APOBEC2 binds Chromatin and Represses Transcription during Myoblast Differentiation

Jose Paulo Lorenzo[1*], Linda Molla[2*], Ignacio L. Ibarra[3,4], Sandra Ruf[1], Poorani Ganesh Subramani[5,6], Jonathan Boulais[5], Dewi Harjanto[2,7], Alin Vonica[8], Javier M. Di Noia[5,6,9], Christoph Dieterich[10], Judith B. Zaugg[3], F. Nina Papavasiliou[1,2]

[1] Division of Immune Diversity, German Cancer Research Center (DKFZ), Heidelberg, Germany

[2] Laboratory of Lymphocyte Biology, The Rockefeller University, New York, NY 10065, USA

[3] Structural and Computational Unit, European Molecular Biology Laboratory, Heidelberg, Germany

[4] Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[5] Institut de Recherches Cliniques de Montréal, 110 av. des Pins Ouest, Montréal, QC, Canada H2W 1R7

[6] Department of Medicine, Division of Experimental Medicine, McGill University, 1001 boul Decarie, Montréal, QC, Canada H4A 3J1

[7] BioNTech US, Cambrige, MA 02139, USA

[8] Department of Biology, The Nazareth College, Rochester, NY 14618, USA

[9] Department of Medicine, Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal, QC, Canada, H3C 3J7

[10] Klaus Tschira Institute for Integrative Computational Cardiology, University Hospital Heidelberg, Heidelberg, Germany

*shared authorship

ABSTRACT

APOBEC2 is a member of the prolific activation induced cytidine deaminase/ apolipoprotein B editing complex (AID/APOBEC) family of DNA or RNA editors. This family of nucleic acid editors has diverse molecular roles ranging from antibody diversification to RNA transcript editing. However, even though APOBEC2 is an evolutionarily conserved zinc-dependent cytidine deaminase, it neither has an established molecular substrate nor function. In this work, we use the C2C12 skeletal myoblast differentiation model to confirm that APOBEC2 is upregulated during differentiation and is critical to proper differentiation. Furthermore, we show that APOBEC2 has none of the attributed molecular functions of the AID/APOBEC family, such as mRNA editing, DNA demethylation, and DNA mutation. Unexpectedly, we reveal that APOBEC2 binds chromatin at promoter regions of actively transcribed genes, and binding correlates with transcriptional repression of non-myogenesis related gene pathways. APOBEC2 occupied regions co-occur with sequence motifs for several transcription factors such as Specificity Protein/Krüppel-like Factor (SP/KLF), and we demonstrate *in vitro* that APOBEC2 binds directly and co-operatively to double stranded DNA containing a SP1 binding site. Finally, protein-proximity data show that APOBEC2 directly interacts with histone deacetylase (HDAC) transcriptional co-repressor complexes.

Taken together, these data suggest a role for APOBEC2 as a transcriptional repressor for muscle differentiation, a novel role that is unique among AID/APOBEC family members.

INTRODUCTION

The AID/APOBEC proteins are zinc-dependent deaminases that catalyze the removal of the amino group from a cytidine base in the context of a polynucleotide chain, resulting in cytidine (C) to uridine (U) transition on DNA or RNA. Members of the AID/APOBEC family are closely related to one another based on homology and conservation of the cytidine deaminase domain containing a zinc-dependent deaminase sequence motif (Conticello, 2008). However, they differ by tissue-specific expression, substrates, and biological functions (reviewed in (Salter et al., 2016)). Physiologically these proteins alter the informational content encoded in the genome through a range of processes: acting as RNA editors to create sequence changes in 3'UTRs of mRNAs, affecting translation (APOBEC1) (Cole et al., 2017; Rayon-Estrada et al., 2017), acting as DNA mutators to create novel gene variants, restrict viruses and retrotransposons (AID and APOBEC3) (reviewed in (Harris & Dudley, 2015; Knisbacher et al., 2016)) and, changing DNA 5mC modification levels, leading to modulation of transcript abundance (AID) (Bochtler et al., 2017; Dominguez & Shaknovich, 2014).

APOBEC2 is an evolutionarily conserved member of the AID/APOBEC family (Conticello, 2008). Substantial evidence highlights the biological relevance of APOBEC2 in metazoans. In mice, APOBEC2 is highly expressed in cardiac and skeletal muscle where it affects muscle development (Anant et al., 2001; W. Liao et al., 1999). Specifically, in the absence of APOBEC2, there is a shift from fast to slow muscle fiber formation, a reduction in muscle mass, and the development of a mild myopathy with age (Sato et al., 2010). In zebrafish, APOBEC2 has been implicated in muscle fiber arrangement (Etard et al., 2010) and in retina and optic axon regeneration (Powell et al., 2012). In frogs, APOBEC2 is important in left-right axis specification during early embryogenesis (Vonica et al., 2011). Mutations and gene expression changes of APOBEC2 have also been linked to cancer development (Bierkens et al., 2013; Lohr et al., 2012; Okuyama et al., 2012).

Even though there is functional evidence of the biological role of APOBEC2 there is little insight as to the mechanism by which APOBEC2 accomplishes these effects. Moreover, there has been no definite demonstration of its predicted cytidine deaminase enzymatic activity or potential physiological substrate. Based on its homology with the other AID/APOBEC family members, it has been hypothesized that APOBEC2 may be involved in RNA editing (W. Liao et al., 1999; Okuyama et al., 2012) or DNA demethylation (Guo et al., 2011; Powell et al., 2012; Rai et al., 2008). It has also been hypothesized that it has lost its deaminase activity altogether and may act biologically by a novel mechanism (Powell et al., 2014). However, there is currently a lack of knowledge on the direct physiological targets of APOBEC2, and its mechanism of action.

To answer some of these questions, here we perform knockdown studies of APOBEC2 during the differentiation of the C2C12 murine myoblast cell line to systematically characterize the transcriptome and DNA methylation patterns of APOBEC2 deficient and control C2C12 cells. While our results do not support APOBEC2 roles either on RNA editing and on DNA methylation, we find that APOBEC2 downregulation leads to substantial gene expression changes affecting programs associated with myogenesis. Moreover, genomic occupancy experiments demonstrate that APOBEC2 interacts with chromatin at promoters of genes that are repressed genes during myoblast differentiation. Furthermore, these genes are derepressed with APOBEC2 knockdown, which allude to APOBEC2 acting as a transcriptional repressor. Notably, these occupied derepressed genes are not directly involved in myogenesis or muscle differentiation. We show that APOBEC2 directly interacts

with Histone Deacetylase 1 (HDAC1) repressor complexes, which supports the molecular function of APOBEC2 as a transcriptional repressor. Furthermore, APOBEC2 chromatin binding regions are enriched for transcription factor (TF) motifs, such as SP/KLF factors. Through biochemical analyses with recombinant protein, we demonstrate that APOBEC2 is both able to directly bind an SP1 motif, and cooperatively enhance SP1 DNA binding affinity. Taken together, our data suggest that APOBEC2 has a direct role in regulating active gene transcription during myoblast differentiation as a transcriptional repressor.

RESULTS

APOBEC2 is required for myoblast to myotube differentiation

The C2C12 myoblast cell line was originally derived from mouse satellite cells activated to proliferate after muscle injury in adult mice (Blau et al., 1985; Yaffe & Saxel, 1977). C2C12 myoblasts are thought to recapitulate the first steps of muscle differentiation in culture and upon differentiation induce APOBEC2 expression (Vonica et al., 2011) (validated in Supplementary Fig. 1), making them a suitable model to investigate putative roles of this suspected cytidine deaminase *in situ*.

To explore the role of APOBEC2 during myogenesis, we reduced APOBEC2 protein levels using short hairpin RNA (shRNA) against APOBEC2 mRNA. Protein reduction was efficient and coincided with reduced myoblast to myotube differentiation, evidenced by the decrease in expression of TroponinT and myosin heavy chain (MyHC), protein markers of late differentiation (Fig. 1a). At the cellular level, an overall cellular differentiation delay was evident: downregulation of APOBEC2 protein levels coincided with reduced myotube formation (Fig. 1b). Thus, APOBEC2 downregulation impaired myoblast-to-myotube differentiation. These observations match those previously reported using mouse embryonic stem cell-derived myogenic precursors (Carrió et al., 2016).

APOBEC2 loss leads to gene expression changes related to muscle differentiation

To study how APOBEC2 loss leads to problems in C2C12 myoblast-to-myotube differentiation, we performed RNA sequencing (RNA-Seq) to compare the transcriptome dynamics of APOBEC2 knockdown and control cells during differentiation. We observed that reduced APOBEC2 levels led to substantial gene expression changes (Fig. 1c). Notably, genes downregulated during differentiation were enriched for muscle development Gene Ontology (GO) terms, whereas genes upregulated were enriched for GO terms related to immune response (Fig. 1c and Supplementary File 1). The decreased expression of muscle differentiation related genes reflects the observed reduced myotube formation. Moreover, genes involved in muscle GO terms were also downregulated at day 0 prior to inducing C2C12s to differentiate. Though undetectable on the immunoblot (Fig. 1a), perturbation of APOBEC2 levels prior to inducing differentiation seems to affect the potential of C2C12 to differentiate into myotubes.

Furthermore, genes related to cell development or differentiation GO terms, particularly immune system development, blood vessel development, and nervous system development, are among those overrepresented in the list of upregulated genes with APOBEC2 deficiency (Supplementary File 2). These spurious non-muscle transcriptional programs possibly reflect transdifferentiation events, which have been observed for C2C12 myoblasts (Watanabe et al., 2004).

We next tested possible molecular mechanisms of how APOBEC2 causes these gene expression changes. Due to the conserved cytidine deaminase domain within the AID/APOBEC family, APOBEC2 is posited to be an RNA editor (W. Liao et al., 1999;

Okuyama et al., 2012) and a DNA demethylase (Guo et al., 2011; Powell et al., 2012; Rai et al., 2008). Upon comparing the transcriptomes of the APOBEC2 knockdown and control C2C12 cells for instances of C-to-U RNA editing, using a previously validated pipeline (Harjanto et al., 2016), we could not identify C-to-U or A-to-I RNA editing events that were APOBEC2 dependent (Supplementary Fig 2a). Similarly, using enhanced reduced-representation bisulfite sequencing (ERRBS), we were unable to observe significant methylation differences between the APOBEC2 knockdown and control C2C12 cells that could account for the gene expression changes (Supplementary Fig 2b,c). Altogether, these results strongly indicate that APOBEC2 is neither involved in mRNA editing (deamination) nor DNA demethylation in differentiating muscle.

APOBEC2 binds chromatin at promoters

Cytidine deaminases of the AID/APOBEC family can bind and mutate DNA either at gene bodies, e.g. exons of the immunoglobulin locus, as catalyzed by AID, reviewed in (Laffleur et al., 2014) or at promoter regions, e.g. local hypermutations as catalyzed by APOBEC3 family members - reviewed in (Chan & Gordenin, 2015). To assess whether APOBEC2 could also bind genomic DNA, we first determined the subcellular localization of APOBEC2 in muscle cells. We observe that APOBEC2 is present in both the cytoplasm and nucleus of differentiated myotubes (Supplementary Figure 3a). Although a weak nuclear localization signal (NLS) can be predicted by cNLS Mapper (Kosugi et al., 2009) (APOBEC2 residues 26 to 57, with a score of 3.7), full length APOBEC2 does not show NLS activity but is homogenously distributed throughout the cell, presumably through passive diffusion (Patenaude et al., 2009). To then assess whether nuclear APOBEC2 could bind chromatin, we utilized sequential salt extraction based on the principle that loosely bound proteins will be dissociated at low salt concentration, while tightly bound ones will not (Porter et al., 2017). Using this technique, we found that a fraction of APOBEC2 within differentiating C2C12 cells, remains bound to chromatin even at high salt concentrations (up to 1 M NaCl). As a comparison histone H4 dissociates completely from DNA at 0.75 M NaCl (Supplementary Fig. 3b). These data suggest a strong association of nuclear APOBEC2 with chromatin in differentiating myoblasts.

To then determine APOBEC2 binding specificity within chromatin, we performed chromatin immunoprecipitation-sequencing (ChIP-Seq) experiments, and calculated enrichment of APOBEC2 at specific loci over input using MACS2 (Feng et al., 2011; Zhang et al., 2008; see Materials and Methods). We performed each experiment in triplicate, and only peaks that were called in at least 2 out of 3 replicates were analyzed. Importantly, we queried APOBEC2 occupancy at two different time points, 14- and 34-hours post-differentiation, that precede the RNA-Seq time points, where we observed changes in gene expression and represent time points of low and higher APOBEC2 protein abundance. Overall, the signal around peak summits of transcription start sites (TSS) is higher at 34 versus 14 hours, reflecting an increase of APOBEC2 in chromatin. In contrast, input controls show lower enrichment over the peak summits (Fig. 2a).

Annotating the APOBEC2 peaks by genomic feature showed that for both time points most of the peaks fall within promoters, defined as regions -2 kilobases (kb) to +2 kb around the TSS (Fig. 2b). Focusing on the promoter bound genes we determined that there are ~1500 genes that are bound by APOBEC2 near their transcription start sites in any of the time points, 590 of which are occupied at both time points (Fig. 2c). Overall, about 79% of the genes that are bound by APOBEC2 in the 14-hour time point remain bound at the 34-hour time point. Further, using Gene Set Enrichment Analysis to determine the distribution of APOBEC2 occupied genes at both time points in a list of ranked expression changes either through differentiation (day 0 to 2) or upon APOBEC2 knockdown (A2 shRNA vs GFP shRNA at day 2). Our results show that APOBEC2 occupied genes are significantly enriched at genes downregulated through differentiation. Moreover, upon APOBEC2 knockdown,

APOBEC2 occupied genes are instead enriched at upregulated genes (Fig. 2c). Interestingly, overall expression of APOBEC2 bound genes through C2C12 differentiation from day 0 to 2 is significantly decreased during differentiation compared to genome wide expression changes (Fig. 2d; -0.3 versus 0.1 mean log2 fold changes; t = -8.2; adjusted P-values < 0.0001). Furthermore, expression of APOBEC2 occupied genes increase upon APOBEC2 knockdown (t-statistic = 4.0; adjusted P-value < 0.001) highlighting a global repressive role of APOBEC2 during differentiation. Altogether, these results suggest that the observed transcriptional changes linked to APOBEC2 deficiency are due to APOBEC2 acting as a transcriptional repressor.

APOBEC2 represses expression of occupied genes

To validate its possible role as a transcriptional repressor, we selected candidate genes repressed by APOBEC2 occupancy. We narrowed it down to a list of genes occupied by APOBEC2 which are downregulated during differentiation (day 2) and upregulated with APOBEC2 knockdown. In this gene list, we did not find an overrepresentation of GO terms relating to muscle differentiation; but rather, we found terms related to development of other lineages (Supplementary Table 1) similar to the non-muscle genes upregulated with APOBEC2 deficiency. We speculate that APOBEC2 is acting as a repressor to direct C2C12 differentiation into the muscle-lineage by repressing specious gene networks related to other lineages – fine-tuning the alleged promiscuity of muscle identity transcription factor, MYOD1 (Lee et al., 2020).

Next, we wanted to determine whether APOBEC2 occupies specific motifs within promoter regions. The members of the AID/APOBEC family that bind to DNA have some local sequence preference with regard to sites they mutate, but do not display rigorous sequence specificity (e.g. akin to TF binding sites (Kohli et al., 2010). To assess the motif signatures in those regions we assessed through motif enrichment the over-representation of transcription factor *8*-mers sequences associated to main transcription factor modules (Mariani et al., 2017). Among 108 TF modules tested against a controlled background of negative sequences (see Methods), we observed 19 of those significantly enriched at APOBEC2 regions in at least one differentiation time point. Specificity protein 1-like factors (SP) and Krüppel-like factor (KLF) motifs were among the top enrichments observed (Fig. 3a), suggesting a co-regulatory role between these factors and APOBEC2. In general, TF specificity groups increase their significance between 14 and 34 hours but with lower overall effect sizes or only at the later time points, suggesting that the interplay between TFs and APOBEC2 occupied regions is specific at 14 hours but broader at later time points, likely as consequence of events related to its early binding. Additionally, we did not observe APOBEC2 related DNA mutation at the occupied peaks, strongly implying that APOBEC2 is not a DNA mutator like other AID/APOBEC family members (Supplementary Fig. 4).

