisher Scientific), snap frozen with liquid nitrogen, and stored at -80 °C.

Transposome adaptor preparation. All transposome adaptors were synthesized at Thermo Fisher Scientific. The oligonucleotide sequences were similar as previously described[18,68]: Tn5MErev, 5'-[phos]CTGTCTCTTATACACATCT-3'; Tn5ME-A, 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'; Tn5ME-B: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'; Tn5ME-A-AF647, 5'-/AlexaFluor647/TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'; Tn5ME-B-AF647: 5'-/AlexaFluor647/GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'. All oligos were resuspended in water to a final concentration of 200 µM each. Equimolar amounts of Tn5MErev/Tn5ME-A, Tn5MErev/Tn5ME-B, Tn5MErev/Tn5ME-A-AF647, and Tn5MErev/Tn5ME-B-AF647 were added together in separate tubes, denatured at 95 °C for 10 min, and cooled slowly to room temperature by removing the heat block. Tn5MEDS-A/Tn5MEDS-B and Tn5MEDS-A-AF647/Tn5MEDS-B-AF647 were combined at equimolar amounts to form 100 µM stocks of Tn5MEDS-A/B and Tn5MEDS-A/B-AF647, aliquoted, and stored at -20 °C.

Electrophoretic mobility shift assay and DNA fragmentation analysis. pSMART HCAmp plasmid (Lucigen) was linearized with EcoRV-HF (NEB) and column-purified. DNA:protein complexes were assembled by incubating 12 pmol enzyme in 2xDB buffer with 15 pmol MEDS-A/B in water for 1 h at room temperature. 200 ng of linearized plasmid was then added to the enzyme mix and brought to a final volume of 20 µL containing 20% dimethylformamide, 20 mM Tris-HCl pH 7.5, and 10 mM $MgCl_2$, with or without 50 mM EDTA. Tagmentation reactions were then incubated for 30 min at 37 °C. For gel shift analysis, reactions were subjected to electrophoresis on a 1% agarose gel in Tris-acetate-EDTA (TAE) buffer using gel loading dye

without SDS (NEB). DNA fragmentation was assessed by adding SDS to a final concentration of 0.2% to the reaction mix after tagmentation and heating at 55 °C for 15 min. Reactions were then subjected to electrophoresis on a 1% agarose gel cast in TAE and ethidium bromide using gel loading dye with SDS (NEB). Images were acquired via a Gel Doc (Bio-Rad) via the Quantity One v4.2.1 software.

ATAC-seq/iDAPT-seq sample preparation. The OmniATAC sample preparation protocol was used as previously described with modifications where indicated below[19]. 10 pmol enzyme (2 µL in 2xDB) was mixed with 12.5 pmol MEDS-A/B (1.25 µL in water) and incubated at room temperature for 1 h. In the meantime, 50,000 cells were centrifuged at 500 x g for 5 min at 4°C. Cells were resuspended in 50 µL lysis buffer 1 (LB1: 10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl2, 0.01% digitonin, 0.1% Tween-20, and 0.1% NP-40) with trituration, incubated on ice for 3 min, and then further supplemented with 1 mL lysis buffer 2 (LB2: 10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl2, and 0.1% Tween-20). Nuclei were pelleted (500 x g, 10 min, 4 °C), resuspended with 50 µL tagmentation reaction mixture (20% dimethylformamide, 10 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5, 33% 1xPBS, 0.01% digitonin, 0.1% Tween-20, and either 10 pmol enzyme equivalent of enzyme:DNA complex or 2.5 µL Nextera Tn5 [Illumina, TDE1 from FC-121-1030] in 50 µL total volume), and incubated at 37 °C for 30 min with agitation on a thermomixer (1,000 rpm). For iDAPT-seq libraries generated from K562 or NB4 cells or genomic DNA, bovine serum albumin (BSA) was added at a final concentration of 1% to lysis (LB1 and LB2) and tagmentation buffers. Tagmentation with naked genomic DNA was performed using 50 ng genomic DNA as substrate. After tagmentation, DNA libraries were extracted with DNA Clean and Concentrator-5 (Zymo) and eluted with 21 µL water.

To determine optimal PCR cycle number for library amplification, quantitative PCR was performed similarly as previously reported on a StepOnePlus Real-Time PCR (Applied Biosystems) with the StepOne v2.3 software[8]. 2 µL of each ATAC-seq or iDAPT-seq library was added to 2x NEBNext Master Mix (NEB) and 0.4x SYBR Green (Thermo Fisher) with 1.25 µM of each primer (Primer 1: 5'-AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG-3'; Primer 2.1: 5'-CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCT CGTGGGCTCGGAGATGT-3') in a final volume of 15 µL, and quantification was assessed using the following conditions: 72 °C for 5 min; 98 °C for 30 s; and thermocycling at 98 °C for 10 s, 63 °C for 30 s and 72 °C for 1 min. Optimal PCR cycle number was determined as the qPCR cycle yielding fluorescence between 1/4 and 1/3 of the maximum fluorescence. The remaining DNA library was then amplified

accordingly by PCR using previously reported barcoded primers for library multiplexing[8], purified with DNA Clean and Concentrator-5 (Zymo), and eluted into 20 µL final volume with water. Libraries were then subject to TapeStation 2200 High Sensitivity D1000 or D5000 fragment size analysis (Agilent) and NextSeq 500 High Output paired-end sequencing (2x75 bp, Illumina) as indicated.

