

Discovery of the Transcriptional Regulatory Role
of the Cytidine Deaminase
APOBEC2
in Skeletal Muscle Differentiation

Jose Paulo Lorenzo

Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
Jose Paulo Lorenzo, MSc, MRes
Born in Mandaluyong, Metro Manila, Philippines
Oral examination: 3 December 2021

Discovery of the Transcriptional Regulatory Role
of the Cytidine Deaminase
APOBEC2
in Skeletal Muscle Differentiation

Referees:

Prof. Dr. Frank Lyko

Prof. Dr. Nina Papavasiliou

Abstract

APOBEC2 is a member of the activation induced cytidine deaminase/apolipoprotein B editing complex (AID/APOBEC) family of nucleic acid deaminases. Even though it has the conserved zinc-dependent deaminase domain, it neither has an established cytidine deaminase activity nor a known nucleic acid ligand. However, APOBEC2 knockout models show defects in skeletal muscle, where APOBEC2 is highly expressed. Interestingly, none of these observations have been attributed to APOBEC2 acting directly on nucleic acids. In this work, I established that APOBEC2 is necessary for myotube differentiation. I showed that APOBEC2 is enriched within the nuclei of these myotubes, where it occupies the regulatory regions of genes related to cell differentiation processes. Moreover, the knockdown of APOBEC2 caused changes in gene expression of its target occupied genes. I then established that APOBEC2 directly interacts with the transcriptional corepressor histone deacetylase 1 (HDAC1), providing a potential mechanism for the observed gene expression changes. Furthermore, I demonstrated that APOBEC2 interacts directly with a single-stranded DNA ligand with an affinity comparable to other APOBEC ligand interactions. Given the evidence that APOBEC2 regulates transcription during myotube differentiation, interacts with the corepressor HDAC1, and binds DNA, I proposed that APOBEC2 is a transcriptional regulator.

Zusammenfassung

APOBEC2 gehört zur Familie der aktivierungsinduzierten Cytidin-Desaminase/Apolipoprotein-B-Editing-Komplexe (AID/APOBEC) der Nukleinsäuredeaminasen. Obwohl es die konservierte Zink-abhängige Desaminase-Domäne besitzt, hat es weder eine etablierte Cytidin-Desaminase-Aktivität noch einen bekannten Nukleinsäure-Ligand. Allerdings zeigen APOBEC2-Knockout-Modelle Defekte in der Skelettmuskulatur, in welcher APOBEC2 stark exprimiert wird. Interessanterweise wurde keine dieser Beobachtungen mit einer direkten Wirkung von APOBEC2 auf Nukleinsäuren in Verbindung gebracht. In dieser Doktorarbeit habe ich nachgewiesen, dass APOBEC2 für die Differenzierung von Myotuben notwendig ist. Darüber hinaus konnte ich zeigen, dass APOBEC2 in den Kernen dieser Myotuben angereichert ist und dort die regulatorischen Regionen von Genen besetzt, die mit Zelldifferenzierungsprozessen zusammenhängen. Außerdem führte die Verminderung der APOBEC2 Anzahl zu Veränderungen in der Genexpression der von ihm besetzten Zielgene. Ich habe dann festgestellt, dass APOBEC2 direkt mit dem transkriptionellen Co-Repressor Histon-Deacetylase 1 (HDAC1) interagiert, was einen möglichen Mechanismus für die beobachteten Veränderungen der Genexpression darstellt. Darüber hinaus konnte ich nachweisen, dass APOBEC2 direkt mit einem Einzelstrang-DNA-Ligand interagiert und zwar mit einer Affinität, die mit anderen APOBEC-Ligand-Interaktionen vergleichbar ist. In Anbetracht der Befunde, dass APOBEC2 die Transkription während der Myotube-Differenzierung reguliert, mit dem Co-Repressor HDAC1 interagiert, und DNA bindet, schlug ich vor, dass APOBEC2 einen Transkriptionsregulator ist.

Table of Contents

Abstract	v
Zusammenfassung	vi
Table of Contents	vii
List of Figures	ix
List of Tables.....	xi
1. Introduction	1
1.1 APOBEC2, a member of the APOBEC family of cytidine deaminases	2
1.1.1 The emergence of the vertebrate-specific APOBEC family.....	3
1.1.2 The APOBEC family and the vertebrate immune system	4
1.1.3 APOBEC roles beyond the immune system	6
1.1.4 Positive and purifying evolutionary pressures on the APOBEC family.....	6
1.2 The physiological role of APOBEC2	8
1.2.1 APOBEC2 loss in animal models leads to muscular dystrophy in skeletal muscle	8
1.2.2 APOBEC2 has roles beyond muscle tissue	9
1.3 APOBEC2 structure suggests an evolutionarily conserved APOBEC function	11
1.3.1 APOBEC2 protein structure bears the conserved cytidine deaminase domain ..	12
1.3.2 The evolutionary conservation of the APOBEC2 protein structure suggests a conserved molecular function	14
1.4 Towards determining the molecular function of APOBEC2	16
2. Aims	17
3. Results	19
3.1 APOBEC2 and myotube differentiation.....	20
3.1.1 RNA-Seq of APOBEC2 knockdown cells pointed to a decrease in skeletal muscle differentiation	21
3.1.2 APOBEC2 knockdown led to a decrease in myotube formation in C2C12 myoblasts	29
3.2 APOBEC2 and transcriptional control	32
3.2.1 APOBEC2 occupied promoters of genes related to cell differentiation.....	32
3.2.2 APOBEC2 occupied genes changed in gene expression during myoblast differentiation.....	36
3.2.3 APOBEC2 regulated non-skeletal muscle-related genes during skeletal muscle differentiation.....	37
3.3 APOBEC2 and corepressor complexes	41
3.3.1 APOBEC2 was found in the chromatin fraction of cells.....	43
3.3.2 APOBEC2 interacted directly with histone deacetylase HDAC1	44
3.3.3 APOBEC2 amino-terminal loss disrupted nuclear localization	47