APOBEC2 binds DNA in vitro

Thus far, the results suggest that APOBEC2 has a novel molecular role unique to the AID/APOBEC family underscored by the ability to bind chromatin. AID/APOBEC family proteins have been shown to be RNA editors, DNA mutators, or DNA demethylases. In all cases, these proteins bind single-stranded nucleic acid and catalyze cytosine deamination within the desired substrate (Salter & Smith, 2018). Our data provide evidence that APOBEC2 binds chromatin and regulates transcription. This implies that APOBEC2 interacts directly with double-stranded DNA (dsDNA) at promoter regions, or that it interacts with transcription regulators that do so, or both.

To determine whether APOBEC2 is capable of binding DNA directly, we generated recombinant APOBEC2 (expressed in insect cells, Supplementary Figure 5a) and tested whether it could bind *in vitro* DNA oligonucleotides that represent APOBEC2 bound

sequences in cells. Specifically, we chose a 30 nucleotide DNA sequence found within an APOBEC2 ChIP-Seq peak containing an SP1 DNA sequence recognition motif since SP/KLF motifs were among the top enriched motifs (Fig. 3a). We chose a gene, ATP2A2, whose occupancy by APOBEC2 coincides with transcriptional repression, suggesting function. We found that recombinant APOBEC2 was able to bind both single-stranded DNA (ssDNA) and dsDNA SP1 probe (Fig. 3b). Members of the AID/APOBEC family are known to bind ssDNA or ssRNA substrates through their cytidine deaminase domain (Salter & Smith, 2018). However, it is remarkable that APOBEC2 is uniquely able to bind dsDNA.

Since APOBEC2 binds sequences with TF binding motifs, we tested for cooperative effects on DNA binding between APOBEC2 and SP1, since SP/KLF motifs were the top enriched motifs (Fig. 3a). Our data show that recombinant SP1 can bind the SP1-motif oligo by itself and the addition of increasing amounts of APOBEC2 increases the binding of either protein (Fig. 3c). Furthermore, the cooperativity is competed only by unlabeled SP1 probe (Fig. 3d, Supplementary Fig. 5b-d). The cooperativity seems specific to APOBEC2 and SP1 as a closely related protein to APOBEC2, Activation-induced cytidine deaminase (AID), did not show cooperativity (Supplementary Fig. 5b). However, we are unable to completely compete the APOBEC2 shift indicating low dissociation of APOBEC2 once bound to the DNA probe.

We then investigated how APOBEC2 cooperatively enhances SP1 DNA binding. We tested a presumed catalytic mutant APOBEC2 (E100A) that showed similar ability to bind (Fig. 3d). This suggests that enhanced binding does not involve base deamination, a finding concordant with the lack of DNA deamination in APOBEC2 occupied regions (Supplementary Fig. 4). Further, there is no detectable direct interaction between APOBEC2 and SP1 as there is neither a detectable supershift by EMSA nor a direct molecular interaction between the two by co-immunoprecipitation (Co-IP) (Fig. 3c, d; Supplementary Fig. 6). These data rather suggest either that binding of APOBEC2 to the probe might generate a more favorable conformation for SP1 binding (similar to ssDNA binding by A3A (Kouno et al., 2017; Shi et al., 2017)), or that a synergistic interaction between SP1 and APOBEC2 is solely transient – undetectable by the Co-IP. We have yet to show cooperativity of APOBEC2 with other candidate TFs (Fig. 3a); however, it has also been reported that APOBEC2 enhances Pou6f2 recognition of an OCT1 DNA probe (Powell et al., 2014).

Previous studies suggest that recombinant APOBEC2 expressed in *E.coli* is incapable of binding DNA *in vitro* (Powell et al., 2014) or deaminating it (Harris et al., 2002; Lada et al., 2011; Mikl et al., 2005; Nabel et al., 2012; Powell et al., 2014); our experiments using recombinant protein produced in eukaryotic cells suggest that a modification or small molecule co-factor may be necessary for APOBEC2 activity – similar to deaminase activity of ADAR2 (Macbeth et al., 2005). Nonetheless, this is the first report of APOBEC2 binding dsDNA directly, in a fashion that is enhanced by cooperative binding with a transcription factor *in vitro*.

APOBEC2 interacts with co-repressor complexes in vivo

Binding of APOBEC2 *in vivo* and *in vitro* suggest that it can bind directly to promoter regions and increase affinity for SP/KLF family factors on its cognate motif. Furthermore, there is a correlation between APOBEC2 promoter occupancy and transcriptional repression in myoblast differentiation. To ascertain the mechanism of action for the observed transcriptional repression, we used occupied peak data from ChIP-Atlas (Oki et al., 2018) to identify transcriptional regulators co-occupied with APOBEC2 occupied genes in C2C12 and myoblast/muscle datasets (*N*=84 factors). We focused our search to regulators enriched at APOBEC2 target genes (versus non-target genes) showing differential expression in normal differentiation and upon APOBEC2 knockdown (normalized as Z-scores). Among the

selected factors enriched at APOBEC2 target genes, three transcriptional regulators stood out: SIN3B, SIN3A and HDAC1 (Fig. 4a; see Supplementary Figure 7). SIN3B, SIN3A, and HDAC1 are components of HDAC co-repressor complexes involved in histone deacetylation, chromatin remodeling, and transcriptional repression (Perissi et al., 2010). Each of these regulators was more enriched at genes downregulated during myoblast differentiation versus upregulated ones at both ChIP-Seq time points (maximum Z-score at 34 h = 4.1 and 0.2 for down- and upregulated genes between day 2 and day 0, respectively; at 14 h = 5.1 and 0.2, day 1 versus day 0). Interestingly, enrichment of these regulators increased in putatively repressed APOBEC2 genes that were significantly upregulated upon APOBEC2 knockdown (maximum Z-score at 34 h = 0.1 and 2.3 for down- and upregulated genes between day 2 and day 0, respectively; at 14 h = 2.7 and 2.3, day 1 versus day 0). This result suggests the requirement of APOBEC2 for the downregulation of genes during myoblast differentiation through interactions with the HDAC1 complex.

Next, we used proximity-dependent protein biotinylation (BioID) to confirm and identify other putative APOBEC2 proximal protein complexes that may indicate direct interactions. After statistical curation we identified 361 proteins that were significantly more tagged by APOBEC2-BirA and/or BirA-APOBEC2 than BirA or BirA-GFP controls in 293 cells. Interestingly, we identified several clusters involved in transcriptional regulation: transcription, splicing, chromatin remodeling and histone modification complexes (Supplementary Fig 7 and Supplementary Table S2). Of special interest was the identification of a cluster nucleate around histone deacetylase HDAC1 (Fig. 4b) that is consistent with the enriched transcriptional repressors found at APOBEC2 peaks with the ChIP-Atlas (SIN3A, SIN3B and HDAC1, Fig. 4a). Using co-IP, we validated that APOBEC2 and HDAC1 interact directly in differentiated C2C12 myoblasts (Fig. 4c).

Taken together, our results suggest a direct role of APOBEC2 in repressing active genes, likely mediated by an HDAC1 co-repressor complex during C2C12 myoblast differentiation (Fig. 4d).

DISCUSSION

There have been many hypothesized roles for the cytidine deaminase APOBEC2. Here we show that the expression of APOBEC2 during myoblast differentiation has consequential effects on myotube formation owing to an unexpected molecular function: transcriptional control. We discovered that APOBEC2 loss, leads to faulty myoblast differentiation and concomitant gene expression changes. We show that these gene expression changes come about through direct DNA binding of APOBEC2 in active promoters – with no observed APOBEC2 related changes in RNA editing or DNA methylation. Prior work has linked APOBEC2 overexpression to RNA editing of specific transcripts (PTEN and EIF4G2) as observed in the healthy livers of transgenic mice which eventually develop hepatocellular carcinoma (Okuyama et al., 2012). Notably, RNA editing was only detectable in the liver at specific transcripts for the transgenic mice. However, based on our transcriptome analysis, we were unable to find evidence of such RNA editing in our myoblast differentiation models. Prior work has reported mild effects of APOBEC2 on DNA methylation specifically at the myogenin (MYOG) promoter (Carrió et al., 2016); from our ChIP-Seq data we do not find occupancy at the MYOG gene. Furthermore, there is conflicting data on the role of the AID/APOBEC family in active DNA demethylation; and, APOBEC2 dependent demethylation has not been found in other cellular contexts (Nabel et al., 2012; Powell et al., 2013).

We find instead that APOBEC2 is capable of binding chromatin *in vivo* and dsDNA *in vitro*, in a concentration dependent manner. Moreover, we observe that APOBEC2 enhances SP1 transcription factor binding to its cognate motif. The mechanism of increased transcription factor DNA affinity was unrelated to the putative deaminase domain as the

deaminase mutant (E100A) had comparable effects to the wildtype protein. The results here, as well as those published by others, suggest that APOBEC2 does not have a nucleic acid deaminase function. Nevertheless, we show that APOBEC2 dsDNA binding leads to repression of APOBEC2 occupied genes through the direct protein-protein interaction of APOBEC2 with HDAC1 co-repressor complexes. It is possible that APOBEC2 regulates SP1 (and potentially other TFs) transcriptional activity as activator or repressor through HDAC1 recruitment at specific genes, similar to E2F and SP1-HDAC1 transcriptional regulation (Doetzlhofer et al., 1999).

We hypothesize that APOBEC2 acts as a modulator of these active promoters during the myogenic program – fine-tuning it for muscle differentiation and repressing other lineage programs. MYOD1 has been shown to bind and activate lineage programs outside the muscle lineage; however, this is mitigated by co-repressors (Lee et al., 2020). Additionally, the interaction with RCOR1 (Fig. 4b) and potentially the REST/NRSF co-repressor complex further supports the idea of APOBEC2 repressing other lineages, such as neurogenesis, to bolster the myogenic lineage in myoblasts (Su et al., 2005). Furthermore, as APOBEC2 expression under healthy conditions is mostly confined to muscle tissue (both skeletal and heart), where it might be acting as a 'many-but-muscle' lineage repressor (Mall et al., 2017).

At the molecular level, the N-terminus of APOBEC2 (which is unique amongst AID/APOBEC family members) contains a region that is glutamate-rich and intrinsically disordered (Krzysiak et al., 2012). Proteins with similar disordered regions have been shown to form liquid phase separated membrane-less compartments (Banani et al., 2017). In particular, this is an emerging paradigm in transcriptional control, where transcription factors with disordered regions form liquid phase separated compartments (Boija et al., 2018; Sabari et al., 2018). We hypothesize that through a similar mechanism APOBEC2 interacts with such phase separated compartments at active chromatin to modulate transcription through HDAC1 co-repressor complexes during myoblast differentiation. Interestingly, from the BioID data, there are direct interactions of APOBEC2 to spliceosome components, especially Serine/Arginine-rich splicing factors, that are also found in nuclear liquid phase separated compartments (Herzel et al., 2017).

The discovery that APOBEC2 has a direct role in transcriptional control impacts how we interpret the phenotypes that have been attributed to it in other biological systems – zebrafish retina and optic nerve regeneration, *Xenopus* embryo development, mouse muscle atrophy, and cancer development. Directly or indirectly these observations likely relate to aberrant transcriptional programs related to tissue development or cell differentiation due to APOBEC2 transcriptional control. It is likely that APOBEC2 has transcriptional roles beyond muscle differentiation.

EXPERIMENTAL PROCEDURES

Raw data submission

High throughput sequencing datasets are all found in: GSE117732 and more specifically: RNA-Seq (GSE117730); ChIP-Seq (GSE117729); ERRBS (GSE117731)
Mass spectrometry data for BioID performed in Flp-In 293 T-REx cells have been deposited in MassIVE under ID (to be submitted by JMDN group)

C2C12 cell culture

C2C12 cells (CRL-1772, ATCC) were maintained in DMEM (30-2002, ATCC) with 10% fetal bovine serum and fed every two days. To differentiate equal number of cells (2.5 x 10^5) were seeded in 6-well plates followed by media change to DMEM with 2% horse serum after

12 hours. For generating single cell clones for RNA- Seq and RRBS experiments C2C12s were sorted using fluorescence-activated cell sorting (FACS) and seeded into a 96 well plate. Each clone was expanded and tested for successful knockdown through immunoblotting.

Protein knockdown using lentivirus infection

C2C12 cells were infected with lentiviruses carrying shRNAs, targeting APOBEC2 (or GFP as a control), which allows for infection of both proliferating and non-proliferating cells and provides a stable repressive effect. All APOBEC2 shRNAs were obtained from The Broad Institute's Mission TRC-1 mouse library and present in pLKO.1-puro construct, which carries the puromycin-resistance gene and drives shRNA transcription from a human U6 promoter. Plasmids used: pLKO.1 - TRC cloning vector (Addgene, # 10878) (Moffat et al., 2006); pLKO.1 puro GFP siRNA (Addgene, # 12273) (Orimo et al., 2005). The design of shRNAs and cloning in pLKO.1-TRC, were done according to the Addgene protocol (Protocol Version 1.0. December 2006). The following shRNAs sequences were used for APOBEC2 knockdown:

A2 shRNA	GCTACCAGTCAACTTCTTCAA
GFP shRNA	GCAAGCTGACCCTGAAGTTCAT

Lentiviruses were produced by co-transfection of pLKO.1-puro short hairpins containing construct, packaging plasmid psPAX2 (Addgene, #12260) and envelope plasmid pMD2.g (Addgene, #12259) in HEK 293T cells. Transfections were done using Lipofectamine 2000 Reagent (Invitrogen) as per manufacturer instructions. Supernatants with lentiviral particles were collected at 24 and 48 hours after transfection, passed through a 0.45 mm filter and applied to C2C12s. For APOBEC2 constitutive knockdown 30% confluent cells were infected with pLKO.1 containing lentiviruses in growth media containing 8 µg/mL polybrene for 12 hours. Two days after lentiviral infection cells were cultured with 4 µg/ml virus-free puromycin containing media for two more days to select cells stably expressing the shRNAs.

Immunoblotting and co-immunoprecipitation

For immunoblotting experiments, C2C12 cells were first washed with cold PBS and lysed in 100 µl RIPA lysis buffer (Santa Cruz, sc-24948) in 6-well plates. They were incubated at 4°C for 15 minutes and then extracts were scrapped into a microfuge tube. Lysates were snap frozen in liquid nitrogen to improve efficiency of lysis. After thawing the lysates on ice and clearing out cell debris by centrifugation, same amount of total protein (ranges between 10-30 µg) was boiled in SDS-PAGE sample buffer and loaded onto each lane of a polyacrylamide gel (Criterion XT Bis-Tris Gel 12%, Bio-Rad). Following electrophoresis, the resolved protein was transferred to a nitrocellulose membrane and subjected to western blot (WB) analysis. The source and dilution for each antibody used were: polyclonal rabbit-APOBEC2 (gift from Alin Vonica MD, PhD), 1:1000; monoclonal mouse-APOBEC2 (clone 15E11, homemade), 1:5000; TroponinT clone JLT- 12 (T6277, Sigma-Aldrich), 1:500; α-tubulin DM1A (Abcam, ab7291), 1:5000; MyHC MF-20 (DSHB), 1:20.