ATAC-seq/iDAPT-seq data preprocessing. Paired-end sequencing reads were trimmed with TrimGalore v0.4.5 to remove adaptor sequence CTGTCTCTTATACACATCT, which arises at the 3' end due to sequenced DNA fragments being shorter than the sequencing length (75 bp). Reads were aligned to the hg38 reference genome using bowtie2 v2.2.9 with options "--no-unal --no-discordant --no-mixed -X 2000". Reads mapping to the mitochondrial genome were subsequently removed, and duplicate reads were removed with Picard v2.8.0. For insert size distribution, transcription start site (TSS) enrichment, and genome track visualization analyses, reads were downsampled to approximately 5 million paired-end fragments. Insert size distributions were determined by counting inferred fragment sizes from read alignments. TSS enrichment was performed by first shifting insert positions aligned to the reverse strand by -5 bp and the forward strand by +4 bp as previously described[8] and then determining the distance of each insertion to the closest Ensembl v94 transcription start site with Homer v4.9. Visualization was performed by mapping insertions to a genome-wide sliding 150 bp window with 20 bp offsets with bedops v2.4.30, followed by conversion to bigwig format with wigToBigWig from UCSC tools v363. Genome tracks were visualized with Integrative Genomics Viewer v2.5.0.

Peaks were aligned by MACS2 v2.1.1 using options "callpeak --nomodel --shift -100 --extsize 200 --nolambda -q 0.01 --keep-dup all", generating either individual peak sets from each library (GM12878 analysis) or a consensus peak set after consolidating all reads (K562, NB4 analyses). For GM12878 analysis, a union of all analyzed peaks was taken as a consensus peak set, and counts of insertions within peaks (downsampled to 5 million reads) were assessed using bedtools v2.26.0 with the multicov function. Correlation analysis was performed with log2 read counts + 1 and visualized using the pheatmap function in R v3.5.0. For K562 and NB4 analyses, consensus peaks overlapping with hg38 blacklist regions were removed (https://www.encodeproject.org/annotations/ENCSR636HFF/), and counts of insertions within peaks were assessed using the bedtools multicov function. Count matrices were processed with DESeq2 for differential insertions with shrunken log2 fold changes, and principal component analyses were performed with counts transformed by the varianceStabilizingTransformation function from DESeq2. Figures were generated with ggplot2 v3.1.1.

Co-immunofluorescence/ATAC-see analysis. ATAC-see was performed similarly as previously described with slight modifications[18]. Enzyme and transposon DNA were mixed at a 1:1.25 enzyme:MEDS-A/B-AF647 molar ratio and incubated at room temperature for 1 h. Adherent cells were grown on glass coverslips (Fisher Scientific, 12-540A) until 80-90% confluent, washed with 1xPBS, fixed with 1% formaldehyde (Electron Microscopy Services) in 1xPBS for 10 min, and washed twice with ice-cold 1xPBS. Immobilized cells were lysed by incubation with LB1 for 3 min followed by LB2 for 10 min at room temperature. Cells were then subject to tagmentation (20% dimethylformamide, 10 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5, 33% 1xPBS, 0.01% digitonin, 0.1% Tween-20, and 80 pmol enzyme equivalent of enzyme:DNA complex in a total volume of 100 µL) for 30 min at 37 °C in a humidified chamber. Subsequently, cells were washed with 50 mM EDTA and 0.01% SDS in 1xPBS three times for 15 min each at 55 °C, lysed for 10 min with 0.5% Triton X-100 in 1xPBS at room temperature, and blocked with 1% BSA and 10% goat serum in PBS-T (1xPBS and 0.1% Tween-20) for 1 h in a humidified chamber. Primary antibody was added to slides in 1% BSA/PBS-T and incubated at 4 °C overnight; slides were then washed and subjected to secondary antibody staining for 1 h. Slides were washed with PBS-T three times for 15 min each, stained with DAPI (Sigma, 1 µg/mL) for 1 min, washed with PBS for 10 min, and mounted with Fluorescence Mounting Medium (Dako). Confocal microscopy images were taken with an LSM 880 Axio Imager 2 or an LSM 880 Axio Observer at 63x magnification (Zeiss). Images were processed with Fiji/ImageJ v2.0.0.

Primary antibodies used were anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) (rabbit, Abcam ab5095, 1:500), anti-H3K27Ac (rabbit, Abcam ab4729, 1:500), anti-H3K9me3 (rabbit, Abcam ab8898, 1:500), anti-SC35 (mouse, SC-35, Abcam ab11826, 1:1000). Secondary antibodies used were Goat anti-Rabbit IgG (H+L) Secondary Antibody, Alexa Fluor 488 conjugate (Thermo Fisher Scientific A11008, 1:1000) and Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 conjugate (Thermo Fisher Scientific A11001, 1:1000).

Quantitative image analyses were performed with CellProfiler v3.1.5. Region of interests (ROIs) were identified from DAPI channel intensity values using minimum cross entropy thresholding, with each ROI corresponding to an individual nucleus. Pearson correlation coefficients were determined by comparing ATAC-see pixel intensities with corresponding immunofluorescence intensity values within each ROI to assess the nucleus-to-nucleus variation in colocalization.