Table of Contents

3.4 Restoring APOBEC2 in knockdown C2C12 cells	49
3.4.1 Expression of APOBEC2 increased muscle marker expression in APOBEC2 knockdown cells	49
3.4.2 Variants of APOBEC2 failed to restore myotube differentiation.....	51
3.5 Exogenous APOBEC2 expression in a fibroblast cell line	56
3.5.1 Expression of APOBEC2 in NIH3T3 fibroblasts led to downregulation of APOBEC2 occupied genes	56
3.5.2 Expressed APOBEC2 had limited nuclear localization	57
3.6 APOBEC2 nucleic-acid ligand interaction	61
3.6.1 SP/KLF GC-rich motifs were enriched in APOBEC2 occupied genomic regions	61
3.6.2 APOBEC2 directly interacted with specific nucleic acid ligand.....	62
4. Discussion.....	67
4.1 On the role of APOBEC2 in transcriptional regulation	68
4.2 On the ability of APOBEC2 to directly interact with DNA as a transcription factor.....	71
4.3 On the role of APOBEC2 in muscle and beyond.....	76
5. Materials and Methods.....	81
5.1 Cell culture	88
5.2 Production of APOBEC2 knockdown C2C12 myoblasts	88
5.3 Molecular cloning of APOBEC2 overexpression constructs.....	89
5.4 Retrovirus production and transduction of cells	90
5.5 Microscopy – fusion index and APOBEC2 localization.....	90
5.6 Analysis of RNA-Seq data from APOBEC2 knockdown cells	91
5.7 Analysis of ChIP-Seq data from differentiating C2C12 myoblasts	92
5.8 Extracting cytoplasmic and nuclear protein fractions from C2C12 cells.....	94
5.9 Co-immunoprecipitation	95
5.10 Western blotting	96
5.11 Reverse transcription-quantitative PCR (RT-qPCR)	96
5.12 Recombinant mouse APOBEC2 production	98
5.13 Microscale Thermophoresis (MST)	99
6. References.....	101
7. Appendix.....	111
7.1 Appendix I.....	112
7.2 Appendix II	114
7.3 Appendix III	115
7.4 Appendix IV	116
8. Acknowledgments.....	119

List of Figures

Figure 1.1. APOBEC family evolution in the vertebrate lineage.	3
Figure 1.2. APOBEC2 protein structure.	12
Figure 1.3. Protein sequence alignment of mouse and human APOBEC2 with other single domain human APOBEC family members.	14
Figure 3.1. Myotube formation in differentiated C2C12 myoblasts.	21
Figure 3.2. Western blots of APOBEC2 and muscle-specific proteins in knockdown cells.	22
Figure 3.3. RNA-Seq differential expression analysis between APOBEC2 knockdown and control cells at early time points of myotube differentiation.	23
Figure 3.4. Gene ontology analysis of downregulated genes in APOBEC2 knockdown cells during early myoblast differentiation.	26
Figure 3.5. Gene ontology analysis of upregulated genes in APOBEC2 knockdown cells during early myoblast differentiation.	29
Figure 3.6. Myotube formation in differentiating APOBEC2 knockdown C2C12 myoblasts.	30
Figure 3.7. APOBEC2 ChIP-Seq binding sites at 14 and 34 hours into myotube differentiation.	33
Figure 3.8. Differentially bound APOBEC2 binding sites at 14 and 34 hours into C2C12 myotube differentiation.	35
Figure 3.9. Gene expression changes of APOBEC2 occupied genes during myotube differentiation, and with APOBEC2 knockdown during differentiation.	37
Figure 3.10. Gene ontology and heatmap of differentially expressed APOBEC2 target genes.	39
Figure 3.11. APOBEC2 protein interactome from BioID experiment performed in HEK293T cells and human B cells.	42
Figure 3.12. APOBEC2 protein expression within differentiated C2C12 myotube protein lysate fractions.	44
Figure 3.13. Co-immunoprecipitation experiments of APOBEC2 with HDAC1 in the nuclear fraction of differentiated C2C12 myotubes.	46
Figure 3.14. N-terminal truncated APOBEC2, del(1-41)A2, localization in cytoplasmic or nuclear fractions from differentiated C2C12 myotubes.	48

List of Figures

Figure 3.15. APOBEC2 exogenous expression in APOBEC2 knockdown C2C12 cell lines.	50
Figure 3.16. Exogenous expression of APOBEC2 in knockdown cells affects muscle marker protein levels.....	51
Figure 3.17. Exogenous expression of APOBEC2 and variants in C2C12 myoblasts.	53
Figure 3.18. Protein sequence alignment of mouse and human APOBEC2.....	54
Figure 3.19. Exogenous expression of APOBEC2 in NIH3T3 fibroblasts affects APOBEC2 target genes.....	57
Figure 3.20. Exogenous APOBEC2 cytoplasmic and nuclear localization in NIH3T3 cells.	58
Figure 3.21. Transcription factor motifs found within APOBEC2 binding sites in C2C12 myoblasts.	62
Figure 3.22. APOBEC2 and single-stranded DNA ligand affinity measurements.....	64
Figure 4.1 APOBEC2 structure alignment with APOBEC3A and APOBEC3G bound to ssDNA substrate.....	73
Figure 4.2. Proposed model of APOBEC2 transcriptional complex interaction.	75
Figure 4.3. APOBEC2 expression in blood cell lineages. Hierarchical hematopoietic system differentiation tree depicting APOBEC2 expression levels at specific cell populations.....	78
Figure 7.1. Heatmap from the list of differentially expressed APOBEC2 target genes. ...	113
Figure 7.2. Mail from Dr. Fowzan Alkuraya regarding APOBEC2 G64R mutation.	114
Figure 7.3 Effect of exogenous expression of APOBEC2 in C2C12 fibroblasts on APOBEC2 target genes.....	115
Figure 7.4. STREME analysis for motif discovery in APOBEC2 target chromatin regions..	117

List of Tables

Table 1. Counts of cells with APOBEC2 nuclear enrichment	59
Table 2. List of reagents and materials	82
Table 3. List of antibodies.....	84
Table 4. List of plasmids	85
Table 5. List of primers and oligonucleotides.....	86
Table 6. List of equipment	87

1. Introduction

1. Introduction

APOBEC2 is a member of the prolific activation induced cytidine deaminase/apolipoprotein B editing complex (AID/APOBEC or APOBEC) family of cytidine deaminases (Conticello, 2008). APOBEC2 was discovered from its sequence similarity to the RNA editing enzyme APOBEC1 (Liao et al., 1999). APOBEC2 was the first among the APOBEC family to have a structure published, which guided the modeling of the structures of the rest of the members (Prochnow et al., 2007). Even though it contains the conserved zinc-dependent cytidine deaminase domain, it has no confirmed deaminase activity on nucleic acids. As an introduction to this current work, I start by discussing APOBEC2 as a member of the APOBEC family. I then discuss the published works on the putative physiological and molecular function of APOBEC2. Further, I review what is known about its molecular structure. Finally, I discuss my ideas on the molecular function and substrate of APOBEC2 that will be addressed in this work.