For co-immunoprecipitation experiments, 4 x 10^6 C2C12 cells were plated and lysed after 4 days in differentiation medium. Cells were trypsinized, washed with cold PBS, and lysed in 1 mL cell lysis buffer: 0.5% Tween 20, 50 mM Tris pH 7.5, 2 mM EDTA, and freshly added 1X Halt protease and phosphatase inhibitor cocktail EDTA-free (Thermo, 78441). Mixture was vortexed and incubated on ice for 10 min, twice. Nuclei were separated from the cytoplasmic fraction by centrifugation (6000 g, 1 minute, 4°C). Nuclei were washed with 1 mL cold PBS before lysing in 250 µL high salt nuclear lysis buffer: 800 mM NaCl, 1% NP40 (Igepal CA-640), 50 mM Tris pH 7.5, and freshly added 1X protease and phosphatase inhibitor cocktail, EDTA-free. Mixture was vortexed and incubated on ice for 10 min, twice. Nuclear lysate was then diluted to a final salt and detergent concentration of 400 mM NaCl and 0.5% NP40

using 250 µL dilution buffer: 50 mM Tris pH 7.5 and 1X protease and phosphatase inhibitor cocktail, EDTA-free. Nuclear lysates were treated with Benzonase (Merck-Millipore, 70664). Nuclear lysates were pre-cleared on 25 µL Dynabeads M-280 Sheep anti-mouse or anti-rabbit IgG (Thermo, 11201D/12203D), depending on primary IgG antibody. Pre-cleared nuclear lysate was then added to 50 µL beads conjugated with 2-4 µg primary IgG antibody: rabbit anti-APOBEC2 (Sigma, HPA017957), rabbit anti-SP1 (Merck, 07-645), or rabbit IgG isotype control. Immunoprecipitation was done overnight at 4C with rotation. Beads were thoroughly washed before resuspending and boiling in SDS-PAGE sample buffer.

Immunofluorescence staining and fusion index of C2C12s

C2C12 cells ($5x10^4$ cells) were seeded in collagen-coated coverslips (BD Biosciences, 356450) in 12- well plate the day before inducting of differentiation with 2% horse serum. They were washed with cold PBS and fixed with paraformaldehyde (4%) in PBS for 10 minutes at 4 °C. This was followed by 2 washes, 5 minutes each at room temperature and blocking solution (0.5% BSA, 1% gelatin, 5% normal goat serum, 0.1% Triton) in PBS for 1 hour at room temperature. This was followed by overnight stain with antibodies in a humidified chamber at 4°C, three washes with cold PBS 5 min each at room temperature. Coverslips were then incubated with secondary antibodies for 1 hour at room temperature, followed by three washes with PBS 5 min at room temperature. Immunofluorescence staining of C2C12 cells was carried out with primary antibodies: MyHC MF20 (DSHB) and FLAG M2 (Sigma, F1804). Nuclei were counterstained and coverslips were mounted with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories, H-1200). Images were taken using confocal Leica TCS SP5 II or widefield Zeiss Cell Observer and image analysis was done with Fiji/ImageJ.

Salt-extraction profiling

C2C12 cells were seeded in equal numbers ($2x10^6$ cells) and induced to differentiate after 12 hours. Five days after differentiation they were lysed in the plate with 100 µl sucrose lysis buffer (320 mM sucrose, 0.5% NP-40, 10 mM Tris pH 8.0, 3 mM $CaCl_2$, 2 mM Mg acetate, 0.1 mM EDTA). Extracts were incubated for 5 minutes on ice and spun at 500 g for 5 minutes to collect the nuclear pellet and supernatant as the cytosol fraction. Nuclear pellets were washed with no-salt Nuclei Buffer (50 mM Tris pH8, 1% NP-40, 10% glycerol). Following the washes the nuclear proteins were extracted at increasing concentrations of NaCl from 250 mM up to 2 M in Nuclei Buffer during which they are homogenized using dounce tissue grinders (Fisher, K8853000000), incubated on ice for 10 minutes and spun at 4°C for 10 additional minutes. Eluted material was collected, resolved on polyacrylamide gel electrophoresis (Criterion XT Bis-Tris Gel 12%, Bio- Rad) and immunoblotted with specific antibodies: Histone H4 (Merck, 05-858R), 1:5000; monoclonal mouse-APOBEC2 (clone 15E11, homemade), 1:5000; α-tubulin DM1A (Abcam, ab7291), 1:5000.

RNA expression analysis

Library preparation and sequencing were done by Rockefeller University Genomics Resource Center [https://www.rockefeller.edu/genomics/] using TruSeq Stranded mRNA Sample Prep kit as per manufacturer's instruction. The procedure includes purification of poly-adenylated RNAs. Libraries were sequenced with 50bp paired-read sequencing on the HiSeq2500 (Illumina). Paired end read alignments and gene expression analysis were performed with the Bioinformatics Resource Center at Rockefeller University. Paired-end reads were aligned to mm10 genome using the subjunc function in the Bioconductor Rsubread (Y. Liao et al., 2013) package and bigWig files for visualization were generated from aligned reads using the Bioconductor rtracklayer (Lawrence et al., 2009) and GenomicAlignments packages (Lawrence et al., 2013). For analysis of differential expression, transcript quantifications were performed using Salmon (Patro et al., 2017) in

quasi-mapping mode. Gene expression values were calculated from transcript quantifications using tximport (Soneson et al., 2015). Gene expression changes were identified at a cut off of 5% FDR (benjamini-hockberg correction) using the Wald test implemented in DESeq2 (Love et al., 2014). Annotation files used: BSgenome.Mmusculus.UCSC.mm10(v1.4.0);org.Mm.db(v3.5.0); TxDb.Mmusculus.UCSC.mm10.knownGene.gtf.gz(v3.4.0)

RNA editing analysis

RNA editing analysis was performed as previously reported elsewhere (Harjanto et al., 2016). Editing detection was performed by comparing C2C12 control samples (GFPsh) to APOBEC2 knockdown samples using RNA-Seq datasets in triplicates for each sample. Minimum filters include quality control thresholds (minimum of five reads covering the putative site with at least two reads supporting the editing event; filtering of reads that contain indels or support an edit in the first or last two base pairs of a read). Stringent filters applied to the APOBEC1 dependent C-to-U edited sites include all of the above and additionally the magnitude of the control vector was at least 15 and the angle between the wild-type and knockout vectors was at least 0.11 radians, as described in the paper referenced in this section.

Enhanced Reduced representation bisulfite sequencing (ERRBS)

ERRBS library preparation, sequencing and read alignment was performed by the Epigenomics Core Facility of Weill Cornell Medicine [epicore.med.cornell.edu/] as previously described (Akalin, Garrett-Bakelman, et al., 2012; Garrett-Bakelman et al., 2015). The procedure includes bisulfite conversion of the DNA. Libraries were sequenced with 50bp single reads (SR) in HiSeq2500 (Illumina). Reads were aligned to a bisulfite converted reference mouse genome with Bismark (Krueger & Andrews, 2011). The methylation context for each cytosine was determined with scripts from the core facility.

Here coverage of specific genomic regions by ERRBS dataset, refers to the percent of features (eg percent of promoters, CpG islands) that contain at least one CpG that is well covered (> 10x). For gene specific annotations the mm10 UCSC knownGene annotations from the UCSC table browser were used and for CpG islands the mm10 cpgIslandExt track of the UCSC table browser. Genomic features were defined as: CpG islands, CpG island shores were defined as 2kb upstream and downstream of a CpG island; Gene promoters (region 2kb upstream and 2kb downstream of the TSS), exons, introns and intergenic regions.

Differential methylation analysis

MethylKit (v1.3.8) (Akalin, Kormaksson, et al., 2012) was used to identify differentially methylated cytosines (DMCs) with q-value less than 0.01 and methylation percentage difference of at least 25% after filtering ERRBS dataset by coverage, normalizing by median and including CpG sites that are covered >10x, in 3 out of 5 biological replicates (lo.count = 10, lo.perc = NULL, hi.count = 1000, hi.perc = 99.9), (destrand=TRUE,min.per.group=3L). eDMRs (v0.6.4.1) (Li et al., 2013) was used to empirically determine differentially methylated regions, using the DMCs identified with methylKit. In order for a region to be defined as a DMR, default parameters (num.DMCs=1, num.CpGs=3, DMR.methdiff=20) of eDMR were used, so that each region has: (1) at least 1 DMC in the region, as determined using methylKit, (2) at least 3 CpGs included in the region and (3) absolute mean methylation difference greater than 20%. For a region to be defined as a significant DMR, default parameters were used (DMR.qvalue = 0.001, mean.meth.diff = 20, num.CpGs = 5, num.DMCs = 3) so that each significant DMRs has (1) 5 CpGs where at least 3 of them are

significant DMCs as determined by methylKit (2) have a minimum 20% methylation change for the region.

Chromatin immunoprecipitation method

C2C12s were plated at ~70% confluence 12 hours prior to inducing differentiation (seed ~2x10^6 cells) maintained in DMEM (ATCC, 30-2002) with 10%FBS. This was followed by media change to DMEM with 2% horse serum (Life Biotechnologies, 26050-088) to induce differentiation. The cells (~5x10^6 /10cm plate) were harvested at 24-hour or 34-hour after inducing differentiation. They were fixed on plate with 1% PFA in PBS for 10 minutes at RT. Glycine was added to a final concentration of 125mM. Cells were washed 2 times with 1x PBS with protease inhibitor cocktail (Roche, 11836170001). They were lysed on the plate with cold Farnham lysis buffer to ~10x10^6 cells /mL (5mM PIPES pH 8.0, 0.5% NP-40, 85mM KCl,1mM EDTA, PIC) and incubated rotating for 15min at 4°C . Lysates were scraped off the plates, pelleted and resuspended in LB2 (10 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, PIC) and incubated rotating for 15 minutes at 4°C and then centrifuged. Pellets were resuspended to 5x10^7 cells/mL in LB3 (10 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium-deoxycholate, 0.5% sodium lauroyl sarcosinate, PIC) until suspension was homogenized. Samples were then sonicated using Covaris ultrasonicator model S220 for 15 minutes with the following settings: 140W peak power, 5% duty, 200 cycles per burst. TritonX-100 to a final concentration of 1% was added to the samples. Samples were clarified by centrifugation at 20,000 g for 10 minutes at 4°C. The supernatant is the soluble chromatin extract. The soluble fragmented chromatin from ~2.5x10^7 was used for each IP. For each IP 100ul Dynabeads (Thermofisher anti-rabbit M280, 11203D) were mixed with 10ul polyclonal rabbit-APOBEC2 antibodies (gift from Alin Vonica MD, PhD) incubating overnight (~16 hours). A magnetic stand was used to separate beads from the lysate and beads were washed one time each with for 5min in: low salt wash (0.1%SDS, 1%Triton X-100, 2 mM EDTA, 20 mM Tris pH8, 150 mM NaCl, PIC), high salt wash (0.1%SDS, 1% Triton X- 100, 2mM EDTA, 20mM Tris pH8, 500mM NaCl, PIC), lithium chloride wash (150mM LiCl, 1% NP-40, 1% NaDOC, 1mM EDTA, 10mM TrispH8, PIC), TE wash (10mM Tris-HCl pH8, 1mM EDTA, 50mM NaCl, PIC). Beads were resuspended in 52 ul of elution buffer (50mM Tris-HCl pH8, 10mM EDTA, 1%SDS) and incubated at 30min at 65°C while shaking to prevent beads from settling. The eluate was transferred to a new tube, inputs of the same volume were incubated for 8 hours at 65°C to reverse the crosslink. The samples were treated with RNAse (Roche, 11119915001) for 1 hour at 37°C, and with Proteinase K for 2 hours at 55°C. Fragmented DNA was purified with Ampure beads (Agencourt AMPure XP beads A63881) as per the manufacturer's instructions.

Chromatin immunoprecipitation sequencing and analysis

The ChIP-Seq included biological triplicates for each group. ChIP-Seq libraries were prepared using NEBNext Ultra DNA Library Prep Kit as per manufacturer's instructions. Libraries were sequenced with 75 base pair single read sequencing on the NextSeq 500 (Illumina). Read alignments and initial analysis were performed with the Bioinformatics Resource Center at Rockefeller University. Single-end reads were aligned to mm10 genome using the subread function in the Bioconductor Rsubread (Y. Liao et al., 2013) package and bigWig files for visualization were generated from aligned reads using the Bioconductor rtracklayer (Lawrence et al., 2009) and GenomicAlignments packages (Lawrence et al., 2013). Quality metrics for the ChIP-Seq data were assessed using ChIPQC bioconductor package (Carroll et al., 2014), according to Encyclopedia of DNA Elements (ENCODE) working standards and guidelines for ChIP experiments (Landt et al., 2012). Reads mapping to more than one genomic location were filtered prior to peak calling using Model-based Analysis of ChIP-Seq (MACS2) (Feng et al., 2011; Zhang et al., 2008) with duplicate filtering applied and input DNA sample as a control. Peaks that are reproducible (present in 2 out of 3) were filtered for known artifact or blacklisted regions (The ENCODE Project Consortium,

2012). For each of the peaks a weighted mean location of peak summits cross biological replicates is calculated (Yang et al., 2014). The list of binding regions 100 base pairs around the mean peak summits was used for downstream analysis. Ngs.plot (v2.61) was used with specific parameters (-G mm10 -D refseq -C -L 1000 -FL 150 -P 4 -SC 0,1 -GO none -RB 0.05) to generate average profiles of ChIP-Seq reads (Shen et al., 2014). ChIPSeeker (v1.14.2) (Yu et al., 2015) and ChIPpeakAnno (3.12.7) (Zhu, 2013; Zhu et al., 2010) were used for downstream analysis after peak calling for annotation of the binding regions to the nearest gene. We created an APOBEC2 occupied gene set, using genes that show consistent APOBEC2 occupancy at both 14-hour and 34-hour time points. GSEA (v3.0) (Subramanian et al., 2005) was used for testing the enrichment of the APOBEC2 occupied gene set in the list of genes that are differentially expressed. Annotation files used: BSgenome.Mmusculus.UCSC.mm10 (v1.4.0) org.Mm.db (v3.5.0) and TxDb.Mmusculus.UCSC.mm10.knownGene.gtf.gz(v3.4.0).

Gene list analysis

Gene list analyses either by statistical overrepresentation test or statistical enrichment test were done through PANTHER (Mi et al., 2019). Briefly, gene lists were filtered based on expression (log2FoldChange, up- or downregulated at specific treatment and time point) and p-adj values (FDR< 10%) and used as input in PANTHER gene list analysis. For statistical overrepresentation tests of upregulated genes with A2 vs GFP shRNA, genes were filtered based on log2FoldChange > 0.58 and FDR < 10% at each time point and used as input list with Mus musculus (all genes in database) as reference/background list. Default parameters were followed for the analyses and are indicated in the corresponding output. For

Prediction of binding motifs

Transcription factor motifs associated to 108 TF modules (Mariani et al., 2017) were mapped against time-point specific sequences harboring APOBEC2, 200 base pairs centered on peak summits. For each time point, we defined a background set of negative sequences using scrambled regions of the positive sequences. Using both sequence sets of positives and negatives, we assessed the presence of strong 8-mers associated to each of those 108 families and their ability to classify between APOBEC2 regions and negative sequences, summarizing a Receiver-Operating Characteristic Area Under the Curve (ROC-AUC) for each of those. Assessment of significance in each case was done using a Wilcoxon rank sums test (one sided). *P*-values were corrected through a Benjamini-Hochberg procedure.