Peroxidase activity assay. 5 pmol enzyme was incubated with 2.5 pmol hemin chloride (Cayman Chemical, dissolved in DMSO) for 1 h at room temperature. This molar ratio was selected given reports of APEX2 maximal heme occupancy between 40-57%. Heme:protein complexes were then subjected to 50 µM Amplex UltraRed (Thermo Fisher Scientific) and 1 mM hydrogen peroxide for 1 min at room temperature in a total volume of 100 µL with 1xPBS. Reactions were then quenched with 100 µL 2x quenching solution (10 mM Trolox, 20 mM sodium ascorbate, and 20 mM NaN_3 in 1xPBS), and fluorescence intensities were measured on a SpectraMax iD3 plate reader with the SoftMax Pro v7.0.3 software, with excitation at 530 nm and emission at 590 nm.

DNA and protein tagging by iDAPT. All iDAPT proteomic labeling assays were performed as described below unless indicated otherwise. 2.5 µmol MEDS-A/B, 2 µmol enzyme, and 1 µmol hemin chloride per channel were incubated at room temperature for 1 h. 1e7 cells per sample were washed (500 x g, 5 min, 4 °C), lysed and triturated in 100 µL LB1 (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl2, 1% BSA, 0.01% digitonin, 0.1% Tween-20, 0.1% NP-40, and 1x cOmplete EDTA-free protease inhibitor cocktail [Roche]) for 3 min, and subsequently supplemented with an additional 1 mL of LB2 (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl2, 1% BSA, 0.1% Tween-20, and 1x protease inhibitor). Nuclei were pelleted (500 x g, 10 min, 4 °C), resuspended with tagmentation reaction mixture (20% dimethylformamide, 10 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5, 33% 1xPBS, 1% BSA, 0.01% digitonin, 0.1% Tween-20, 500 µM biotin-phenol, 1x protease inhibitor, and 2 µmol enzyme equivalent of enzyme:DNA:heme complex in a total volume of 500 µL), and incubated at 37 °C for 30 min with agitation on a thermomixer (1,000 rpm). 5 µL of tagmentation mix was saved for quality assessment as described above for ATAC-seq/iDAPT-seq sample preparation. The remaining nuclear suspension was then washed 2x with 1xPBS supplemented with 500 µM biotin-phenol, 1% BSA, 0.1% Tween-20, and 1x protease inhibitor (3000 x g, 5 min, 4 °C) and labeled with 1 mM hydrogen peroxide and 500 µM biotin-phenol for 1 min in 1xPBS with 1x protease inhibitor in a volume of 500 µL. Peroxidation reactions were quenched with 500 µL 2x quenching buffer (10 mM Trolox, 20 mM sodium ascorbate, 20 mM NaN_3, and 1x protease inhibitor in 1xPBS). Labeled nuclei were then pelleted, washed with 1x quenching buffer, resuspended in 500 µL RIPA containing protease inhibitors, and frozen at -80 °C. Lysates were thawed on ice, sonicated via a Sonic Dismembrator 100 (Fisher Scientific, setting 3, 15 s, 4 pulses), and incubated on ice for 30 min after the addition of 1 µL benzonase (EMD Millipore). Lysates were clarified by centrifugation (15,000 x g, 20 min, 4 °C), quantified via the detergent-compatible

Bradford assay (Thermo Fisher Scientific), and subjected to either Western blotting or quantitative mass spectrometry analyses as described below. For NB4 cell analysis, an additional endogenous peroxidase blocking step was added after nuclear extraction and before tagmentation: nuclei were resuspended in 500 µL 1xPBS containing 1% BSA, 0.03% hydrogen peroxide, and 0.1% NaN_3 and incubated on ice for 30 min. Nuclei were pelleted and washed 4x with 1xPBS/1% BSA (3000 x g, 5 min, 4 °C). Residual hydrogen peroxide was monitored by colorimetric assessment of supernatant via Quantofix peroxides test stick (Sigma).

Western blotting analysis. Whole cell or nuclear lysates were generated by resuspending cells or nuclei in RIPA (Boston BioProducts) supplemented with 1x cOmplete EDTA-free protease inhibitor cocktail (Roche). Lysates were incubated on ice for 30 min, sonicated via a Sonic Dismembrator 100 (Fisher Scientific) at setting 3 with 3-4 pulses of 15 s on/off on ice, and treated with benzonase for an additional 30 min on ice. Lysates were clarified by centrifugation (15,000 x g, 20 min, 4 °C) and their concentrations quantified via the detergent-compatible Bradford assay (Thermo Fisher Scientific). All Western blots were run on NuPAGE 4-12% Bis-Tris protein gels (Thermo Fisher Scientific) and transferred to 0.2 µm nitrocellulose membranes (GE Healthcare). Membranes were blocked with 3% milk in PBS-T and incubated overnight with primary antibody and subsequently with secondary antibody after brief washing with PBS-T. Chemiluminescence was determined by applying ECL Western Blotting detection reagent (GE Healthcare) to membranes and imaging on an Amersham Imager 600 (GE Healthcare). Membranes were stripped with Restore PLUS Stripping Buffer (Thermo Fisher Scientific).

Primary antibodies used were anti-FLAG M2 (mouse, Sigma-Aldrich, F1804, 1:2000), anti-PCNA (mouse, PC10, Santa Cruz Biotechnology sc-56, 1:1000), and anti-PML (rabbit, Bethyl A301-167A, 1:1000). Secondary antibodies used were Rabbit IgG, HRP-linked $F(ab')_2$ fragment (GE Healthcare NA9340, from donkey, 1:5000) and Mouse IgG, HRP-linked whole Ab (GE Healthcare NA931, from sheep, 1:5000). Streptavidin-HRP (Cell Signaling Technology #3999S, 1:1000) was also used for probing.