1.1 APOBEC2, a member of the APOBEC family of cytidine deaminases

The vertebrate-specific APOBEC family is composed of several subfamilies: AICDA (AID), APOBEC1, APOBEC2, APOBEC3 (and its paralogs), and APOBEC4 (Salter et al., 2016). Humans have seven paralogs of APOBEC3: APOBEC3A-D and 3F-H (Münk et al., 2012). Members of the APOBEC family have diverse roles – immunoglobulin loci mutation in B cell maturation (AID), mRNA editing (APOBEC1), and defense-related viral DNA and retrotransposon editing (APOBEC3 subfamily) (Ratcliff & Simmonds, 2021). Members of the family have a conserved cytidine deaminase fold, composed of a His-X-Glu-X(25-30)-Pro-Cys-X(2-4)-Cys amino acid sequence (Salter & Smith, 2018). This deaminase domain coordinates a zinc ion that catalyzes the nucleophilic attack on the C4 carbon of cytidine leading to deamination (Salter & Smith, 2018). Tracing the evolutionary origins of this conserved deaminase domain provides insights into the molecular function of the APOBEC family and provides threads to follow in unraveling the putative function of APOBEC2.

1.1.1 The emergence of the vertebrate-specific APOBEC family

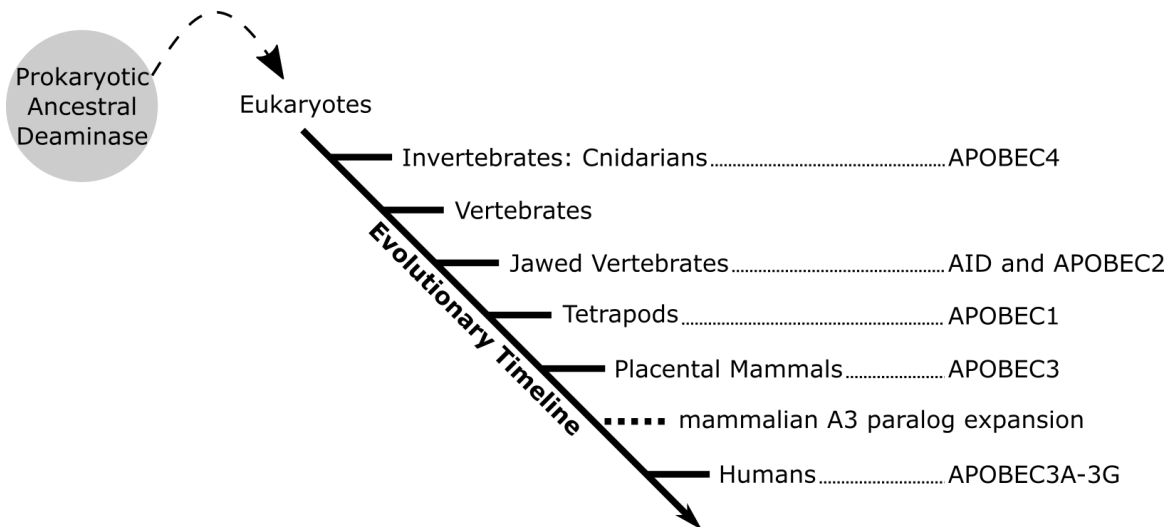


Figure 1.1. APOBEC family evolution in the vertebrate lineage. Ticks on the evolutionary timeline depict points where the vertebrate lineage diverged, such as the split between jawless and jawed vertebrates. The emergence of APOBEC members coincided with these splits. The timeline is not to scale.

The APOBEC family belongs to the deaminase superfamily involved in nucleotide catabolism and notably in the deamination of polynucleotides such as DNA and RNA (Iyer et al., 2011). Members of the deaminase superfamily contain the conserved zinc-coordinating deaminase fold motif (Conticello et al., 2005). Within the deaminase superfamily, the polynucleotide deaminases: APOBEC family, adenosine deaminase RNA specific (ADAR), and adenosine deaminase tRNA specific (ADAT) family, form separate clades based on the deaminase domain from the rest of the catabolic cytidine deaminases acting on free bases, nucleosides or nucleotides (Conticello et al., 2005).

This conserved deaminase domain could be traced back to deaminase domains of prokaryotic toxins that acted on foreign nucleic acids. These prokaryotic defense deaminases are believed to have jumped into eukaryotes through a lateral gene transfer from an endosymbiont (Iyer et al., 2011). The tRNA adenosine deaminases found in protozoans to metazoans are also thought to be the origin of the APOBEC family, given their structural similarities and the ability of these tRNA adenosine deaminases to deaminate cytosine in trypanosome DNA (Conticello, 2008; Rubio et al., 2007). Ancestral prokaryotes likely co-opted these polynucleotide editing deaminases to expand tRNA decoding capacity. And eventually, in eukaryotes, the deaminases were appropriated in the diverse functions of the vertebrate-specific APOBEC family.

1. Introduction

The APOBEC family, involved in cytosine to uracil editing in single-stranded nucleotides, is believed to have emerged in jawed vertebrates (gnathostomes) with AID and APOBEC2 as the most ancient members found in cartilaginous and bony fish (Conticello, 2008) (**Figure 1.1**). Through what is believed to have been AID gene duplication events, APOBEC1 emerged at the tetrapod-lungfish divergence; and then, APOBEC3 emerged later in placental mammals (Conticello et al., 2005; Krishnan et al., 2018). APOBEC3 paralog expansion then occurred in placental mammals with bats and humans having the most diverse repertoire of paralogs (Hayward et al., 2018; Münk et al., 2012). More recently, homologs of APOBEC4 found in invertebrate cnidarians suggest that APOBEC4 is the most ancient member separate from the vertebrate AID/APOBEC2-branch (Krishnan et al., 2018).

The process of tracing the origins of the APOBEC from invertebrates to vertebrates is still ongoing. As mentioned, the presence of APOBEC4 homologs expands what is known about the origins of the APOBEC family (Krishnan et al., 2018). Furthermore, APOBEC-like deaminases have also been described in jawless vertebrates (lampreys) and invertebrate sea urchins (Holland et al., 2018; Liu et al., 2018; Rogozin et al., 2007). Interestingly, these APOBEC-like deaminases play a role in the immune system of these species.

1.1.2 The APOBEC family and the vertebrate immune system

The origin of the APOBEC family from a deaminase domain of a bacterial toxin acting as a primitive immune system against foreign nucleic acids underscores the immunological roles played by the APOBEC family. Even in invertebrate sea urchins, the APOBEC-like deaminases are induced upon pathogen challenge as part of an immune response (Liu et al., 2018). Members of the APOBEC family have distinct roles in the vertebrate immune system.