Enrichment of ChIP-seq peaks on APOBEC2 differentially expressed target genes

Using the ChIP-Atlas as a reference, we downloaded all datasets associated to myoblast or C2C12 cells (*N*=54). For each dataset, we intersected ChIP-seq peaks from APOBEC2 in each timepoint and replicate (command *fisher* in *bedtools)*, obtaining a 2 x 2 contingency matrix. The number of overlaps was linked to the closest gene using 2000bp with respect to TSS annotations in the mouse genome. The proportion of genes associated to a differential expression comparison was done by dividing the number of APOBEC2 peaks proximal to a DE-gene with a peak from a ChIP-Atlas dataset by the total number of DE-genes in that comparison. This was repeated for each gene expression contrast. Mean log2 fold change estimates for each ChIP-Atlas peak dataset were obtained by calculating the distribution of log2 fold changes between non-target DE-genes and target DE-genes, in each time point, using the three APOBEC2 ChIP-seq replicates. With those, we calculated a global mean and standard deviation across all ChIP-Atlas factors, reporting a Z-score for dataset, time point and differentially expression comparison between APOBEC2 target and non-target genes.

BioID samples preparation

APOBEC2 was cloned into pDEST-pcDNA5-BirA-FLAG N-term and pDEST-pcDNA5-BirA-FLAG C-term (gifts from Dr. Anne-Claude Gingras). Stable Flp-In 293 T-REx cell populations expressing inducible BirA-FLAG-APOBEC2 and APOBEC2-BirA-FLAG were generated by transfection and hygromycin selection. Stable Flp-In 293 T-REx cell lines with eGFP-BirA-FLAG and BirA-FLAG were gifts from Dr. Jean-Francois Cote (IRCM, Montreal, Canada). BioID samples were processed as described previously (Couzens et al., 2013).

BioID MS Data analysis

Mass spectrometry was performed at the IRCM proteomics platform. Samples were injected into Q Exactive Quadrupole Orbitrap (Thermo Fisher), and raw files were analyzed with the search engines Mascot and XTandem! (Craig & Beavis, 2004) through the iProphet pipeline (Shteynberg et al., 2011) integrated in Prohits (Liu et al., 2010), using the human RefSeq database (version 57) supplemented with "common contaminants" from the Max Planck Institute (http://maxquant.org/downloads.htm), the Global Proteome Machine (GPM; http://www.thegpm.org/crap/index.html) and decoy sequences. The search parameters were set with trypsin specificity (two missed cleavage sites allowed), variable modifications involved Oxidation (M) and Deamidation (NQ). The mass tolerances for precursor and fragment ions were set to 15 ppm and 0.6 Da, respectively, and peptide charges of +2, +3, +4 were considered. The resulting searches results were individually processed by PeptideProphet (Keller et al., 2002), and peptides were assembled into proteins using parsimony rules first described in ProteinProphet (Nesvizhskii et al., 2003) using the Trans-Proteomic Pipeline (TPP). TPP settings were the following: -p 0.05 -x20 -PPM –d “DECOY”, iprophet options: pPRIME and PeptideProphet: pP.

BioID interactions scoring

The estimation of interactions scorings was performed on proteins with iProphet protein probability ≥ 0.9 and unique peptides ≥ 2, by combining two algorithmic approaches : SAINTexpress (Teo et al., 2014) (version 3.6.1) and DESeq (Anders & Huber, 2010). Two biological replicates of APOCE2-BirA-FLAG and two of BirA-FLAG-APOBEC2 were jointly compared to their respective negative controls. These controls comprised of pulldowns of either the BirA*-Flag or the cytoplasmic EGFP-BirA*-Flag vector, each in 4 biological replicates. The SAINT analysis was performed with the following settings: nControl:4 (2 fold compression), nCompressBaits:2 (no bait compression). Interactions displaying a BFDR ≤ 0.01 were considered statistically significant. We also used DESeq2 (version 1.2.1335) an R package that applies negative binomial distribution to calculate enrichments over controls. DESeq was run using default settings and significant preys were selected by applying a ≤0.1 p-value cut-off. The combined list of significant preys obtained from DEseq and SAINT was defined as potential APOBEC2 proximity interactors.

BioID annotations and network analyses

Graphical representations of protein networks were generated with Cytoscape (Shannon, 2003) (version 3.7.2). The APOBEC2 interactome was annotated with the Gene Ontology Annotation database (Huntley et al., 2015) (GOA version 171) and the CORUM (Giurgiu et al., 2019) protein complexes database (version 3.0). Network augmentations of our BioID screens was performed by extracting prey-prey interactions from the human BioGRID (Chatr-aryamontri et al., 2017) network (version 3.5.176), and from Cytoscape’s PSICQUIC built-in web service client (October 2019 release) after searching against the IntAct (Orchard et al., 2014) and iRefIndex (Razick et al., 2008) databases. Clusters were extracted by the Markov CLustering Algorithm (MCL) from Cytoscape’s ClusterMaker2 (Morris et al., 2011) application (version 1.3.1).

DNA editing detection

We aligned all short reads from input and IP experiments to the mouse genome (GRCm38 EnsEMBL 90) using HiSAT v2.1.0 with default settings. We removed all non-unique mappers and marked all read duplicates with picard.sam.markduplicates.MarkDuplicates (v 2.5.0). We compared all samples to the reference genome using JACUSA v1.2.4 in call-1 mode with program parameters: call-1 -s -c 5 -P UNSTRANDED -p 10 -W 1000000 -F 1024 --filterNM_1 5 -T 1 -a D,M,Y -R. Diverging positions are reported if the LLR ratio exceeds 1.0. Briefly, read count distributions at every genomic position (coverage >5) are contrasted with the expected read count based on the reference base. For the pairwise comparison of all input samples stratified by conditions, we used JACUSA v.1.2.4 in call-2 mode with program parameters: call-2 -s -c 5 -P UNSTRANDED,UNSTRANDED -p 10 -W 1000000 -F 1024 --filterNM_1 5 --filterNM_2 5 -T 1 -a D,M,Y -u DirMult:showAlpha -R. Briefly, read count comparison from replicate input samples are contrasted with one another: A2 shRNA knockdown vs GFP shRNA knockdown. Diverging positions are reported if the LLR ratio exceeds 1.0.

Recombinant mouse APOBEC2 production

Bac-to-Bac Baculovirus expression system (ThermoFisher) was used for generating recombinant N terminal His tagged mouse APOBEC2 bacmid DNA and virus was generated by transfection into Sf9 cells. Mouse His tagged APOBEC2 was expressed in High-Five insect cells by infection of recombinant baculovirus. High-Five cells were harvested 48 hours after infection. Protein purification was carried out at 4°C. The cells were harvested by centrifugation at 7,500 × g and 4 °C and lysed with a cell disrupter (Avestin) in a buffer containing 20 mM Tris, pH 8.0, 300 mM NaCl, 5 mM β-mercaptoethanol, 0.5 mM phenylmethylsulfonyl fluoride (Sigma), 2 μM bovine lung aprotinin (Sigma), and complete EDTA-free protease inhibitor mixture (Roche). After centrifugation at 35000 g for 45 minutes, the cleared lysate was loaded onto an Ni-NTA column (Qiagen) and eluted with an imidazole gradient. Protein-containing fractions were pooled, dialyzed against a buffer containing 20 mM Tris, pH 8.0, 100 mM NaCl, and 5 mM DTT. Protein was bound to a heparin column (GE Healthcare) and eluted with an NaCl gradient. Protein-containing fractions were pooled, concentrated, and purified on a HiLoad Superdex 200 16/60 gel filtration column (GE Healthcare) in a buffer containing 20 mM Hepes, pH 7.5, 150 mM NaCl, and 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP).

Electromobility shift assay (EMSA)

Oligonucleotides (oligos) were ordered from Sigma-Aldrich:

SP1 oligo F: GGC GGC GCG GCC CCG CCC CCT CCT CCG GC
SP1 oligo R: GCC GGA GGA GGG GGC GGG GCC GCG CCG CC
GC oligo F: CCT CCC GCA CCG TCC CCG GCC CCT GGC AA
GC oligo R: TTG CCA GGG GCC GGG GAC GGT GCG GGA GG
STAT3 oligo F: CTC CTC CCA CTT CCT AGA AGA TCC GCC TT
STAT3 oligo R: AAG GCG GAT CTT CTA GGA AGT GGG AGG AG

Single-stranded (ss) oligos were labeled with [γ-^{32}P] ATP and annealed with complementary oligos to form double-stranded (ds) oligos. Specified recombinant protein, mouse APOBEC2, human SP1 (Sigma, SRP2030), or mouse AID (Enzymax), and ss or ds oligos were mixed in binding reaction buffer (50 mM KCl, 10 mM HEPES, 5 mM MgCl, 5% glycerol, 100 ng poly(dI-dC) and 5 mM DTT) for 40 minutes at room temperature. An equal volume of 50% glycerol was added to the reactions before running on 5% Criterion TBE gel (3450048, Bio-Rad) with 0.5X TBE buffer (1610733, Bio-Rad) for 1.5 hours at 90 volts with the tank submerged on ice. Gels were then dried and imaged with a phosphorimager system.

Acknowledgements

The authors thank Dr Diego Mourao-Sa for his insights, support and encouragement during the completion of these studies; Dr Thomas Carroll, for reproducing RNA-Seq and ChIP-Seq bioinformatics analyses; Dr. Pete Stavropoulos, Erik Debler, Philipp Schmiege and Nicholas Economos for purifying APOBEC2 antibodies.

Funding

This work is supported through funding by the Helmholz Foundation to the DKFZ (NP). Financial support was provided through the David Rockefeller Graduate Program (LM). Work in the JMDN lab is supported by the Canadian Institutes of Health Research grant PJT 155944.

Author Contribution

LM and JPL designed the study, performed experiments, analyzed data, and wrote the manuscript with supervision from FNP.SR and JPL designed and performed EMSA experiments together. DH performed RNA editing analysis. ILR performed motif prediction and enrichment of ChIP-seq peaks analysis with supervision from JBZ.PGS performed experimental work and JB performed analysis related to BioID data with supervision from JMDN.AV provided key reagent.CD performed DNA editing detection analysis.All authors wrote, read, and approved the final manuscript.

REFERENCES

Akalin, A., Garrett-Bakelman, F. E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., … Figueroa, M. E. (2012). Base-Pair Resolution DNA Methylation Sequencing Reveals Profoundly Divergent Epigenetic Landscapes in Acute Myeloid Leukemia. *PLoS Genetics*, *8*(6), e1002781. http://doi.org/10.1371/journal.pgen.1002781

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, *13*(10), R87. http://doi.org/10.1186/gb-2012-13-10-r87

Anant, S., Mukhopadhyay, D., Sankaranand, V., Kennedy, S., Henderson, J. O., & Davidson, N. O. (2001). ARCD-1, an apobec-1-related cytidine deaminase, exerts a dominant negative effect on C to U RNA editing. *American Journal of Physiology-Cell Physiology*, *281*(6), C1904–C1916. http://doi.org/10.1152/ajpcell.2001.281.6.C1904

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. http://doi.org/10.1186/gb-2010-11-10-r106

Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: Organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, *18*(5), 285–298. http://doi.org/10.1038/nrm.2017.7

Bierkens, M., Krijgsman, O., Wilting, S. M., Bosch, L., Jaspers, A., Meijer, G. A., … Steenbergen, R. D. M. (2013). Focal aberrations indicate EYA2 and hsa-miR-375 as oncogene and tumor suppressor in cervical carcinogenesis. *Genes, Chromosomes and Cancer*, *52*(1), 56–68. http://doi.org/10.1002/gcc.22006

Blau, H. M., Pavlath, G. K., Hardeman, E. C., Chiu, C. P., Silberstein, L., Webster, S. G., … Webster, C. (1985). Plasticity of the differentiated state. *Science (New York, N.Y.)*, *230*(4727), 758–66. http://doi.org/10.1126/science.2414846

Bochtler, M., Kolano, A., & Xu, G.-L. (2017). DNA demethylation pathways: Additional players and regulators. *BioEssays*, *39*(1), e201600178. http://doi.org/10.1002/bies.201600178

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V., … Young,

R. A. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*, *175*(7), 1842-1855.e16. http://doi.org/10.1016/j.cell.2018.10.042

Carrió, E., Magli, A., Muñoz, M., Peinado, M. A., Perlingeiro, R., & Suelves, M. (2016). Muscle cell identity requires Pax7-mediated lineage-specific DNA demethylation. *BMC Biology*, *14*(1), 30. http://doi.org/10.1186/s12915-016-0250-9

Carroll, T. S., Liang, Z., Salama, R., Stark, R., & de Santiago, I. (2014). Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Frontiers in Genetics*, *5*. http://doi.org/10.3389/fgene.2014.00075

Chan, K., & Gordenin, D. A. (2015). Clusters of Multiple Mutations: Incidence and Molecular Mechanisms. *Annual Review of Genetics*, *49*(1), 243–267. http://doi.org/10.1146/annurev-genet-112414-054714

Chatr-aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N. K., … Tyers, M. (2017). The BioGRID interaction database: 2017 update. *Nucleic Acids Research*, *45*(D1), D369–D379. http://doi.org/10.1093/nar/gkw1102

Cole, D. C., Chung, Y., Gagnidze, K., Hajdarovic, K. H., Rayon-Estrada, V., Harjanto, D., … Bulloch, K. (2017). Loss of APOBEC1 RNA-editing function in microglia exacerbates age-related CNS pathophysiology. *Proceedings of the National Academy of Sciences*, *114*(50), 13272–13277. http://doi.org/10.1073/pnas.1710493114

Conticello, S. G. (2008). The AID/APOBEC family of nucleic acid mutators. *Genome Biology*, *9*(6), 229. http://doi.org/10.1186/gb-2008-9-6-229

Couzens, A. L., Knight, J. D. R., Kean, M. J., Teo, G., Weiss, A., Dunham, W. H., … Gingras, A.-C. (2013). Protein Interaction Network of the Mammalian Hippo Pathway Reveals Mechanisms of Kinase-Phosphatase Interactions. *Science Signaling*, *6*(302), rs15–rs15. http://doi.org/10.1126/scisignal.2004712

Craig, R., & Beavis, R. C. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*, *20*(9), 1466–1467. http://doi.org/10.1093/bioinformatics/bth092

Doetzlhofer, A., Rotheneder, H., Lagger, G., Koranda, M., Kurtev, V., Brosch, G., … Seiser, C. (1999). Histone Deacetylase 1 Can Repress Transcription by Binding to Sp1. *Molecular and Cellular Biology*, *19*(8), 5504–5511. http://doi.org/10.1128/MCB.19.8.5504

Dominguez, P. M., & Shaknovich, R. (2014). Epigenetic Function of Activation-Induced Cytidine Deaminase and Its Link to Lymphomagenesis. *Frontiers in Immunology*, *5*(DEC). http://doi.org/10.3389/fimmu.2014.00642

Etard, C., Roostalu, U., & Strähle, U. (2010). Lack of Apobec2-related proteins causes a dystrophic muscle phenotype in zebrafish embryos. *Journal of Cell Biology*, *189*(3), 527–539. http://doi.org/10.1083/jcb.200912125

Feng, J., Liu, T., & Zhang, Y. (2011). Using MACS to Identify Peaks from ChIP-Seq Data. *Current Protocols in Bioinformatics*, *34*(1), 2.14.1-2.14.14. http://doi.org/10.1002/0471250953.bi0214s34

Garrett-Bakelman, F. E., Sheridan, C. K., Kacmarczyk, T. J., Ishii, J., Betel, D., Alonso, A., … Melnick, A. M. (2015). Enhanced Reduced Representation Bisulfite Sequencing for Assessment of DNA Methylation at Base Pair Resolution. *Journal of Visualized Experiments*, (96). http://doi.org/10.3791/52246

Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., … Ruepp, A. (2019). CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Research*, *47*(D1), D559–D563. http://doi.org/10.1093/nar/gky973

Guo, J. U., Su, Y., Zhong, C., Ming, G., & Song, H. (2011). Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell*, *145*(3), 423–434. http://doi.org/10.1016/j.cell.2011.03.022

Harjanto, D., Papamarkou, T., Oates, C. J., Rayon-Estrada, V., Papavasiliou, F. N., & Papavasiliou, A. (2016). RNA editing generates cellular subsets with diverse sequence within populations. *Nature Communications*, *7*, 12145. http://doi.org/10.1038/ncomms12145