Streptavidin enrichment and tandem mass tag labeling. 250 µg (K562) or 150 µg (NB4) lysate was reduced with 5 mM DTT and then added to 60 µL (K562) or 90 µL (NB4) Pierce streptavidin bead slurry equilibrated 2x with RIPA buffer. Lysate/bead mixture was incubated with end-to-end rotation overnight at 4 °C. Beads were washed 3x with RIPA, 2x with 200 mM EPPS pH 8.5, and resuspended with 100 µL 200 mM EPPS pH 8.5, with beads resuspended and incubated with end-to-end rotation for 5 min per wash. 1 µL mass spectrometry-grade LysC

(Wako) was added to each tube and incubated at 37 °C for 3 h with mixing, and an additional 1 μL mass spectrometry-grade trypsin (Thermo Fisher Scientific) was added, followed by overnight incubation at 37 °C with mixing. Beads were magnetized, and eluate was collected and subjected to downstream TMT labeling.

Peptides were processed using the SL-TMT method[24]. TMT reagents (0.8 mg) were dissolved in anhydrous acetonitrile (40 μL), of which 10 μL was added to each peptide suspension (100 μL) with 30 μL of acetonitrile to achieve a final acetonitrile concentration of approximately 30% (v/v). Following incubation at room temperature for 1 h, the reaction was quenched with hydroxylamine to a final concentration of 0.3% (v/v). The TMT-labeled samples were pooled at a 1:1 ratio across all samples. The pooled sample was vacuum centrifuged to near dryness and subjected to C18 solid-phase extraction (SPE) (Sep-Pak, Waters).

Off-line basic pH reversed-phase (BPRP) fractionation. We fractionated the pooled TMT-labeled peptide sample using BPRP HPLC[69]. We used an Agilent 1200 pump equipped with a degasser and a photodiode array (PDA) detector (set at 220 and 280 nm wavelength) from ThermoFisher Scientific (Waltham, MA). Peptides were subjected to a 50-min linear gradient from 9% to 35% acetonitrile in 10 mM ammonium bicarbonate pH 8 at a flow rate 600 μL/min over an Agilent 300Extend C18 column (3.5 μm particles, 4.6 mm ID and 220 mm in length). The peptide mixture was fractionated into a total of 96 fractions, which were consolidated into 24 super-fractions[70]. Samples were subsequently acidified with 1% formic acid and vacuum centrifuged to near dryness. Each consolidated fraction was desalted via StageTip, dried again via vacuum centrifugation, and reconstituted in 5% acetonitrile, 5% formic acid for LC-MS/MS processing.

LC-MS/MS proteomic analysis. Samples were analyzed on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to a Proxeon EASY-nLC 1200 liquid chromatography (LC) pump (Thermo Fisher Scientific). Peptides were separated on a 100 μm inner diameter microcapillary column packed with 35 cm of Accucore C18 resin (2.6 μm, 150 Å, ThermoFisher). For each analysis, approximately 2 μg of peptides were separated using a 150 min gradient of 8 to 28% acetonitrile in 0.125% formic acid at a flow rate of 450-500 nL/min. Each analysis used an MS3-based TMT method[71,72], which has been shown to reduce ion interference compared to MS2 quantification[73]. The scan sequence began with an MS1 spectrum (Orbitrap analysis, resolution 120,000, 350−1400 Th, automatic gain control (AGC) target 2e5, maximum injection time 100 ms). The top ten precursors were then selected for

MS2/MS3 analysis. MS2 analysis consisted of: collision-induced dissociation (CID), quadrupole ion trap analysis, automatic gain control (AGC) 1.4e4, NCE (normalized collision energy) 35, q-value 0.25, maximum injection time 120 ms), and isolation window at 0.7. Following acquisition of each MS2 spectrum, we collected an MS3 spectrum in which multiple MS2 fragment ions are captured in the MS3 precursor population using isolation waveforms with multiple frequency notches. MS3 precursors were fragmented by HCD and analyzed using the Orbitrap (NCE 65, AGC 1.5e5, maximum injection time 150 ms, resolution was 50,000 at 400 Th).

Proteomic data analysis. Mass spectra were processed using a Sequest-based pipeline[74]. Spectra were converted to mzXML using a modified version of MSConvert. Database searching included all entries from the human UniProt database. This database was concatenated with one composed of all protein sequences in the reversed order. Searches were performed using a 50-ppm precursor ion tolerance for total protein level analysis. The product ion tolerance was set to 0.9 Da. TMT tags on lysine residues and peptide N-termini (+229.163 Da) and carbamidomethylation of cysteine residues (+57.021 Da) were set as static modifications, while oxidation of methionine residues (+15.995 Da) was set as a variable modification.

Peptide-spectrum matches (PSMs) were adjusted to a 1% false discovery rate (FDR)[75,76]. PSM filtering was performed using a linear discriminant analysis (LDA), as described previously[74], while considering the following parameters: XCorr, ΔCn, missed cleavages, peptide length, charge state, and precursor mass accuracy. For TMT-based reporter ion quantitation, we extracted the summed signal-to-noise (S:N) ratio for each TMT channel and found the closest matching centroid to the expected mass of the TMT reporter ion. PSMs with poor quality, MS3 spectra with more than eight TMT reporter ion channels missing, MS3 spectra with TMT reporter summed signal-to-noise of less than 100, missing MS3 spectra, or isolation specificity < 0.7 were excluded from quantification[77].