AID is specifically expressed in B lymphocytes, where it catalyzes the deamination of immunoglobulin genes during B cell activation; it is directly involved in antibody diversification through somatic hypermutation and class switch recombination (Muramatsu et al., 1999, 2000). AID deaminates single-stranded DNA intermediates with specific secondary structures found in the switch or variable regions of immunoglobulin loci (Qiao et al., 2017). Interestingly, the APOBEC-like deaminases found in lampreys act on genes of immunoglobulin-like receptors found in jawless fish similar to the role of AID

in jawed vertebrates (Holland et al., 2018; Rogozin et al., 2007). It is theorized that in an ancestral immune cell, perhaps a phagocytic precursor to human B lymphocytes, the deaminase function targeting foreign DNA was co-opted to target endogenous receptor genes for enhanced pathogen recognition (Rogozin et al., 2007).

Though emerging later from AID, APOBEC3 plays a more similar role to the ancestral cytidine deaminase with the targeting of pathogenic viruses in innate immunity. The human paralogs of the APOBEC3 enzymes (APOBEC3A-D and 3F-H) are the most studied in terms of their role in the restriction of exogenous retroviruses, genomic retroelements, and some DNA viruses through deaminase-dependent and independent mechanisms (Harris & Dudley, 2015). The general mechanism of APOBEC3 viral restriction is through the targeted deamination of single-stranded cDNA intermediates of retroviral genomes – first elucidated in the role of APOBEC3G in human immunodeficiency virus type 1 (HIV-1) restriction (Harris et al., 2003). However, HIV-1 and other viruses owe their success through evolving mechanisms to evade APOBEC3-mediated viral restriction. The multiple paralogs of APOBEC3 in humans reflect this genetic arms race between pathogenic viruses and the human immune system (Harris & Dudley, 2015). This arms race is prominent in bats that are known reservoirs for emerging human viruses; this is highlighted by the numerous APOBEC3 paralogs found within bats compared to other mammals (Hayward et al., 2018). A salient point since as I write this work humanity is still struggling against a pandemic that started in December 2019 that was caused by a coronavirus, SARS-CoV-2, believed to have originated in bats (Zhou et al., 2020). This is further emphasized by the preponderance of C-to-U mutations in the SARS-CoV-2 genome; although, APOBEC3-mediated hypermutation of coronavirus genomes has not been demonstrated (Ratcliff & Simmonds, 2021). In the case of humans, likely as well as other mammals, the different APOBEC3 paralogs have evolved to target diverse exogenous and endogenous viral entities.

The roles of AID in adaptive B cell immunity and of APOBEC3 in innate antiviral immunity were co-opted from the ability of the ancestral APOBEC deaminase to mutate polynucleotides. In the case of AID, the original role of targeting foreign pathogens was co-opted to targeting endogenous immunoglobulin receptors. The other members of the APOBEC family seem to have been co-opted for other physiological roles.

1. Introduction

1.1.3 APOBEC roles beyond the immune system

APOBEC1, the founding member of the APOBEC family, was discovered as the deaminase responsible for a small intestine-specific C-to-U modification in the apolipoprotein B (APOB) RNA transcript leading to a tissue-specific truncated APOB protein (Teng et al., 1993). Unlike AID and APOBEC3 that deaminate cytosine in single-stranded DNA, APOBEC1 has an established physiological role in deaminating RNA. Moreover, beyond the *ApoB* RNA transcript, APOBEC1 targets diverse transcripts that are consequential in hepatocellular carcinoma and monocyte function (Harjanto et al., 2016; Rayon-Estrada et al., 2017; Yamanaka et al., 1997). The apparent role of APOBEC1 in the monocytic lineage echoes the role of an APOBEC precursor in an ancestral phagocytic immune cell, where again, deaminase activity was co-opted for endogenous RNA transcript editing for functional diversity in immune cells.

Unlike the rest of the APOBEC members discussed so far, APOBEC2 and APOBEC4 have no established deaminase activity. APOBEC4 was more recently discovered and forms a separate clade from the vertebrate-specific APOBEC family that may further clarify the ancestry of the APOBEC family (Krishnan et al., 2018; Rogozin et al., 2005). Aside from the expression in rodent and human testes, very little is currently known about the functional role of APOBEC4 (Rogozin et al., 2005). Further studies on the biological role of APOBEC4 may reveal yet unknown ancestral functions of the APOBEC family. Conversely, several studies have been done on the biological function of APOBEC2.

Before I discuss the studies on the biological function of APOBEC2, I would like to tackle the sequence conservation of the APOBEC members across the vertebrate lineage. The contrasting sequence conservation of each member and studies on the evolutionary history and known biological roles of the APOBEC family provide a basis for pursuing the biological function of APOBEC2.

1.1.4 Positive and purifying evolutionary pressures on the APOBEC family

The amino acid sequence changes in the APOBEC family received attention due to the ongoing genetic conflict between the APOBEC3 subfamily and viral pathogens. Each member of the family has varying ratios between synonymous and non-synonymous amino acid substitutions within primates and the vertebrate lineage, where a higher ratio of

non-synonymous to synonymous mutation (positive selection) is observed in APOBEC1 and APOBEC3 proteins (Krishnan et al., 2018; Sawyer et al., 2004). However, in the case of AID, APOBEC2 and APOBEC4, there is a strong negative or purifying selection pressure, where non-synonymous mutations are disfavored due to detrimental effects on biological function.

Positive selection, or rapid evolution, is prominent within the APOBEC3 subfamily likely due to their direct role in retroviral restriction. However, the positive selection observed for APOBEC1 is surprising given its conserved role in targeting endogenous RNA transcripts such as *Apob* mRNA. There is evidence, although conflicting, for a role of APOBEC1 in retroviral restriction which may explain the positive selection (Harris & Dudley, 2015). However, similar evidence for an antiviral role is present for AID which is under strong purifying selection throughout the vertebrate lineage. It can be hypothesized that the positive selection experienced by APOBEC1 is due to its reliance on protein cofactors for targeting and effecting its biological function (Fossat et al., 2014; Lellek et al., 2000; Mehta et al., 2000). Perhaps for APOBEC1, the deaminase domain is conserved, while changes in the accessory domains are dictated by protein-protein interaction across species.