Harris, R. S., & Dudley, J. P. (2015, May 1). APOBECs and virus restriction. *Virology*. Academic Press Inc. http://doi.org/10.1016/j.virol.2015.03.012

Harris, R. S., Petersen-Mahrt, S. K., & Neuberger, M. S. (2002). RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. *Molecular Cell*, *10*(5), 1247–1253. http://doi.org/10.1016/S1097-2765(02)00742-6

Herzel, L., Ottoz, D. S. M., Alpert, T., & Neugebauer, K. M. (2017). Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. *Nature Reviews Molecular Cell Biology*, *18*(10), 637–650. http://doi.org/10.1038/nrm.2017.63

Huntley, R. P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M. J., & O'Donovan, C. (2015). The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Research*, *43*(D1), D1057–D1063. http://doi.org/10.1093/nar/gku1113

Keller, A., Nesvizhskii, A. I., Kolker, E., & Aebersold, R. (2002). Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by {MS}/{MS} and Database Search. *Analytical Chemistry*, *74*(20), 5383–5392. http://doi.org/10.1021/ac025747h

Knisbacher, B. A., Gerber, D., & Levanon, E. Y. (2016). DNA Editing by APOBECs: A Genomic Preserver and Transformer. *Trends in Genetics*, *32*(1), 16–28. http://doi.org/10.1016/j.tig.2015.10.005

Kohli, R. M., Maul, R. W., Guminski, A. F., McClure, R. L., Gajula, K. S., Saribasak, H., … Stivers, J. T. (2010). Local Sequence Targeting in the AID/APOBEC Family Differentially Impacts Retroviral Restriction and Antibody Diversification. *Journal of Biological Chemistry*, *285*(52), 40956–40964. http://doi.org/10.1074/jbc.M110.177402

Kosugi, S., Hasebe, M., Tomita, M., & Yanagawa, H. (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proceedings of the National Academy of Sciences*, *106*(25), 10171–10176. http://doi.org/10.1073/pnas.0900604106

Kouno, T., Silvas, T. V, Hilbert, B. J., Shandilya, S. M. D., Bohn, M. F., Kelch, B. A., … Schiffer, C. A. (2017). Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nature Communications*, *8*(1), 15024. http://doi.org/10.1038/ncomms15024

Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, *27*(11), 1571–1572. http://doi.org/10.1093/bioinformatics/btr167

Krzysiak, T. C., Jung, J., Thompson, J., Baker, D., & Gronenborn, A. M. (2012). APOBEC2 Is a Monomer in Solution: Implications for APOBEC3G Models. *Biochemistry*, *51*(9), 2008–2017. http://doi.org/10.1021/bi300021s

Lada, A. G., Frahm Krick, C., Kozmin, S. G., Mayorov, V. I., Karpova, T. S., Rogozin, I. B., & Pavlov, Y. I. (2011). Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochemistry (Moscow)*, *76*(1), 131–146. http://doi.org/10.1134/S0006297911010135

Laffleur, B., Denis-Lagache, N., Péron, S., Sirac, C., Moreau, J., & Cogné, M. (2014). AID-induced remodeling of immunoglobulin genes and B cell fate. *Oncotarget*, *5*(5), 1118–1131. http://doi.org/10.18632/oncotarget.1546

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., … Snyder, M. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research*, *22*(9), 1813–1831. http://doi.org/10.1101/gr.136184.111

Lawrence, M., Gentleman, R., & Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*, *25*(14), 1841–1842. http://doi.org/10.1093/bioinformatics/btp328

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., … Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*, *9*(8), e1003118. http://doi.org/10.1371/journal.pcbi.1003118

Lee, Q. Y., Mall, M., Chanda, S., Zhou, B., Sharma, K. S., Schaukowitch, K., … Wernig, M. (2020). Pro-neuronal activity of Myod1 due to promiscuous binding to neuronal genes.

Nature Cell Biology, *22*(4), 401–411. http://doi.org/10.1038/s41556-020-0490-3

Li, S., Garrett-Bakelman, F. E., Akalin, A., Zumbo, P., Levine, R., To, B. L., … Mason, C. E. (2013). An optimized algorithm for detecting and annotating regional differential methylation. *BMC Bioinformatics*, *14*(Suppl 5), S10. http://doi.org/10.1186/1471-2105-14-S5-S10

Liao, W., Hong, S.-H., Chan, B. H.-J., Rudolph, F. B., Clark, S. C., & Chan, L. (1999). APOBEC-2, a Cardiac- and Skeletal Muscle-Specific Member of the Cytidine Deaminase Supergene Family. *Biochemical and Biophysical Research Communications*, *260*(2), 398–404. http://doi.org/10.1006/bbrc.1999.0925

Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, *41*(10), e108–e108. http://doi.org/10.1093/nar/gkt214

Liu, G., Zhang, J., Larsen, B., Stark, C., Breitkreutz, A., Lin, Z.-Y., … Gingras, A.-C. (2010). ProHits: integrated software for mass spectrometry–based interaction proteomics. *Nature Biotechnology*, *28*(10), 1015–1017. http://doi.org/10.1038/nbt1010-1015

Lohr, J. G., Stojanov, P., Lawrence, M. S., Auclair, D., Chapuy, B., Sougnez, C., … Golub, T. R. (2012). Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(10), 3879–84. http://doi.org/10.1073/pnas.1121343109

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. http://doi.org/10.1186/s13059-014-0550-8

Macbeth, M. R., Schubert, H. L., Vandemark, A. P., Lingam, A. T., Hill, C. P., & Bass, B. L. (2005). Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science (New York, N.Y.)*, *309*(5740), 1534–9. http://doi.org/10.1126/science.1113150

Mall, M., Kareta, M. S., Chanda, S., Ahlenius, H., Perotti, N., Zhou, B., … Wernig, M. (2017). Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature*, *544*(7649), 245–249. http://doi.org/10.1038/nature21722

Mariani, L., Weinand, K., Vedenko, A., Barrera, L. A., & Bulyk, M. L. (2017). Identification of Human Lineage-Specific Transcriptional Coregulators Enabled by a Glossary of Binding Modules and Tunable Genomic Backgrounds. *Cell Systems*, *5*(3), 187-201.e7. http://doi.org/10.1016/j.cels.2017.06.015

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., & Thomas, P. D. (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols*, *14*(3), 703–721. http://doi.org/10.1038/s41596-019-0128-8

Mikl, M. C., Watt, I. N., Lu, M., Reik, W., Davies, S. L., Neuberger, M. S., & Rada, C. (2005). Mice Deficient in APOBEC2 and APOBEC3. *Molecular and Cellular Biology*, *25*(16), 7270–7277. http://doi.org/10.1128/MCB.25.16.7270-7277.2005

Moffat, J., Grueneberg, D. A., Yang, X., Kim, S. Y., Kloepfer, A. M., Hinkle, G., … Root, D. E. (2006). A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. *Cell*, *124*(6), 1283–1298. http://doi.org/10.1016/j.cell.2006.01.040

Morris, J. H., Apeltsin, L., Newman, A. M., Baumbach, J., Wittkop, T., Su, G., … Ferrin, T. E. (2011). clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics*, *12*(1), 436. http://doi.org/10.1186/1471-2105-12-436

Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T., … Kohli, R. M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature Chemical Biology*, *8*(9), 751–758. http://doi.org/10.1038/nchembio.1042

Nesvizhskii, A. I., Keller, A., Kolker, E., & Aebersold, R. (2003). A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Analytical Chemistry*, *75*(17), 4646–4658. http://doi.org/10.1021/ac0341261

Oki, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., … Meno, C. (2018). ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Reports*, *19*(12), 1–10. http://doi.org/10.15252/embr.201846255

Okuyama, S., Marusawa, H., Matsumoto, T., Ueda, Y., Matsumoto, Y., Endo, Y., … Chiba, T. (2012). Excessive activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) contributes to liver and lung tumorigenesis. *International Journal of Cancer*, *130*(6), 1294–301. http://doi.org/10.1002/ijc.26114

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., … Hermjakob, H. (2014). The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, *42*(D1), D358–D363. http://doi.org/10.1093/nar/gkt1115

Orimo, A., Gupta, P. B., Sgroi, D. C., Arenzana-Seisdedos, F., Delaunay, T., Naeem, R., … Weinberg, R. A. (2005). Stromal Fibroblasts Present in Invasive Human Breast Carcinomas Promote Tumor Growth and Angiogenesis through Elevated SDF-1/CXCL12 Secretion. *Cell*, *121*(3), 335–348. http://doi.org/10.1016/j.cell.2005.02.034

Patenaude, A. M., Orthwein, A., Hu, Y., Campo, V. A., Kavli, B., Buschiazzo, A., & Di Noia, J. M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nature Structural and Molecular Biology*, *16*(5), 517–527. http://doi.org/10.1038/nsmb.1598

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. http://doi.org/10.1038/nmeth.4197

Perissi, V., Jepsen, K., Glass, C. K., & Rosenfeld, M. G. (2010). Deconstructing repression: evolving models of co-repressor action. *Nature Reviews Genetics*, *11*(2), 109–123. http://doi.org/10.1038/nrg2736

Porter, E. G., Connelly, K. E., & Dykhuizen, E. C. (2017). Sequential Salt Extractions for the Analysis of Bulk Chromatin Binding Properties of Chromatin Modifying Complexes. *Journal of Visualized Experiments*, *2017*(128). http://doi.org/10.3791/55369

Powell, C., Cornblath, E., & Goldman, D. (2014). Zinc-binding Domain-dependent, Deaminase-independent Actions of Apolipoprotein B mRNA-editing Enzyme, Catalytic Polypeptide 2 (Apobec2), Mediate Its Effect on Zebrafish Retina Regeneration. *Journal of Biological Chemistry*, *289*(42), 28924–28941. http://doi.org/10.1074/jbc.M114.603043

Powell, C., Elsaeidi, F., & Goldman, D. (2012). Injury-Dependent Muller Glia and Ganglion Cell Reprogramming during Tissue Regeneration Requires Apobec2a and Apobec2b. *Journal of Neuroscience*, *32*(3), 1096–1109. http://doi.org/10.1523/JNEUROSCI.5603-11.2012

Powell, C., Grant, A. R., Cornblath, E., & Goldman, D. (2013). Analysis of DNA methylation reveals a partial reprogramming of the Muller glia genome during retina regeneration. *Proceedings of the National Academy of Sciences*, *110*(49), 19814–19819. http://doi.org/10.1073/pnas.1312009110

Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., & Cairns, B. R. (2008). DNA Demethylation in Zebrafish Involves the Coupling of a Deaminase, a Glycosylase, and Gadd45. *Cell*, *135*(7), 1201–1212. http://doi.org/10.1016/j.cell.2008.11.042

Rayon-Estrada, V., Harjanto, D., Hamilton, C. E., Berchiche, Y. A., Gantman, E. C., Sakmar, T. P., … Papavasiliou, F. N. (2017). Epitranscriptomic profiling across cell types reveals associations between APOBEC1-mediated RNA editing, gene expression outcomes, and cellular function. *Proceedings of the National Academy of Sciences*, *114*(50), 13296–13301. http://doi.org/10.1073/pnas.1714227114

Razick, S., Magklaras, G., & Donaldson, I. M. (2008). iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinformatics*, *9*(1), 405. http://doi.org/10.1186/1471-2105-9-405

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., … Young, R. A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science (New York, N.Y.)*, *361*(6400), eaar3958. http://doi.org/10.1126/science.aar3958

Salter, J. D., Bennett, R. P., & Smith, H. C. (2016). The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends in Biochemical Sciences*, *41*(7), 578–94. http://doi.org/10.1016/j.tibs.2016.05.001

Salter, J. D., & Smith, H. C. (2018). Modeling the Embrace of a Mutator: APOBEC Selection of Nucleic Acid Ligands. *Trends in Biochemical Sciences*, *43*(8), 606–622. http://doi.org/10.1016/j.tibs.2018.04.013

Sato, Y., Probst, H. C., Tatsumi, R., Ikeuchi, Y., Neuberger, M. S., & Rada, C. (2010). Deficiency in APOBEC2 Leads to a Shift in Muscle Fiber Type, Diminished Body Mass, and Myopathy. *Journal of Biological Chemistry*, *285*(10), 7111–7118. http://doi.org/10.1074/jbc.M109.052977

Shannon, P. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498–2504. http://doi.org/10.1101/gr.1239303

Shi, K., Carpenter, M. A., Banerjee, S., Shaban, N. M., Kurahashi, K., Salamango, D. J., … Aihara, H. (2017). Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nature Structural & Molecular Biology*, *24*(2), 131–139. http://doi.org/10.1038/nsmb.3344

Shteynberg, D., Deutsch, E. W., Lam, H., Eng, J. K., Sun, Z., Tasman, N., … Nesvizhskii, A. I. (2011). iProphet: Multi-level Integrative Analysis of Shotgun Proteomic Data Improves Peptide and Protein Identification Rates and Error Estimates. *Molecular & Cellular Proteomics*, *10*(12), M111.007690. http://doi.org/10.1074/mcp.M111.007690

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, *4*, 1521. http://doi.org/10.12688/f1000research.7563.1

Su, X., Kameoka, S., Lentz, S., & Majumder, S. (2005). Activation of REST/NRSF Target Genes in Neural Stem Cells Is Sufficient To Cause Neuronal Differentiation. *Molecular and Cellular Biology*, *25*(7), 2872–2872. http://doi.org/10.1128/MCB.25.7.2872.2005

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. http://doi.org/10.1073/pnas.0506580102

Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., Gingras, A.-C., & Choi, H. (2014). SAINTexpress: Improvements and additional features in Significance Analysis of INTeractome software. *Journal of Proteomics*, *100*, 37–43. http://doi.org/10.1016/j.jprot.2013.10.023

Vonica, A., Rosa, A., Arduini, B. L., & Brivanlou, A. H. (2011). APOBEC2, a selective inhibitor of TGFβ signaling, regulates left-right axis specification during early embryogenesis. *Developmental Biology*, *350*(1), 13–23. http://doi.org/10.1016/j.ydbio.2010.09.016

Watanabe, Y., Kameoka, S., Gopalakrishnan, V., Aldape, K. D., Pan, Z. Z., Lang, F. F., & Majumder, S. (2004). Conversion of myoblasts to physiologically active neuronal phenotype. *Genes and Development*, *18*(8), 889–900. http://doi.org/10.1101/gad.1179004

Yaffe, D., & Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*, *270*(5639), 725–7. http://doi.org/10.1038/270725a0

Yang, Y., Fear, J., Hu, J., Haecker, I., Zhou, L., Renne, R., … McIntyre, L. M. (2014). LEVERAGING BIOLOGICAL REPLICATES TO IMPROVE ANALYSIS IN CHIP-SEQ EXPERIMENTS. *Computational and Structural Biotechnology Journal*, *9*(13), e201401002. http://doi.org/10.5936/csbj.201401002

Yu, G., Wang, L.-G., & He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, *31*(14), 2382–2383. http://doi.org/10.1093/bioinformatics/btv145

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., … Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137.

http://doi.org/10.1186/gb-2008-9-9-r137
Zhu, L. J. (2013). Integrative Analysis of ChIP-Chip and ChIP-Seq Dataset. In *Methods in Molecular Biology* (pp. 105–124). Humana Press. http://doi.org/10.1007/978-1-62703-607-8_8
Zhu, L. J., Gazin, C., Lawson, N. D., Pagès, H., Lin, S. M., Lapointe, D. S., & Green, M. R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, *11*(1), 237. http://doi.org/10.1186/1471-2105-11-237