PSM intensities were normalized by taking the median intensity of streptavidin and trypsin PSMs per sample as a normalization factor, as these proteins are added to each sample in equal amounts post-enrichment. Normalized PSMs were then log2-transformed and collapsed to proteins by arithmetic average, with priority given to uniquely mapping peptides. Hierarchical clustering, Pearson correlation, and principal component analyses were performed at the protein level. The limma package in R was used to determine differential protein abundances.

Protein enrichment analyses. Gene set enrichment analyses of iDAPT-MS datasets were performed with the fgsea package (10,000 permutations) in R, using UniProt protein

identifications ranked by their log2 fold changes from limma[78]. Gene sets used for analyses: CORUM (v3.0) protein complex annotations[32], Human Protein Atlas (v19) subcellular localization annotations with reliability demarcated as "Enhanced" or "Supported"[27], BioGrid (v3.5.178) multi-validated protein interaction annotations[35], ReactomeDB (v70) pathway to gene mappings from fgsea via the "reactomePathways" function[79], and CisBP transcription factors from the "human_pwms_v2" dataset curated as in the chromVARmotifs package in R[36,47]. All gene identities were converted to UniProt prior to analysis via biomaRt in R. Protein interaction networks were visualized with igraph v1.2.4.

Four classes of nuclear proteins were collated: histones, chromatin remodelers, transcription factors, and RNA-binding proteins. Histone UniProt IDs were collated from Histone DB 2.0[80] and UniProt with search query "Nucleosome core"[81]. Chromatin remodeler proteins were obtained from UniProt IDs associated with "GO:0006338" ("chromatin remodeling")[82] and CORUM protein complex components associated with the five primary chromatin remodelers[32]: NuRD, SWI, ISWI, INO80, SWR1. High-confidence RNA binding proteins were obtained from hRBPome[83], and transcription factors were obtained from Lambert et al[3].

K562 RNA-seq[25] (ENCFF664LYH and ENCFF855OAF), whole cell proteome[33], and nuclear proteome[34] datasets were downloaded and converted to UniProt IDs. RNA-seq genes were filtered for those with nonzero read counts (transcripts per million) in both replicates[25]. The whole cell proteomic dataset was filtered by removing peptides with missing quantitations[33]. The nuclear proteome dataset was preprocessed by removing peptides with multiple UniProt IDs and collating remaining UniProt IDs across all salt extraction conditions[34]. For determination of proteins associated with specific extraction conditions, we followed a procedure as reported by Federation et al.: peptide intensities were normalized by total intensities for a given sample, collapsed to protein intensities by arithmetic mean, scaled to maximum intensities of 1, and subjected to k-means clustering analysis using $k = 8$ for clustering[34]. Protein annotations from Alajem et al. were converted from mouse to human homologs via biomaRt in R, and gene sets (1000U, 45U, 3U) were compiled taking the sets of protein IDs with scores greater than 95 in either ES or NPC sample types[13]. Additional publicly available open chromatin proteome datasets were downloaded, and gene identities were converted to UniProt IDs[12,14]. Because published datasets differ in their analytical depths from our iDAPT-MS datasets, we converted gene identifiers to Human Protein Atlas subcellular enrichment proportions for better comparison. Specifically, the proportion for each subcellular localization term and for each dataset was calculated as the (*number of proteins overlapping between the subcellular term and the dataset*) / (*number of proteins overlapping between all annotated Human Protein Atlas*

proteins and the dataset). These proportions were used as features for principal component analysis.

CUT&RUN sample preparation. pAG/MNase (Addgene #123461) was expressed in Rosetta2 cells (EMD Millipore), purified with the Pierce His Protein Interaction Pull-Down kit (Thermo), and stored at either -80 °C for long-term storage or -20 °C for working stocks[84]. CUT&RUN was performed similarly as previously reported[26]. 500,000 K562 cells per assay were washed three times (room temperature, 3 min, 600 x g) in wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, and 1x cOmplete EDTA-free protease inhibitor cocktail [Roche]). Concavalin A beads were activated by washing beads in binding buffer (20 µM HEPES pH 7.5, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$). 10 µL activated Concavalin A beads were added to 100 µL cell suspension and incubated with rotation for 10 min at room temperature. Supernatant was removed, and 100 µL wash buffer containing 0.01% digitonin (dig-wash buffer) was added. Antibodies were added at 1:50 concentration, and tubes were incubated with rotation overnight at 4 °C. Beads were washed with dig-wash buffer, pAG/MNase was added at a final concentration of 2 µg/mL, and suspensions were incubated for 1 h at 4 °C. Beads were further washed with wash buffer, resuspended in 100 µL wash buffer, and chilled to 0 °C in an ice-water bath. 2 µL 0.1 M $CaCl_2$ was added to each tube, and tubes were incubated for 1 h at 0 °C. 100 µL stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.05% digitonin, 100 µg/mL RNase A, 50 µg/mL GlycoBlue) was added, and tubes were incubated for 15 min 37 °C to release DNA fragments. Supernatant was collected, SDS (0.1% final) and proteinase K (250 µg/mL final) were added to each 200 µL sample, and tubes were incubated for 1 h at 50 °C. DNA was isolated by phenol/chloroform extraction, and libraries were constructed using the NEBNext Ultra kit (NEB) as previously described[52]. Libraries were then subject to TapeStation 2200 High Sensitivity D1000 fragment size analysis (Agilent) and NextSeq 500 High Output paired-end sequencing (2x42 bp, Illumina). Primary antibodies used for CUT&RUN were: ERH (Bethyl, A305-402A; 1:50), WBP11 (Bethyl, A304-855A; 1:50), and normal rabbit IgG (EMD Millipore, #12-370; 1:50).