Among the members, AID, APOBEC2, and APOBEC4 are under strong purifying selection, indicative of a conserved and vital biological function (Krishnan et al., 2018). For AID, its sequence conservation reflects its central role in immunoglobulin gene diversification. Given that we currently know so little about the biological function of APOBEC4 and its presence in invertebrate species, it is tempting to speculate that it serves an unknown ancestral deaminase function. For APOBEC2, this thesis aims to determine its strongly conserved molecular function that has been carried through the vertebrate lineage.

1. Introduction

1.2 The physiological role of APOBEC2

APOBEC2 was discovered due to its sequence similarity to APOBEC1 and its conserved APOBEC cytidine deaminase domain (Anant et al., 2001; Liao et al., 1999). APOBEC2 is strongly enriched in striated muscle, both skeletal and cardiac, where a potential physiological role for APOBEC2 was uncovered.

1.2.1 APOBEC2 loss in animal models leads to muscular dystrophy in skeletal muscle

Animal models, where APOBEC2 expression was removed, were used to elucidate the physiological role of APOBEC2. These APOBEC2 knockout and knockdown models revealed changes in skeletal muscle and, though less prominently, cardiac muscle.

For mouse (*Mus musculus*) models, the loss of APOBEC2 did not lead to observable effects in health, breeding, or survival (Mikl et al., 2005). However, closer inspection of APOBEC2 knockout mice revealed subtle changes in leg muscles: a shift in proportions of slow and fast-twitch fibers, and a mild myopathy observed in older mice (Sato et al., 2010). More recently, in the same animal models, abnormalities in skeletal muscles were also observed due to defects in mitochondrial function (Sato et al., 2017). Remarkably, even though APOBEC2 is also highly expressed in cardiac muscle, there has been no mention of a heart defect in APOBEC2 knockout mice. Although, heart defects may similarly be subtle.

In the zebrafish (*Danio rerio*) model, the knockdown of APOBEC2 resulted in observable defects in both skeletal and cardiac muscle. Zebrafish have two APOBEC2 isoforms, Apobec2a and Apobec2b, that are expressed in cardiac and skeletal muscle; decreased expression of both isoforms led to substantial defects in skeletal muscle structure and heart dysfunction in zebrafish embryos (Etard et al., 2010). It is interesting to note that APOBEC2 loss in zebrafish exhibited a more evident muscular dystrophy than in mice. This could be due to the biological difference between species, or possibly the lack of other APOBEC proteins (APOBEC1 and APOBEC3) in zebrafish. However, redundancy of function between APOBEC subfamilies, except perhaps between APOBEC3 paralogs, has not been described. Furthermore, expression of other APOBEC members in striated muscle has not been detected (Mikl et al., 2005).

Both animal models, zebrafish and mice, point to a conserved physiological function of APOBEC2 within the context of striated muscle. Furthermore, in zebrafish and other models, evidence exists for a role of APOBEC2 beyond the striated muscle context.

1.2.2 APOBEC2 has roles beyond muscle tissue

One of the initial studies showed evidence of APOBEC2 expression outside muscle tissue in humans and mice (Anant et al., 2001). However, most others have shown that expression seems to be exclusive to the skeletal and cardiac muscle (Liao et al., 1999; Mikl et al., 2005). Looking at transcript and protein expression across human tissues (Human Protein Atlas project), APOBEC2 expression is enriched in cardiac and skeletal muscle; however, it is also expressed in some neuronal cells of the eye, several blood cell lineages, spermatids, and epithelial cells from different tissues (Uhlén et al., 2015). Furthermore, APOBEC2 loss seems to affect tissues other than muscle.

In the zebrafish model, loss of APOBEC2 affected retinal regeneration and embryo development. Both isoforms of APOBEC2 (*Apobec2a* and *Apobec2b*) in zebrafish were expressed in a retinal regeneration model (Powell et al., 2012, 2013). In this model, loss of APOBEC2 impaired regeneration of the retina that required dedifferentiation of Müller glia to reform the lost retinal tissue, which included neural and glial cells (Powell et al., 2012). Moreover, during zebrafish and frog (*Xenopus*) embryo development, APOBEC2 also plays a role in left-right axis specification. Knockdown of APOBEC2 during early embryogenesis led to inversions in heart folding and gut looping – indicators of heterotaxia (Vonica et al., 2011). These studies point to a broader role of APOBEC2 in tissue regeneration and development.

APOBEC2 has also been linked to carcinogenesis and cancer progression. In a mouse model of APOBEC2 overexpression, APOBEC2 was linked to the development of hepatocellular and lung carcinomas, where APOBEC2 was thought to deaminate gene transcripts linked to tumorigenesis (Okuyama et al., 2012). Additionally, inflammation in the liver, a potential precursor to carcinogenesis, leads to upregulation of APOBEC2 (Matsumoto et al., 2006). In diffuse large B cell lymphoma, APOBEC2 was among the significantly mutated genes in patient samples (Lohr et al., 2012). Lastly, in cervical cancer, the APOBEC2 gene was among the frequently lost genes in high-grade cervical intraepithelial neoplasia, which correlated with its decreased expression in squamous cell carcinomas (Bierkens et al., 2013). It is uncertain whether APOBEC2 is a driver or

1. Introduction

passenger mutation in these cancers. In the case of the hepatocellular carcinoma model, it was proposed to be acting as a driver through deamination of the target transcripts. However, whether APOBEC2 has bonafide deaminase activity has been contentious.

1.3 APOBEC2 structure suggests an evolutionarily conserved APOBEC function

There have been several conflicting reports regarding the deaminase activity of APOBEC2. In its initial discovery, APOBEC2 was thought to be able to deaminate free cytosines to uracil (Anant et al., 2001; Liao et al., 1999). However, subsequent works showed that, unlike other APOBEC family members, APOBEC2 was unable to mutate DNA in bacterial mutator assays and showed no deaminase activity (Etard et al., 2010; Harris et al., 2002; Mikl et al., 2005; Powell et al., 2014). The initial observed deaminase activity may have been an artifact from bacterial deaminases during APOBEC2 protein purification (Mikl et al., 2005). It has also been speculated that APOBEC2 plays a role in active DNA demethylation through coupling deamination with DNA repair (Guo et al., 2011; Rai et al., 2008). However, subsequent work provided no direct evidence for the role of APOBEC2 in DNA demethylation (Powell et al., 2013). Furthermore, due to steric hindrance, methylated cytosines were unlikely substrates for cytidine deaminases (Nabel et al., 2012). These contentious results are intriguing given that APOBEC2 retains the evolutionarily conserved deaminase domain of the APOBEC family.