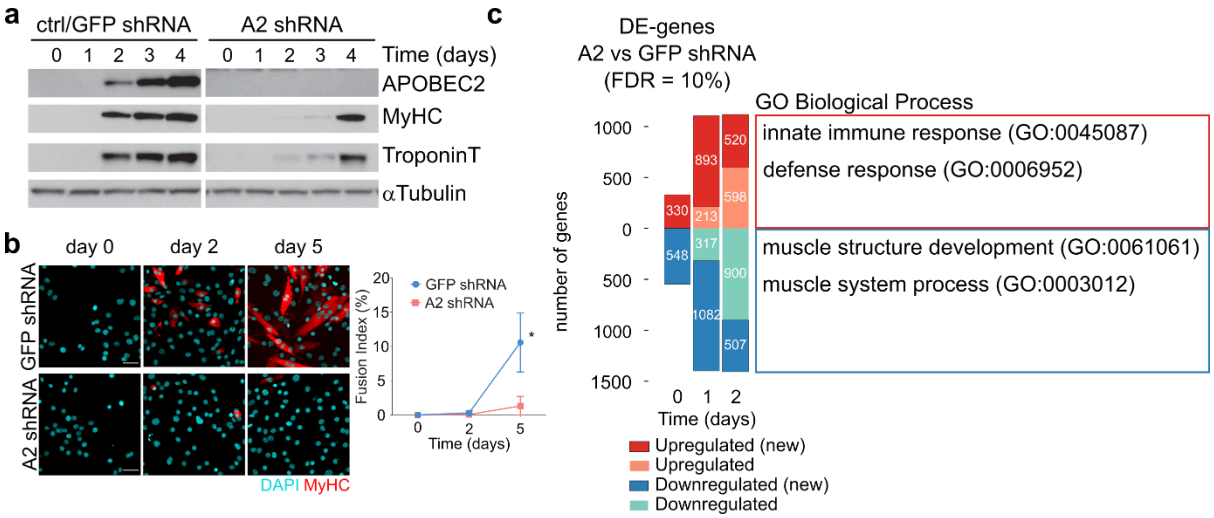


Figure 1: APOBEC2 expression is required for myoblast differentiation, and loss of APOBEC2 leads to gene expression changes.

a) C2C12 lysates in DM at day 0 to 4 were analyzed by Western blot. MyHC (myosin heavy chain) and TroponinT were used as markers of late differentiation; αTubulin, as loading control.

b) C2C12s were fixed and immunostained using an antibody specific to MyHC (red), DAPI (cyan) was used to stain for DNA. Scale bar = 50 µm. Line plot shows quantification of fusion index. Statistics: t test. Six fields of view were measured, and data is shown as means. Error bars indicate SD (n=3), *p < 0.05.

c) Number of differentially expressed genes (DE-genes) between APOBEC2 knockdown (A2 shRNA) relative to GFP shRNA control. Colors indicate up- (red) and down-regulated (blue) genes. A false discovery rate (FDR) cutoff of 10% was used to determine DE-genes. Dark red and blue indicate newly differentially expressed genes at a given time point. Common GO Biological Process terms across day 0 to 2 from statistical enrichment test ranking genes by log2FoldChange (see Supplementary File 1 for complete output tables of the tests).

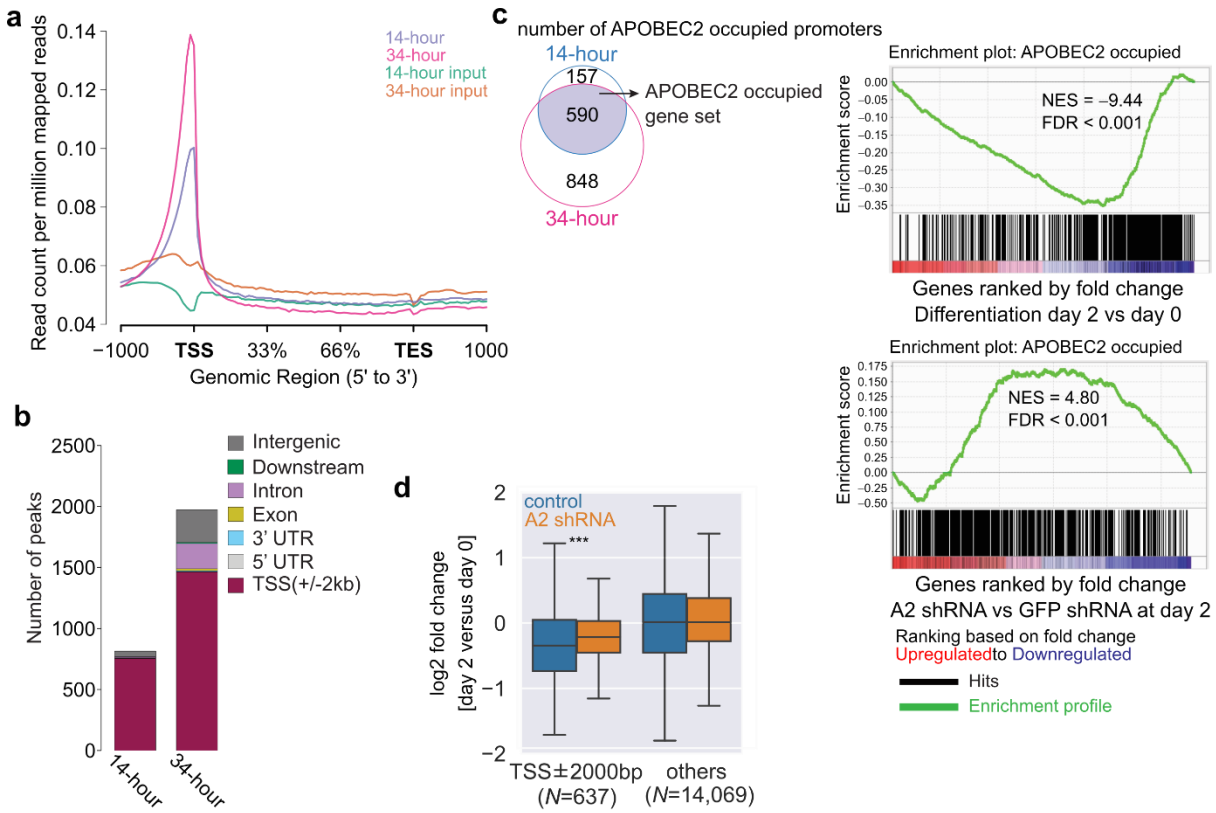


Figure 2. APOBEC2 ChIPSeq in differentiating C2C12 myoblasts.

a) The mean normalized APOBEC2 signal (plotted as read counts per million mapped reads) across all annotated genes. This plot shows the global differences in APOBEC2 binding between the two time points in TSS. Both time points are in biological triplicates.

b) Genomic annotations of APOBEC2 consensus binding regions in each of the time points (14- and 34-hour). Binding regions are annotated based on genomic feature. The priority of assignment to a genomic feature when there is annotation overlap is: Promoter (2kb around the TSS), 5' UTR, 3' UTR, Exon, Intron, Downstream (within 3kb downstream of the end of the gene).

c) The Venn diagram represents the number of genes that have APOBEC2 occupancy on their promoters at 14- and 34-hour time points. The genes that show consistent APOBEC2 occupancy in their promoters at both time points were used to create an APOBEC2_OCCUPIED gene set as shown in the Venn diagram. Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was used to test the enrichment of the APOBEC2_OCCUPIED gene set in the list of genes that are differentially expressed through differentiation (top right graph) or the ones that are differentially expressed due to APOBEC2 knockdown at day 2 (bottom right graph). The enrichment profile over the whole ranked gene set is shown in green with normalized enrichment score (NES) and FDR values. Gene hits are shown as black lines. A positive NES indicates gene set enrichment at the top (positive/up fold change) of the ranked list; a negative NES indicates gene set enrichment at the bottom (negative/down fold change) of the ranked list.

d) Expression changes for genes in control (GFP control shRNA) and A2 knockdown (A2 shRNA) C2C12 during differentiation. Genes are grouped by the presence of an APOBEC2

ChIP-seq peak nearby Transcription Start Sites (TSS) at not more than 2000bp from it, or genome wide background (others). Asterisks indicate adjusted P values, by two-sided t-test corrected using Benjamini Hochberg procedure (***=0.001).

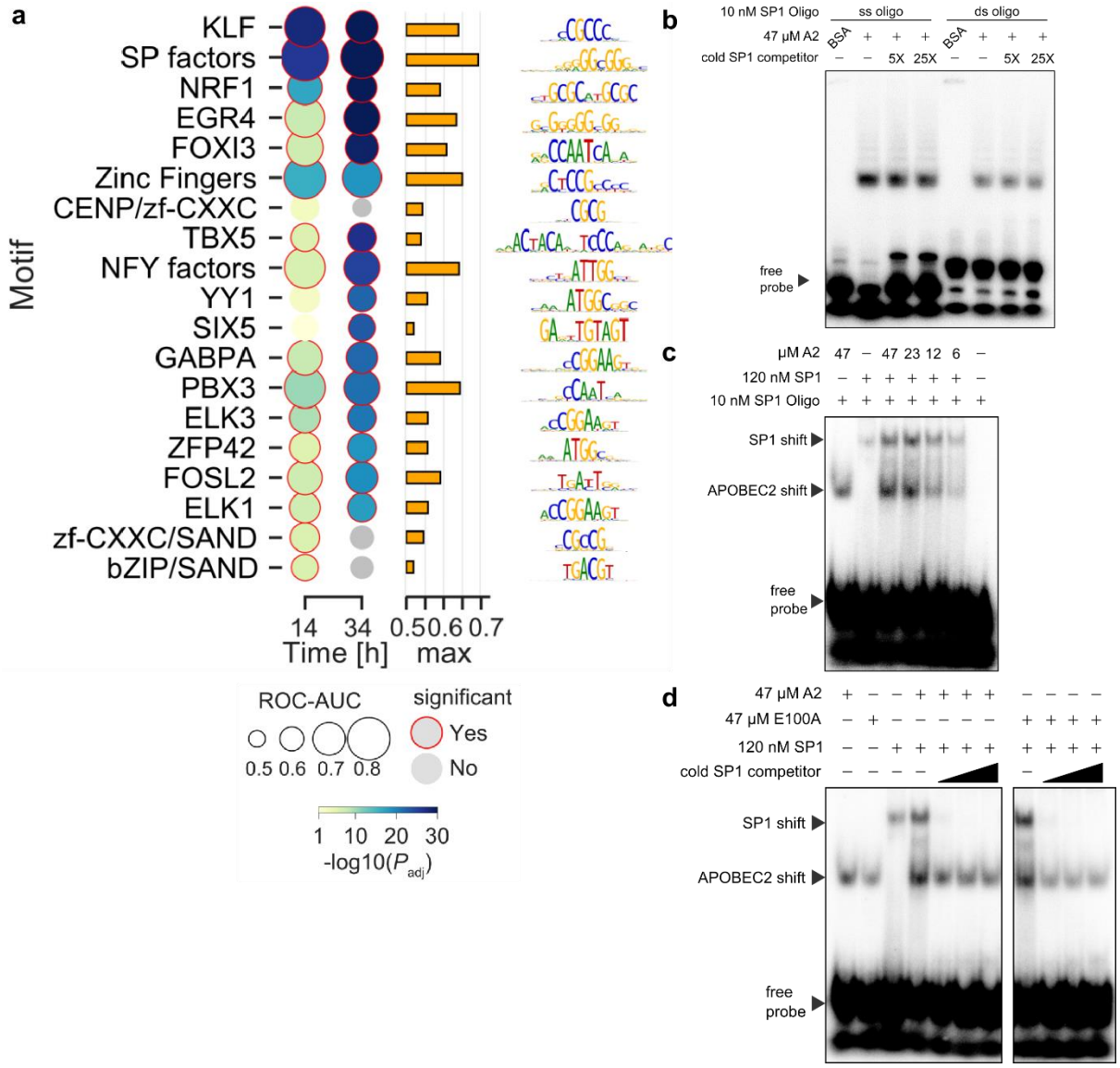


Figure 3. APOBEC2 binding at specific transcription factor DNA motifs.

a) Enrichment of *8*-mers associated to 19 known transcription factors specificities (*motifs*) in APOBEC2 ChIP-seq regions, compared against negative genomic regions (Methods). Receiver Operating Characteristic Area under the Curve (ROC-AUC, circle sizes) values are descendently sorted on the y-axis by minimum adjusted *P* value observed in both time points (color). All 18 shown transcription factors are significant on at least one time point (adjusted *P* value < 0.1, using one-sided Wilcoxon rank sum test and Benjamini-Hochberg correction. Red circles). Right barplot indicates maximum ROC-AUC value, row-wise. Right sequence logos depict unsupervised alignment of all *8*-mers referred to that transcription factor motif (Mariani et al., 2017).

b) Electrophoretic mobility shift assay (EMSA) of recombinant APOBEC2 on either ss or ds SP1 DNA probe competed with corresponding cold (unlabeled) probe at 5X or 25X excess.

c) EMSA of recombinant APOBEC2 on a SP1 DNA motif with increasing amounts of recombinant APOBEC2.

d) EMSA of recombinant APOBEC2 on a SP1 DNA motif with cold competitor and mutant APOBEC2 E100A (E100A). Cold SP1 competitor lanes include an excess of unlabeled SP1 probe: 5X, 25X, and 100X excess.

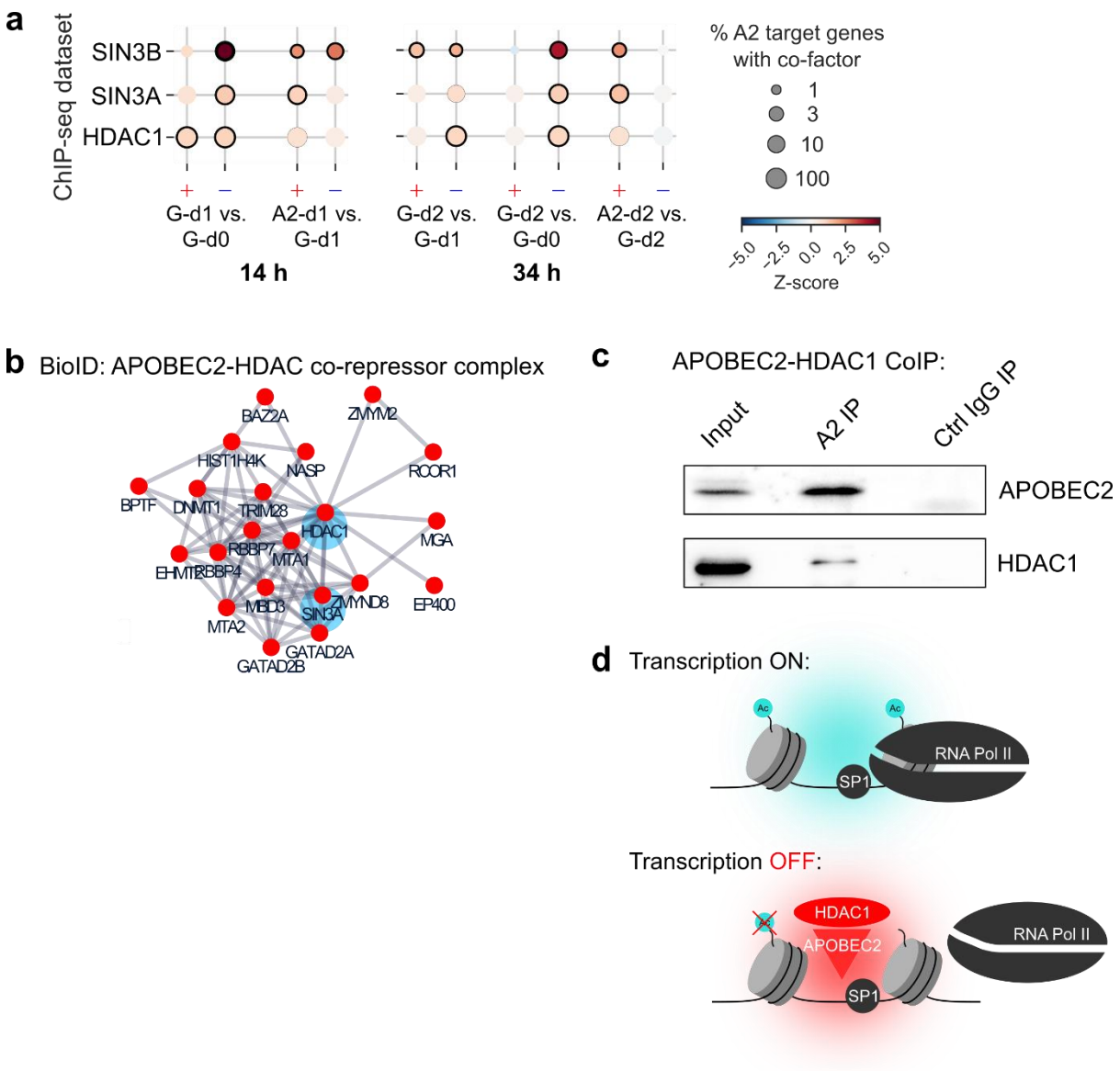


Figure 4. APOBEC2 recruits the HDAC1 co-transcriptional repressor complex.