Antibodies used for CUT&RUN were validated by immunoprecipitation followed by Western blotting analysis. K562 cells were lysed in RIPA, and 1.5 µL antibody was added to 500 µg protein lysate and incubated overnight at 4 °C. The next day, lysates were incubated with 20 µL Pierce protein A magnetic beads (Thermo) for 2 h at 4 °C, beads were washed in RIPA buffer, and bound protein was boiled in 2x LDS sample buffer for 10 min. Resulting protein lysates were subjected to Western blotting analysis as described above. Primary antibodies used for

Western blotting were: ERH (Atlas Antibodies, HPA002567; 1:1,000) and WBP11 (Bethyl, A304-857A; 1:1,000).

CUT&RUN analysis. Paired-end sequencing reads were trimmed with TrimGalore v0.4.5 to remove adaptor sequence GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT with additional removal of fragments smaller than 25 bp. Reads were aligned to the hg38 reference genome using bowtie2 v2.2.9 with options "--no-unal --no-discordant --no-mixed --dovetail -I 25 -X 700". Reads mapping to the mitochondrial genome were subsequently removed, and duplicate reads were removed with Picard v2.8.0. Reads smaller than 120 bp were retained for subsequent analysis. Visualization was performed by mapping insertions to a genome-wide sliding 150 bp window with 20 bp offsets with bedops v2.4.30, followed by conversion to bigwig format with wigToBigWig from UCSC tools v363. Genome tracks were visualized with Integrative Genomics Viewer v2.5.0. Open chromatin regions were defined as 1% FDR-thresholded MACS2 peaks obtained from K562 iDAPT-seq relative to genomic DNA input as described above. CUT&RUN signal was determined relative to these peak regions and normalized by the signal intensity between +1950 and +2000 bp distal to the peak summit, representing background enrichment. CUT&RUN peaks were called by MACS2 v2.1.1 using options "callpeak -q 0.01 --keep-dup all". CUT&RUN and ChIP-seq peak overlap analyses were performed with bedtools v2.26.0 using the intersect function.

ATAC-seq/iDAPT-seq transcription factor analysis. Motif enrichment analysis was performed with ChromVAR as previously described using the human_pwms_v2 set of curated CisBP transcription factor motifs[36,47]. ChromVAR motif deviations from the computeDeviations function were used for principal component analysis, and FDR-adjusted p-values were obtained with the differentialDeviations function with default settings.

Bivariate footprinting analysis was performed similarly as previously described with slight modifications[10,85]. CisBP motifs curated from the ChromVAR human_pwms_v2 dataset[36,47] or motifs for ZEB2[86] and EBF3[87] were matched within peaks using matchMotifs from motifmatchr in R. Motif alignments were extended by 250 bp on each side, and adjusted transposon insertions were mapped to the corresponding regions. Motif flank height was determined by the average insertion rate between positions +1 to +50 bp, immediately flanking the motif. Background insertions were determined by the average insertion rate between positions +200 to +250 bp, distal to the positioned motif. Footprint height was determined by the 10% trimmed mean of the insertion rate within the 10-11 bp positioned around the center of the motif. Footprint depth

(FPD) was determined as the log2 count ratio of footprint height over flank height; flanking accessibility (FA) was determined as the log2 count ratio of flank height over background. The norm of the orthogonal projection of FA and FPD scores onto the -45° line was used as a raw footprinting score. A linear regression model was implemented (*footprinting score ~ transcription factor + transcription factor:treatment*), from which the t-statistic of the interaction term per transcription factor motif (*transcription factor:treatment*) was used as the composite footprinting score, and the corresponding p-value, adjusted to false discovery rate with the Benjamini-Hochberg method, was used to assess significance.

For analysis of transcription factor activity at steady-state, composite footprinting scores were modeled by a two-state Gaussian mixture model with mixtools in R, and class A footprinted motifs (strong footprinting) were determined to be those with greater than 50% probability of being in the Gaussian distribution further away from the origin. Class C footprinted motifs (no/negative footprinting) were determined as those with weak statistical significance (FDR > 5%) or negative enrichment (composite footprinting score < 0). Positive and significant footprinted motifs not in class A were demarcated as class B footprinted motifs (weak footprinting). Consensus transcription factor classifications were determined by concordance between K562 and NB4 steady-state footprinting analyses, limited to those transcription factors exhibiting positive significant enrichment from both corresponding iDAPT-MS datasets.

For classification of transcription factors upon ATRA treatment, FDR < 5% thresholds of iDAPT-MS abundance and iDAPT-seq footprinting profiles were used to discriminate between classes.