1. Introduction

1.3.1 APOBEC2 protein structure bears the conserved cytidine deaminase domain

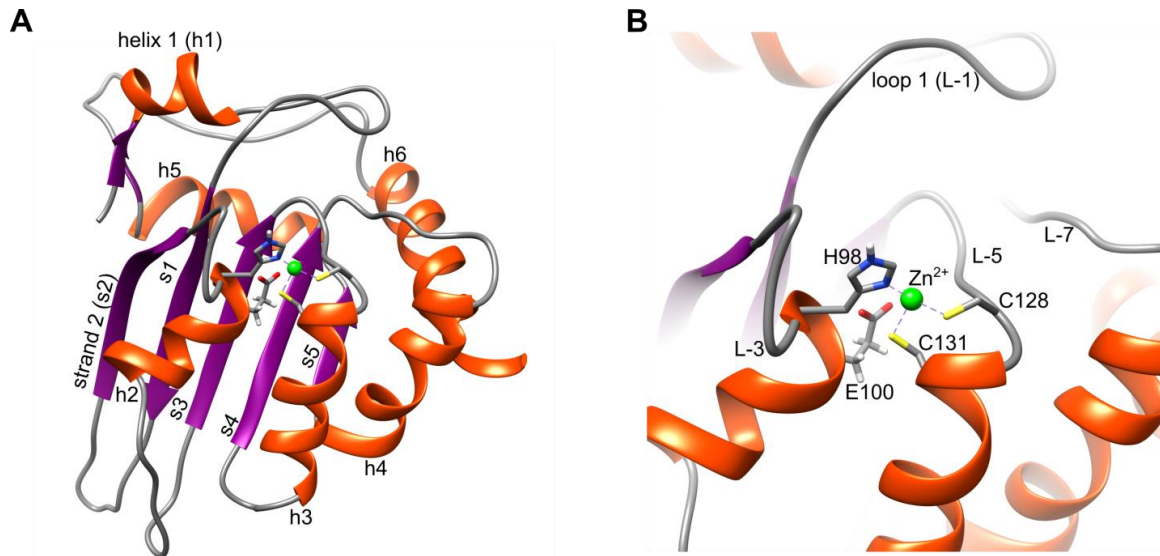


Figure 1.2. APOBEC2 protein structure.

A) APOBEC2 (mouse) structure derived from solution nuclear magnetic resonance (NMR) spectroscopy (PDB: 2RPZ, model 13). The first 45 residues were not included in the structure. The α -helices are labeled (h1 to h6) and colored in orange. The strands comprising the β -sheet are labeled (s1 to s6) and colored in purple. The image was prepared using UCSF Chimera (Pettersen et al., 2004).

B) Focused view from (A) of the core residues comprising the catalytic domain of APOBEC2. The zinc ion (Zn^{2+}) is colored green. The zinc-coordinating residues, H98, C128, and C131, and the catalytic E100 residue are depicted as sticks. The loops surrounding the catalytic domain are labeled, loop 1 (L-1), L-3, L-5, and L-7.

The APOBEC family is characterized by the APOBEC cytidine deaminase domain composed of five beta(β)-strands that form a hydrophobic β -sheet surrounded by six alpha(α)-helices with the order: α -helix 1 (h1)- β -strand 1 (s1)-s2-h2-s3-h3-s4-h4-s5-h5-h6 (Harris & Dudley, 2015; Salter & Smith, 2018). Within this structure are the highly conserved zinc-coordinating deaminase fold motifs: H-X-E and P-C-X(2-4)-C (Salter & Smith, 2018). Within the APOBEC3 subfamily are APOBEC3B, 3D, 3G, and 3H that contain two of these cytidine deaminase domains in tandem (Salter & Smith, 2018). As mentioned earlier, APOBEC2 was the first to have its crystal structure published, which has helped elucidate the molecular structures and functions of the rest of the family (Prochnow et al., 2007).

The first published structure of human APOBEC2 showed exactly how the APOBEC cytidine deaminase domain folded (**Figure 1.2**). In the initial crystal structure, where APOBEC2 formed a tetramer within the protein crystal, each monomer folded into the characteristic six α -helices surrounding the five-stranded β -sheet (Prochnow et al.,

2007). The APOBEC2 structure also shows how the APOBEC cytidine deaminase domain coordinates zinc at the His-X-Glu motif found on h2 and the adjacent Pro-Cys-X(2-4)-Cys motif found on h3. Published later, the full-length structure of human APOBEC2 (residues 1-224) in solution had key differences to the prior crystal structure of APOBEC2 (residues 41-224), which removed the initial amino(N)-terminal portion upstream of h1 (Krzysiak et al., 2012). First, the N-terminal region that was removed proved to be a highly flexible negatively-charged unstructured region that would interfere with multimerization (Krzysiak et al., 2012). Thus, APOBEC2 likely exists as a monomer in solution. Second, the loop between h1 and s1 (loop 1, L-1) that obscures the active site in the human APOBEC2 crystal structure is flexible in solution and could freely cover or expose the active site (Krzysiak et al., 2012). Lastly, the first α -helix (h1) is also a highly mobile domain (Krzysiak et al., 2012). The flexibility of these domains may affect and dictate intermolecular ligand or substrate interactions.