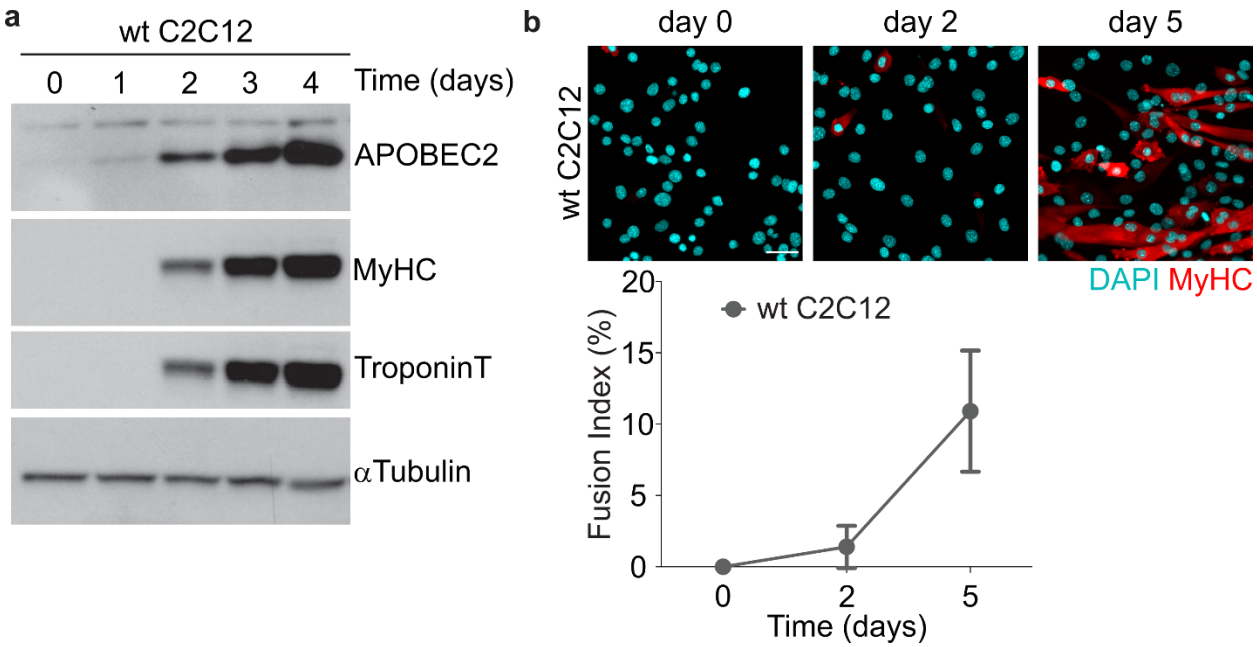
a) Enrichment of APOBEC2 target genes with differential expression for HDAC1 complex ChIP-seq peaks. Circles sizes indicate fraction of differentially up- (+) or down- (-) regulated genes with ChIP-seq peaks for SIN3B, SIN3A and HDAC1 (for all factors see Supplementary Figure 7). Z-scores are calculated by comparing the proportion of differentially expressed genes with peaks for a factor between APOBEC2 target genes and all other genes. (*left*) 14 h = APOBEC2 target genes at 14-hour ChIP-Seq time point (three replicates) and (*right*) 34 h = APOBEC2 target genes at 34-hour time point (three replicates). Differential gene expression comparisons are labels for GFP shRNA samples (G) or APOBEC2 shRNA (A2) at days 0, 1, and 2 (d0, d1 and d2, respectively). Circles are bold when observed Z-scores are greater than 1.0.

b) HDAC co-repressor complex cluster identified from protein-protein interactions of APOBEC2 taken from proximity-dependent protein biotinylation (BioID). Each red node corresponds to a protein that was identified through BioID mass spectrometry to interact with APOBEC2. The grey edges denote the known interactions of these proteins with each other. SIN3A and HDAC1enriched at APOBEC2 occupied genes are highlighted with a blue circle

(see **a**). This is one cluster among several identified clusters from the total 361 identified proteins APOBEC interacts with (see Supplementary Table 2).

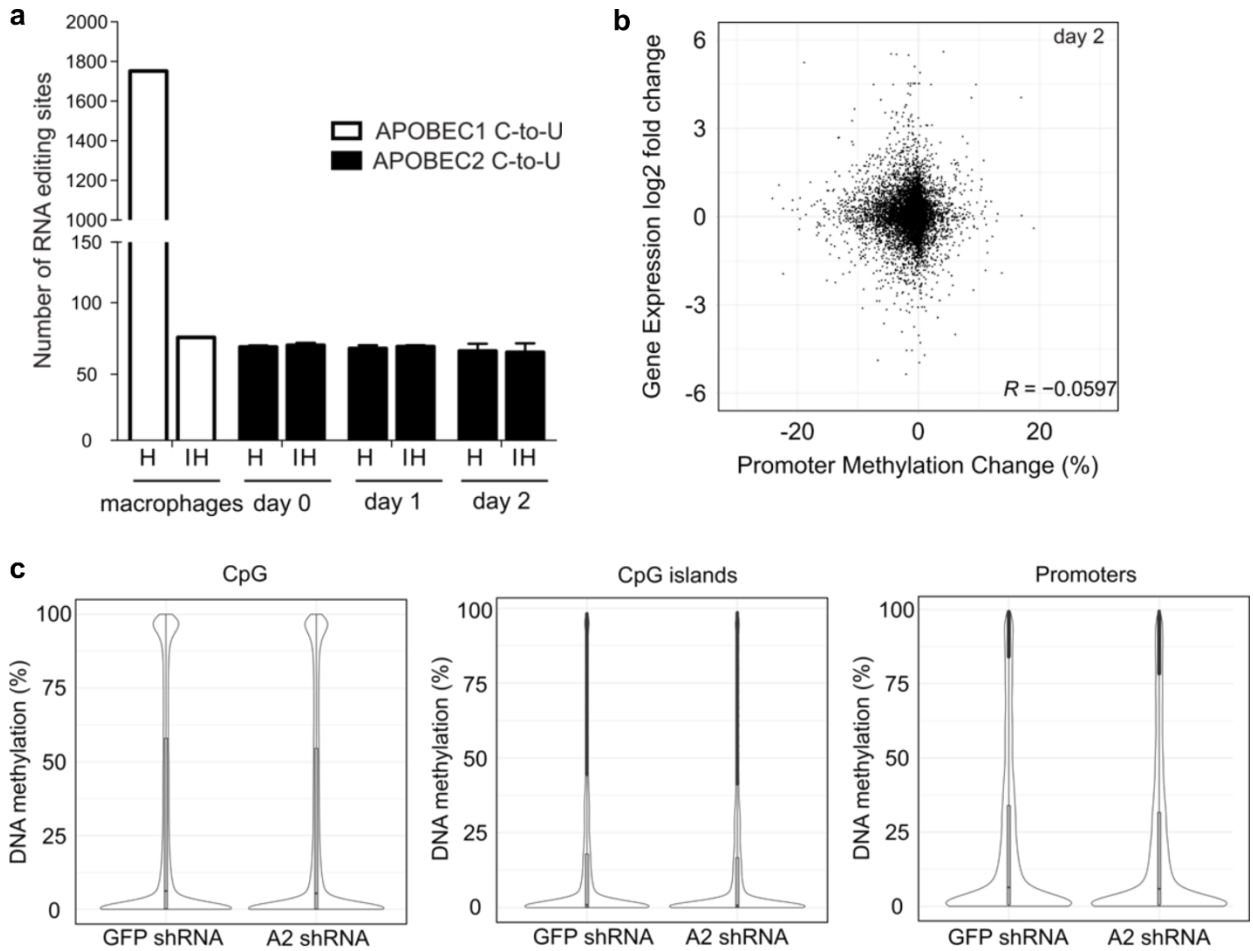
c) Co-immunoprecipitation (Co-IP) of APOBEC2 with HDAC1 in C2C12 myoblasts differentiated to myotubes for 4 days. Nuclear protein lysates (Input) were incubated with beads conjugated to either APOBEC2 antibody (A2 IP) or IgG isotype control antibody (Ctrl IgG IP). Proteins were then eluted, ran on an SDS-PAGE gel, and blotted with APOBEC2 or HDAC1 antibodies.

d) Proposed molecular function of APOBEC2 as a co-transcriptional repressor complex that acts on active/open chromatin to repress transcription through HDAC1 histone deacetylation during myoblast differentiation.



Supplementary Figure 1:
a) Whole cell extracts of mouse wildtype (wt) C2C12 myoblasts and myotubes were analyzed by Western blotting using anti-APOBEC2 antibodies. MyHC and TroponinT were used as markers of late differentiation, alpha-tubulin was used as loading control.

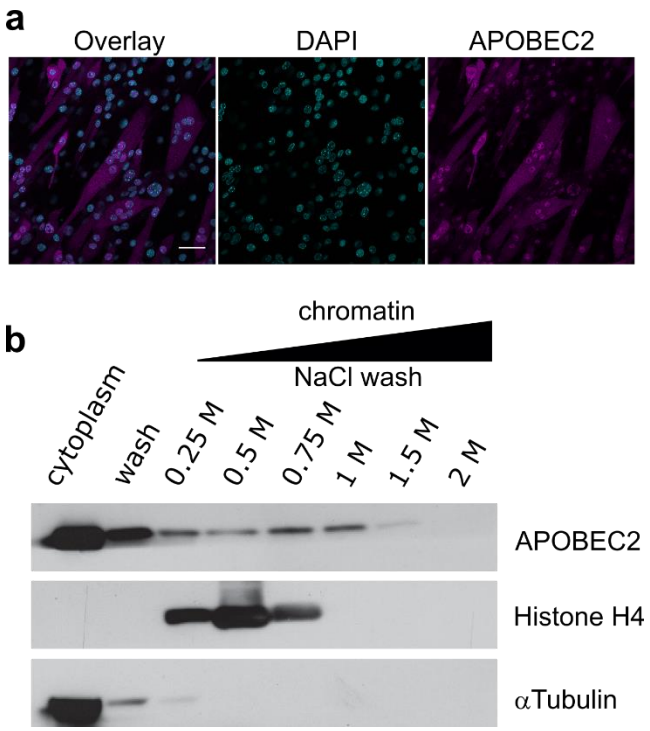
b) C2C12 cells were cultured in differentiation medium for 0, 2 and 5 days, fixed, and stained with antibody to MyHC (red). Nuclei were visualized by DAPI staining (blue). Below the quantification of differentiation expressed as fusion index, which is the percentage MyHC-positive myotubes with >2 nuclei. Results are presented as means from quantification of at least 6 images/sample. Error bars indicate SD. Scale bar = 50 µm.



Supplementary Figure 2:
a) Candidate C-to-U RNA editing sites called from APOBEC2 knockdown samples, control (GFPsh) at day 0, 1, and 2 in DM in C2C12s and wild-type and APOBEC1-/- macrophages (positive control). Hits (H) represent candidate editing sites present in control (GFPsh) but not in APOBEC2 knockdown dataset (in positive control hits are the # of sites in APOBEC1-/- dataset.) Inverse hits (IH) represent putative edited sites yielded when the inverse comparison is made, thus edit sites present in the APOBEC2 knockdown dataset but not in the control (GFPsh) (for the positive control edit sites that are present in APOBEC1-/- dataset but not in wild- type). Data are represented as means ± SD using outputs of 3 RNA-Seq datasets.

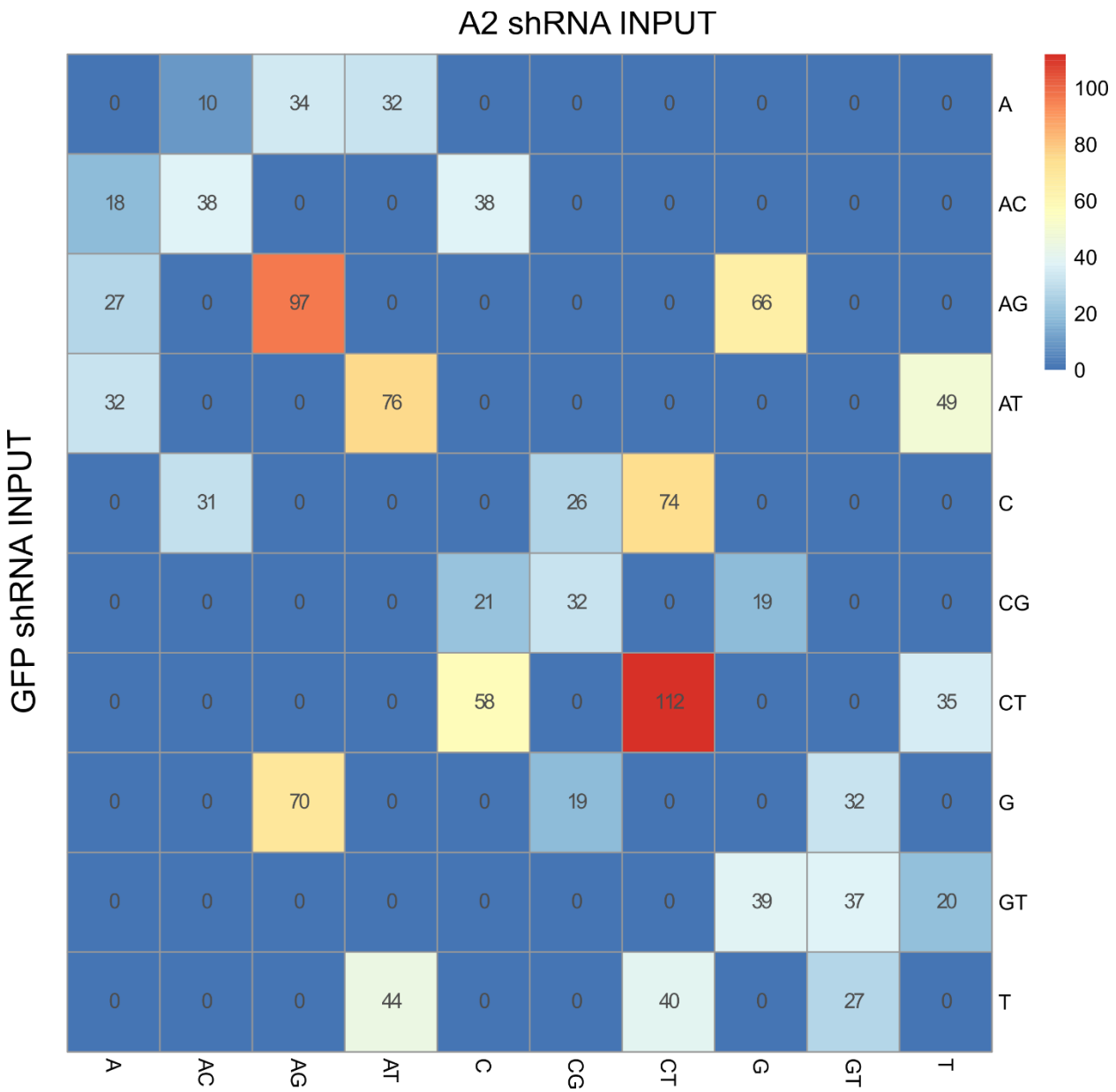
b) Methylation changes across all the represented promoters in the ERRBS dataset compared with the expression changes of the same genes in RNA-Seq dataset. Shown here are datasets from the day 2 timepoint. R = Pearson's correlation coefficient.

c) Distribution of DNA methylation frequencies in C2C12s as determined by ERRBS for individual CpGs, CpG islands and promoters. Promoters are defined at -/+ 2Kb around the TSS in Ensemble annotations. CpG islands were taken from the cpgIslandExt track of the UCSC table browser. Violin plots represent the distribution of DNA methylation frequencies for each feature. Median and first and third quartiles are shown with the box plots.