ChIP-seq analysis. ENCODE ChIP-seq transcription factor datasets were downloaded from the ENCODE data portal[25] (https://www.encodeproject.org/). ENCODE K562 ChIP-seq datasets are listed in Supplementary Table 3. In brief, ChIP-seq bed files aligned to hg38 and annotated as "optimal IDR peaks" were downloaded, and iDAPT-seq peaks overlapping with ChIP-seq peaks were collated. ChIP-seq enrichment within open chromatin was determined by gene set enrichment analysis using iDAPT-seq differential peaks ranked by log2 fold change using the fgsea package in R.

Colocalization of ChIP-seq epitopes on open chromatin was determined using the Jaccard similarity coefficient, with colocalization determined if ChIP-seq peaks from different epitopes overlap a given iDAPT-seq peak.

Granulocytic differentiation analysis. NB4 cells treated either with DMSO or 1 μM ATRA were washed with 2% fetal bovine serum prior to staining. Anti-human CD11b-PE-Cy7 antibody conjugate (Clone: ICRF44, Biolegend Catalog #301321; 1:100) and anti-human CD11c-APC antibody conjugate (Clone: B-ly6, BD Pharmingen #559877; 1:100) were incubated with samples for 20 min and then washed to remove excess antibody. Stained samples were analyzed on a Beckman Coulter CytoFLEX LX flow cytometer with the CytoExpert v2.3.1.22 software. Data were analyzed with FlowJo v10.0.7.

Cell proliferation assay. NB4 cells were seeded at a density of 5e5 cells/mL subjected to either DMSO or 1 μM ATRA. After 48 h, 50 μL cell suspension was added to 50 μL CellTiter-Glo reagent, incubated for 10 min at room temperature, and assayed for luminescence with a SpectraMax iD3 plate reader.

Genetic dependency analysis. Genetic dependency map (DepMap) scores generated from CRISPR/Cas9 pooled screening (Avana) were downloaded (19Q3, https://depmap.org/portal/). DepMap scores from hematopoietic cancer cell lines were collated, and the distribution of dependency scores was modeled as a two-state Gaussian mixture model with mixtools in R. Gene dependency was determined as the threshold corresponding to 50% probability of being in either distribution. Essential genes across hematopoietic cell lines were those genes representing dependencies across at least 50% of profiled hematopoietic cell lines.

RNA-seq analysis. Raw sequencing reads (GSM1288651, GSM1288652, GSM1288653, GSM1288654, GSM1288659, GSM1288660, GSM1288661, GSM1288662, GSM2464389, GSM2464392) were aligned to a reference transcriptome generated from the Ensembl v94 database with salmon v0.14.1 using options "--seqBias --useVBOpt --gcBias --posBias --numBootstraps 30 --validateMappings". Length-scaled transcripts per million were acquired using the tximport function, and log2 fold changes and false discovery rates were determined by DESeq2 in R, with batch as a covariate. Principal component analysis was performed with counts transformed by the varianceStabilizingTransformation function from DESeq2, and shrunken log2 fold changes were determined with DESeq2, which were used to rank genes for gene set enrichment analysis. For comparison of RNA-seq and mass spectrometry datasets, gene symbols and Ensembl gene IDs were matched to UniProt IDs via biomaRt.

Statistical analysis. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. All statistical analyses were performed in R[88]. Two-tailed statistical tests were used unless stated otherwise. Multiple comparison adjustments were performed as noted.

Data availability. iDAPT-seq/ATAC-seq and CUT&RUN datasets are deposited in GEO (GSE158350). iDAPT-MS proteomics data are deposited to the ProteomeXchange Consortium via the PRIDE partner repository (PXD022252). Raw confocal image files (.czi) are deposited to the Dryad repository at https://doi.org/10.5061/dryad.4xgxd257p.

Raw iDAPT-seq/ATAC-seq sequencing data (GSE158350, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158350) are associated with the following figures: **Fig 1b-c** and **Extended Data Fig 2** (GM12878 ATAC-seq, iDAPT-seq); **Fig 2g-h, Fig 3,** and **Extended Data Figs 5, 7-8** (K562 iDAPT-seq); **Fig 4g, Extended Data Figs 7-8,** and **Supplementary Figs 5-9** (NB4 iDAPT-seq). Raw CUT&RUN sequencing data (GSE158350, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158350) are associated with the following figures: **Fig. 2c** and **Extended Data Fig 5**. Raw mass spectrometry data (PXD022252, https://www.ebi.ac.uk/pride/archive/projects/PXD022252) are associated with the following figures: **Fig 2, Fig 3, Extended Data Figs. 3, 6, 8,** and **Supplementary Figs 3-4** (K562 iDAPT-MS); **Fig 4, Extended Data Figs. 4, 6, 8-10,** and **Supplementary Figs. 4, 6-10** (NB4 iDAPT-MS). Preprocessed mass spectrometry data are available as supplementary tables (Supplementary Tables 1-2). Raw confocal microscopy image data (https://doi.org/10.5061/dryad.4xgxd257p) are associated with the following figures: **Fig 1d-e, 2d,** and **Extended Data 6d-e**.