1. Introduction

1.3.2 The evolutionary conservation of the APOBEC2 protein structure suggests a conserved molecular function

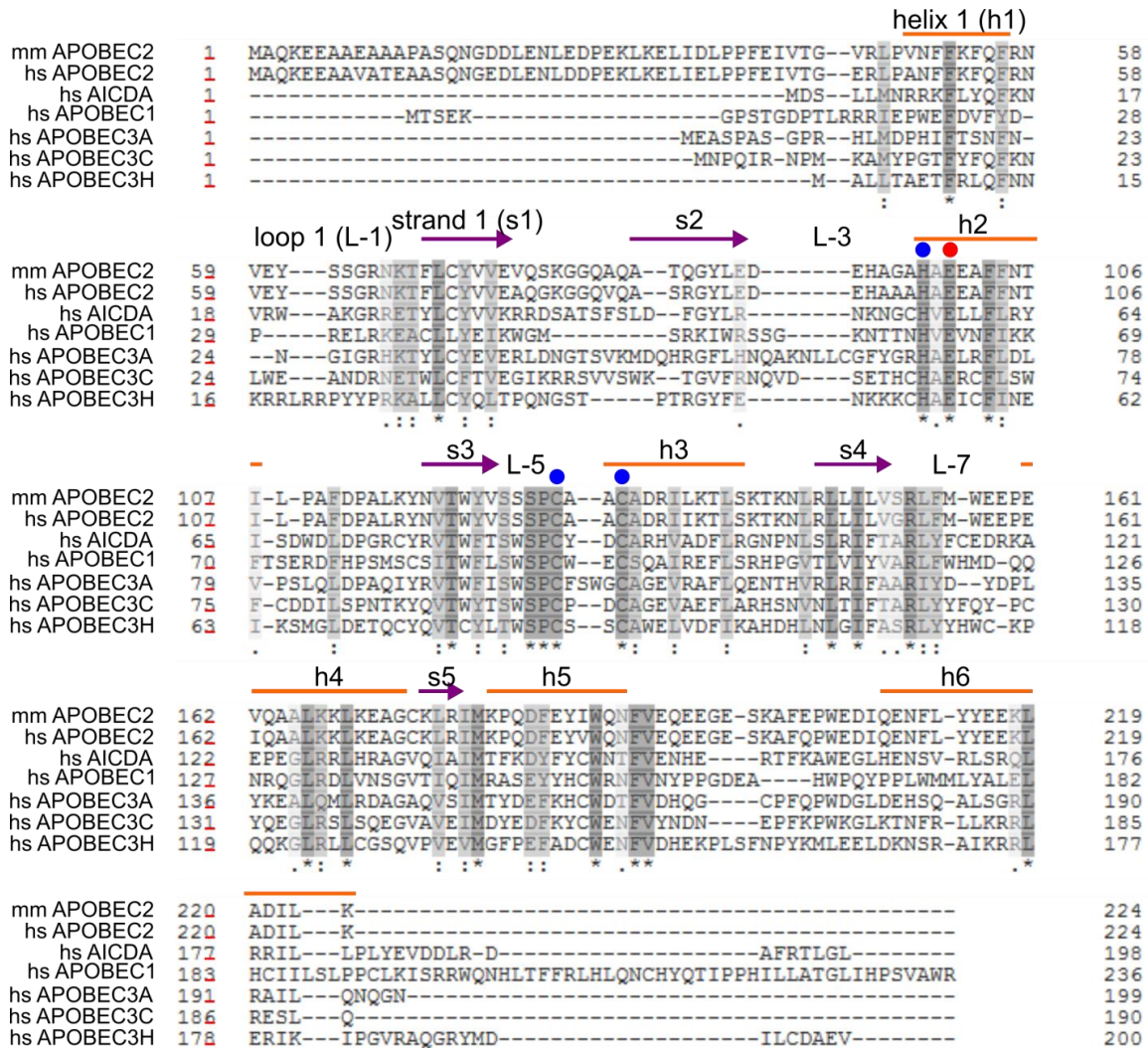


Figure 1.3. Protein sequence alignment of mouse and human APOBEC2 with other single domain human APOBEC family members. Alpha(α)-helices (h1 to h6) are marked with orange bars. Beta(β)-strands (s1 to s5) making up the core β -sheet are marked with purple arrows. Locations of loops 1, 3, 5, and 7 (L-1 to L-7) are indicated. The annotation of the domains was based on the APOBEC2 structure (PDB: 2RPZ). The conserved zinc-binding residues are marked with blue circles, while the catalytic glutamic acid residue is marked with a red circle. Protein sequences were obtained from UniProt and alignment was done using Clustal Omega. Conserved residues are marked and highlighted in gray. hs: Homo sapiens; ms: Mus musculus.

As previously discussed, the cytidine deaminase fold is highly conserved across the APOBEC family members. The zinc-binding residues and the catalytic glutamic acid residue are retained throughout members of the family; moreover, residues within loops, helices, and strands are also highly conserved (**Figure 1.3**).

Comparing the known structure of APOBEC2 to the published structures of APOBECs in complex with their nucleic acid substrate provides ideas on how APOBEC2

would interact with its yet unknown molecular substrate. Structures of APOBEC3A (PDB ID: 5KEG and 5SWW) and APOBEC3G (PDB ID: 6BUX) show the importance of amino acid residues on loops 1, 3, 5, and 7 in interactions with the ssDNA substrate (Kouno et al., 2017; Maiti et al., 2018; Shi et al., 2017). Furthermore, the structure of AID, in complex with the target deoxycytidine, shows how positively charged residues in loop 7 and negatively charged residues in loop 1 and h6 could support interaction with structured ssDNA from open dsDNA (Qiao et al., 2017). Further, a comparison of the evolutionary variation and conservation in loop length and residues within these substrate-interacting loops reflect their importance in substrate specification; AID and APOBEC2 have conserved residues at these loops while the APOBEC1 and APOBEC3 subfamilies have high variability (Krishnan et al., 2018). The evolutionarily conserved cytidine deaminase structure of APOBEC2 and the conservation and variability within the substrate interacting domains suggest that APOBEC2 interacts with a yet undetermined nucleic acid ligand. Furthermore, like in the case of AID, the purifying selection suggests this interaction is biologically critical.

1. Introduction

1.4 Towards determining the molecular function of APOBEC2

The APOBEC family traces its origins to tRNA cytidine deaminases and ancestral cytidine deaminase domains of toxins used by ancestral prokaryotes. These ancestral functions have been appropriated by members of the APOBEC family into diverse and unique roles in immunological and other physiological contexts. Prior studies in animal models point to a physiological role of APOBEC2 in muscle tissue where it is highly enriched, and it may also play a role in regenerative or developmental contexts. Furthermore, APOBEC2 has been linked to inflammation and carcinogenesis. However, the specific molecular role APOBEC2 plays in these physiological or pathological contexts is still unclear.

The existence of the enzymatic activity of APOBEC2 has been questionable so far. However, the sequence and structural conservation of the deaminase domain of APOBEC2 suggest a molecular function that is similar to the rest of the family members. Furthermore, the purifying selection, where detrimental changes in sequence were disfavored, strongly suggests a conserved and essential molecular function.

In this current work, my primary goal is to identify this evolutionarily conserved molecular function. Prior work in Prof. Dr. Nina Papavasiliou's lab had laid the foundation for pursuing the molecular function of APOBEC2 within the context of differentiated myotubes. The work before mine focused on the hypothesized roles of APOBEC2 in RNA deamination and DNA demethylation to explain physiological or gene transcriptional changes caused by APOBEC2 loss in muscle. However, based on the experimental results and the consensus in the field, no such enzymatic activity exists for APOBEC2. Nonetheless, the results of my work and the ideas I present here continued and built upon this work. Through my efforts, I identified the putative molecular function for APOBEC2 in transcriptional regulation affecting muscle differentiation.