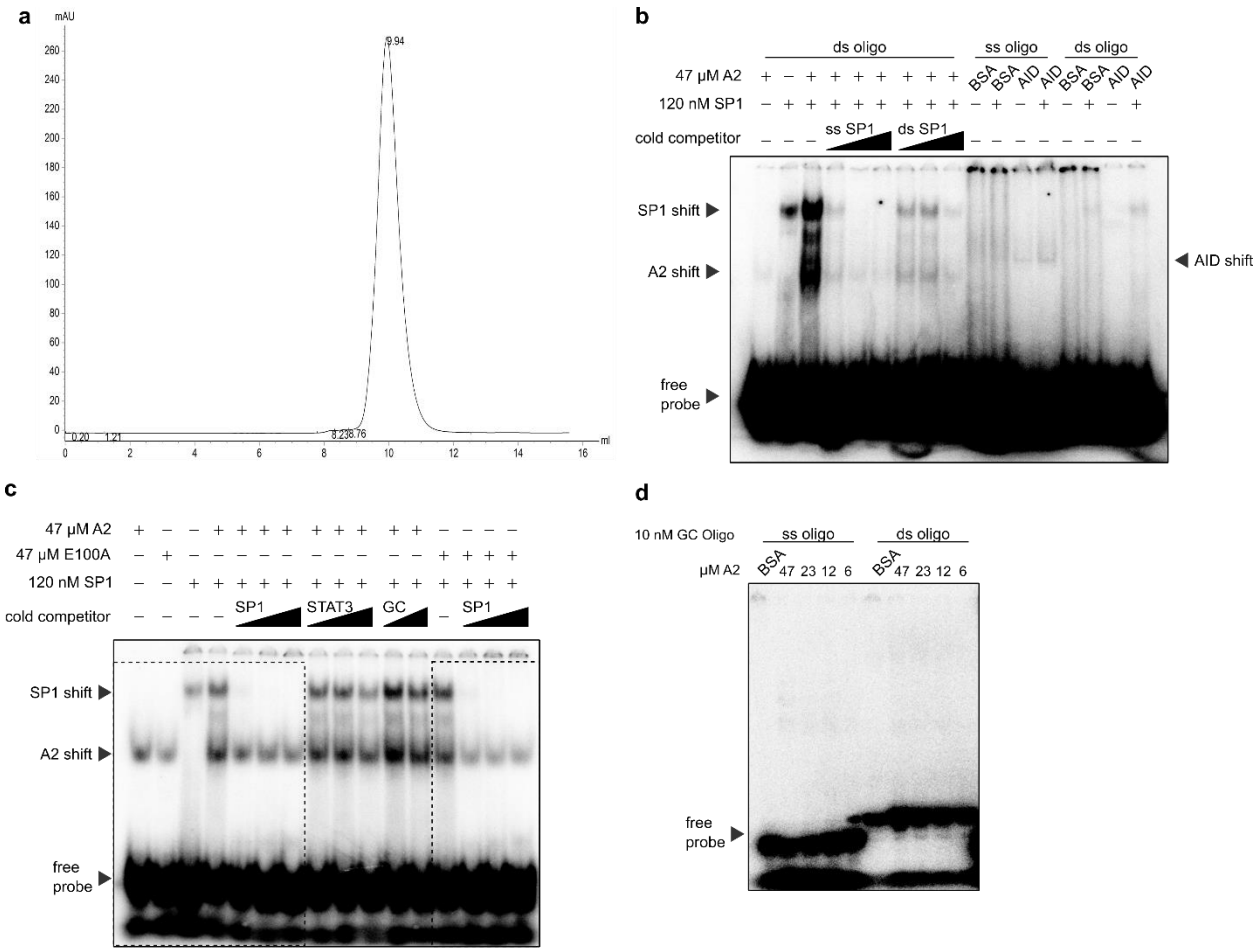
Supplementary Figure 3:
a) Immunostaining of APOBEC2 (magenta) and DAPI-positive (blue) nuclei in C2C12 cultured for 5 days in DM. Scale bar = 50 µm.

b) The NaCl-elution profiles of endogenous Apobec2 and histone H4 are shown. alpha-tubulin is a cytoplasmic marker. The amount of indicated proteins in eluates was measured by Western blotting.



Supplementary Figure 4: DNA Editing in APOBEC2 knockdown versus control (GFP shRNA)

Pairwise heatmap - This is a head-to-head comparison of variant positions between the 2 input sample sets. The symmetry of the heatmap indicates that there is no preference for a unidirectional base substitution process.



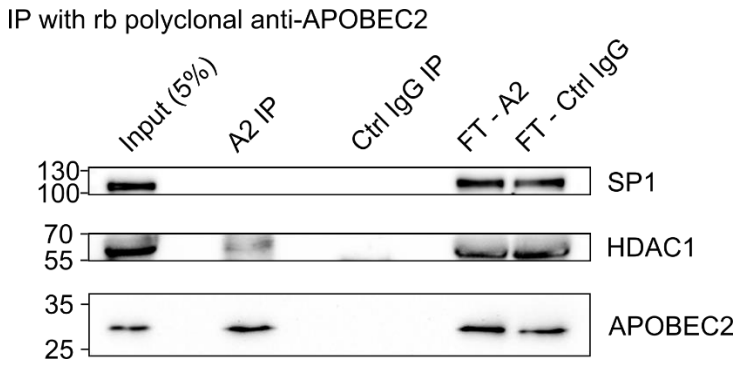
Supplementary Figure 5: APOBEC2 EMSA

a) Fast protein liquid chromatography (FPLC) profile (Superdex 75 10/300 GL, GE Healthcare) of recombinant 6x-His-APOBEC2 produced in High Five (Thermo) insect cells.

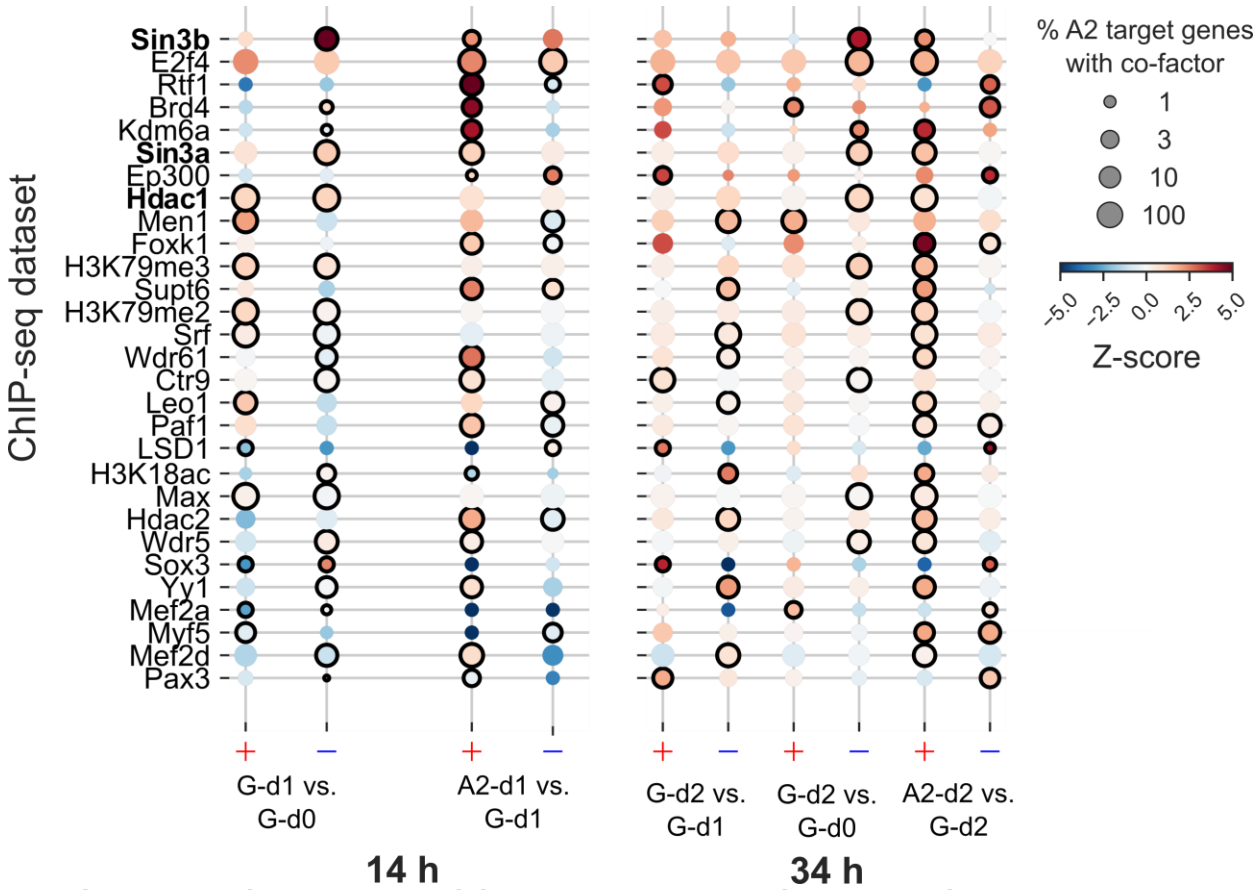
b) Lanes 1 to 9**:** EMSA using purified recombinant APOBEC2 and SP1 on 10 nM SP1 oligo probe. Binding on labeled ds SP1 probe was competed with excess unlabeled ss SP1 or ds SP1 probe corresponding to 5X, 25X and, 100X excess. Lanes 10 to 17: EMSA using purified recombinant SP1 and AID on 10 nM ss or ds SP1 probe.

c) Boxed in dashed lines (lanes 1 to 7; 13 to 16) are presented in Figure 3d. Lanes 8 to 12: EMSA using purified recombinant APOBEC2 and SP1 on 10 nM SP1 oligo probe competed with either ds STAT3 probe (5X, 25X and 100X excess) and GC-rich motif (GC) probe (5X and 25X excess).

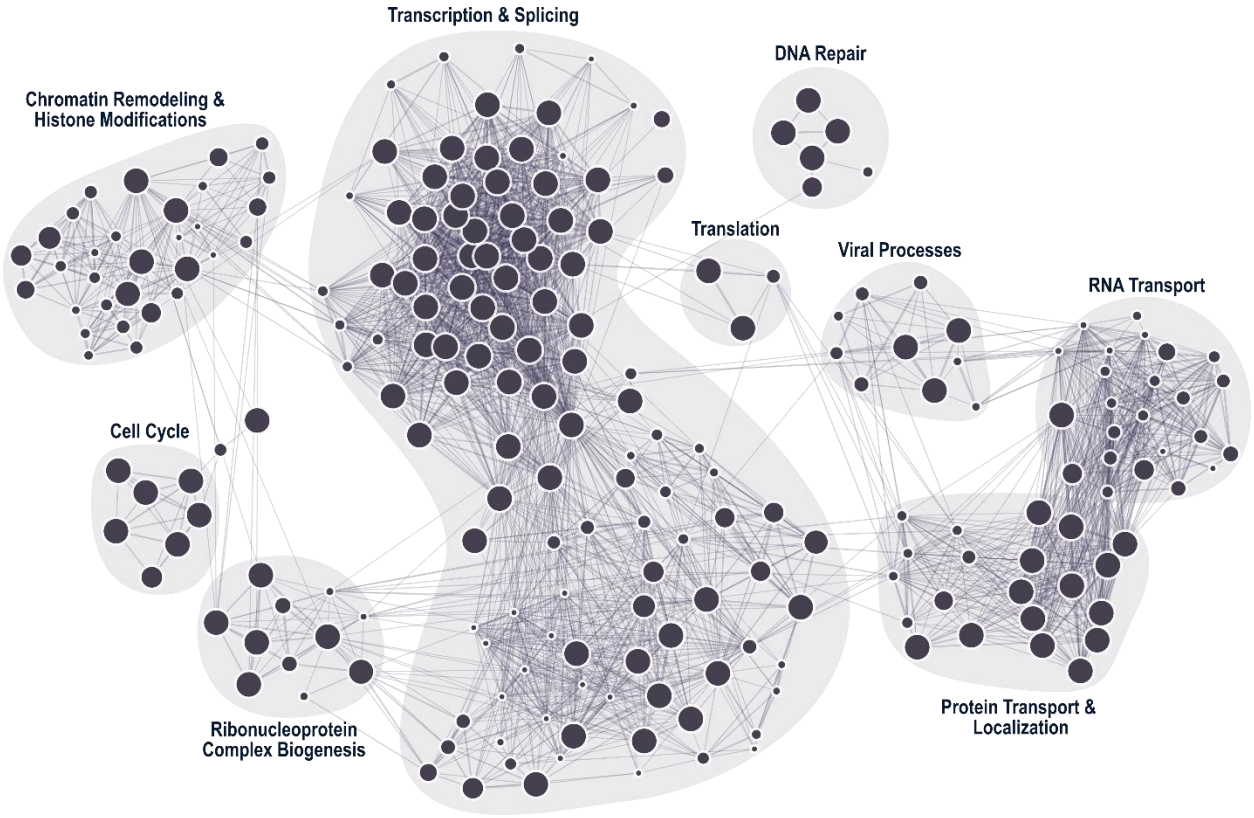
d) EMSA using purified recombinant APOBEC2 on either 10 nM ss or ds GC probe with decreasing amounts of APOBEC2, 47 to 6 µM, and BSA alone.



Supplementary Figure 6: APOBEC2-SP1 CoIP
Nuclear lysate from C2C12 cells differentiated for 4 days was applied to beads conjugated with anti-APOBEC2 antibody. Input represents 5% of amount used for CoIP. A2 IP (APOBEC2 antibody beads) or Ctrl IgG (isotype control antibody beads) represents eluate from the respective CoIP. Immunoblots were stained with SP1, HDAC1, and APOBEC2 antibodies. FT: flow through from beads representing unbound protein.



Supplementary Figure 7: Co-enrichment at APOBEC2 ChIP-seq peaks
Enrichment of APOBEC2 target genes with differential expression for HDAC1 complex ChIP-seq peaks. Circles sizes indicate fraction of differentially up- (+) or down- (-) regulated genes with ChIP-seq peaks for all ChIP-Seq antigens. Z-scores are calculated by comparing of the proportion of differentially expressed genes with peaks for a factor between APOBEC2 target genes and all other genes. (*left*) 14 h = APOBEC2 target genes at 14-hour ChIP-Seq time point (three replicates) (*right*) 34 h = APOBEC2 target genes at 34-hour time point (three replicates). Differential gene expression comparisons are labels for GFP shRNA samples (G) or APOBEC2 shRNA (A2) at days 0, 1, and 2 (d0, d1 and d2, respectively). Circles are bold when observed Z-scores are greater than 1.0.



Supplementary Figure 8: BioID GO Enrichment Map
Gene ontology enrichment map based on APOBEC2 BioID hits. Gene ontology terms are clustered by molecular function.