Publicly available sequencing datasets used are as follows: GM12878 ATAC-seq: https://www.ncbi.nlm.nih.gov//geo/query/acc.cgi?acc=GSE47753 (SRR891268, SRR891269, SRR891270, SRR891271), https://www.ncbi.nlm.nih.gov/bioproject/PRJNA482539 (SRR7586167, SRR7586168), https://www.ncbi.nlm.nih.gov/bioproject/PRJNA305986 (SRR2999312, SRR2999313, SRR2999314, SRR2999315), https://www.ncbi.nlm.nih.gov/bioproject/PRJNA380283 (SRR5427884, SRR5427885, SRR5427886, SRR5427887); ENCODE K562 ChIP-seq: https://www.encodeproject.org/, with unique identifiers listed in Supplementary Table 3; ENCODE K562 RNA-seq: https://www.encodeproject.org/files/ENCFF664LYH/@@download/ENCFF664LYH.tsv and https://www.encodeproject.org/files/ENCFF855OAF/@@download/ENCFF855OAF.tsv; NB4 +/-

ATRA RNA-seq: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53258 (GSM1288651, GSM1288652, GSM1288653, GSM1288654), https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53259 (GSM1288659, GSM1288660, GSM1288661, GSM1288662), and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93877 (GSM2464389, GSM2464392).

Publicly available proteome datasets used are as follows: whole cell proteome: https://gygi.med.harvard.edu/sites/gygi.med.harvard.edu/files/documents/protein_quant_current_normalized.csv.gz; nuclear proteome and differential salt fractionation: https://ars.els-cdn.com/content/image/1-s2.0-S2211124720301303-mmc2.xlsx, Alajem et al.: https://www.cell.com/cms/10.1016/j.celrep.2015.02.064/attachment/daebc867-0c82-45ef-837b-b408682c76cf/mmc2.xlsx; Torrente et al.: https://doi.org/10.1371/journal.pone.0024747.s004 and https://doi.org/10.1371/journal.pone.0024747.s006; Kulej et al.: https://www.mcponline.org/highwire/filestream/35613/field_highwire_adjunct_files/5/TABLE_S5_Host_chromatin_bound_proteome.xlsx.

Additional public reference datasets are as follows: hg38 reference genome: ftp://ftp.ensembl.org/pub/release-94/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz; hg38 blacklist regions: https://www.encodeproject.org/files/ENCFF356LFX/@@download/ENCFF356LFX.bed.gz; CORUM v3.0 complexes: http://mips.helmholtz-muenchen.de/corum/download/allComplexes.txt.zip; Human Protein Atlas v19: https://www.proteinatlas.org/download/subcellular_location.tsv.zip; BioGrid v3.5.178: https://downloads.thebiogrid.org/File/BioGRID/Release-Archive/BIOGRID-3.5.178/BIOGRID-MV-Physical-3.5.178.tab2.zip; Lambert et al. transcription factors: https://www.cell.com/cms/10.1016/j.cell.2018.01.029/attachment/ede37821-fd6f-41b7-9a0e-9d5410855ae6/mmc2.xlsx; HistoneDB 2.0: https://www.ncbi.nlm.nih.gov/research/HistoneDB2.0/HistoneDB/static/browse/dumps/seqs.txt; hRBPome: http://caps.ncbs.res.in/hrbpome/downloads/high_confidence_proteins.fasta; DepMap 19Q3: https://ndownloader.figshare.com/files/16757666. CisBP transcription factors (http://cisbp.ccbr.utoronto.ca/) were obtained via the command data("human_pwms_v2") in R package "chromVARmotifs": https://github.com/GreenleafLab/chromVARmotifs. ReactomeDB v70 pathway annotations (https://reactome.org/) were obtained via the "reactomePathways" command in R package "fgsea": https://bioconductor.org/packages/release/bioc/html/fgsea.html. Gene Ontology (http://geneontology.org/) was queried from org.Hs.eg.db using the "select"

function from AnnotationDbi in R. UniProt IDs (https://www.uniprot.org/) were either downloaded from the UniProt website or collated via biomaRt in R (https://www.bioconductor.org/packages/release/bioc/html/biomaRt.html).

Code availability. R code used in this manuscript is deposited at https://github.com/jonathandlee12/iDAPT-MS.

Methods-Only References

67. Chen, X., Zaro, J. L. & Shen, W.-C. Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev.* **65**, 1357–69 (2013).
68. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
69. Wang, Y. *et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11**, 2019–26 (2011).
70. Paulo, J. A. *et al.* Quantitative mass spectrometry-based multiplexing compares the abundance of 5000 S. cerevisiae proteins across 10 carbon sources. *J. Proteomics* **148**, 85–93 (2016).
71. Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **8**, 937–940 (2011).
72. McAlister, G. C. *et al.* MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).
73. Paulo, J. A., O'Connell, J. D. & Gygi, S. P. A Triple Knockout (TKO) Proteomics Standard for Diagnosing Ion Interference in Isobaric Labeling Experiments. *J. Am. Soc. Mass Spectrom.* **27**, 1620–5 (2016).
74. Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
75. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol. Biol.* **604**, 55–71 (2010).
76. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–14 (2007).
77. McAlister, G. C. *et al.* Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal. Chem.* **84**, 7469–7478 (2012).
78. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
79. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
80. Draizen, E. J. *et al.* HistoneDB 2.0: a histone database with variants—an integrated resource to explore histones and their variants. *Database* **2016**, baw014 (2016).
81. The Uniprot Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids*

Res. **47**, D506–D515 (2019).
82. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
83. Ghosh, P., Murugavel, P. & Sowdhamini, R. hRBPome : a central repository of all known human RNA-binding proteins Keywords. (2018). doi:https://doi.org/10.1101/269043
84. Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *Elife* **8**, (2019).
85. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, (2018).
86. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–89 (2010).
87. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gkz1001
88. R Core Team. R: A language environment for statistical computing. (2014).