2. Aims

Determining the molecular function of APOBEC2 would enrich our understanding of the immunologically important APOBEC family. Beyond scientific curiosity and academic pursuit, medical applications could also be designed based on the ability of these enzymes to edit nucleic acids.

In this thesis, I sought to determine the molecular function of APOBEC2 in the context of chromatin and transcriptional regulation in muscle cells. In doing so, I aimed (1) to elucidate the unknown molecular role of APOBEC2 in muscle differentiation; and (2) to uncover the molecular mechanism behind this role.

3. Results

3. Results

3.1 APOBEC2 and myotube differentiation

In investigating the putative molecular function of APOBEC2 (A2), we used the C2C12 mouse myoblast cell line C2C12 as a model. These cells recapitulate myoblast to myotube differentiation in vitro upon induction to differentiate (Blau et al., 1985). I induced the C2C12 myoblasts to differentiate into myotubes through serum starvation by replacing the medium containing 10% fetal calf serum with a medium containing 2% horse serum instead – a standard protocol for C2C12 myotube differentiation (McMahon et al., 1994). I confirmed differentiation by quantifying myotube formation by staining for myosin heavy chain (MyHC), a muscle-specific protein found in skeletal myotubes. I checked differentiation at three time points: day 0, as the starting point, and 2 and 5 days after inducing the myoblasts to differentiate (**Figure 3.1.A,B**). The number of myotubes, characterized as multinucleated MyHC-positive (MyHC⁺) cells, were increasing as the time of differentiation progressed. The fusion index (a ratio of the number of multinucleated MyHC⁺ cells to the total number of nuclei) was increasing; day 0: 0.121 ± 0.512 ; day 2: 2.901 ± 2.475 ; and day 5: 15.328 ± 4.814 (mean \pm standard deviation). There was also a noticeable increase in APOBEC2-positive cells; in particular, only single MyHC⁺ cells and myotubes had enriched cytoplasmic and nuclear staining of APOBEC2. Moreover, On day 5, there was noticeable nuclear enrichment of APOBEC2 in the myotubes (**Figure 3.1.A**). The observed increase in APOBEC2-positive cells during C2C12 myoblast to myotube differentiation was similar to the observed increase in APOBEC2 protein levels in differentiating mouse primary myoblasts (Sato et al., 2010). Thus, given that C2C12 myoblasts recapitulate myotube formation in vitro, it was chosen as a suitable model to study the putative molecular function of APOBEC2.

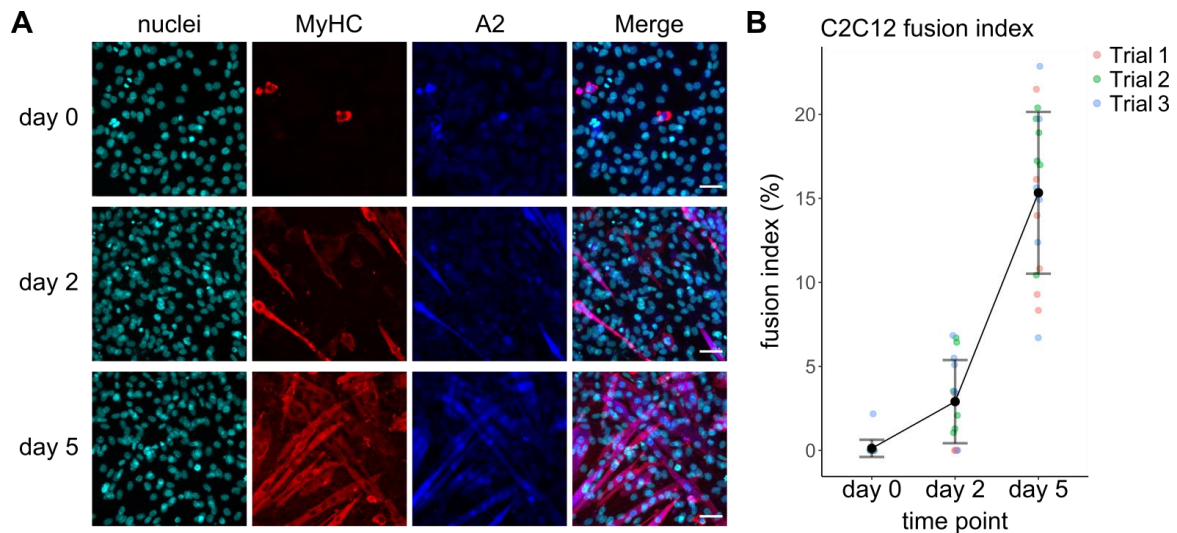


Figure 3.1. Myotube formation in differentiated C2C12 myoblasts.

A) Immunofluorescence images of differentiating C2C12 myoblasts at three time points of differentiation, day 0, day 2, and day 5. Each column represents a single channel corresponding to the nuclei, myosin heavy chain (MyHC), APOBEC2 (A2), and merged channels (Merge). Each row represents the time points. Images were taken with a widefield fluorescence microscope, processed uniformly, and representative of the groups. Scale bar: 50 μ m.

B) Measured fusion indices of differentiating C2C12 myoblasts at three time points of differentiation, day 0, day 2, and day 5. Fusion index represents the ratio of multinucleated (nuclei ≥ 2) MyHC positive cells over the total number of nuclei in each field of view. Black dots represent the mean of the fusion index at each time point from 3 trials with 6 fields of view each; the black line connects the means. Error bars represent the standard deviation. The colored dots, grouped by trial, represent the measured fusion index from each field of view.

3.1.1 RNA-Seq of APOBEC2 knockdown cells pointed to a decrease in skeletal muscle differentiation

In pursuing the molecular function of APOBEC2, the first task I did was look through the results left to me by Linda Molla, a prior Ph.D. student. Short hairpin RNA (shRNA)-mediated transcriptional silencing was used to specifically decrease APOBEC2 levels in C2C12 myoblasts. C2C12 myoblast cell lines were generated to stably express shRNA against APOBEC2 (shA2) or green fluorescent protein (shGFP) as a non-targeting control. During differentiation, the shA2 C2C12 myoblasts showed the expected decrease in APOBEC2 protein levels compared to the shGFP control cells (**Figure 3.2**). Notably, the shA2 cells also showed a reduction in levels of muscle-specific proteins - myosin heavy chain (MyHC) and Troponin T. This observation prompted me to look further into a defect in myoblast to myotube differentiation in the shA2 cells.

3. Results

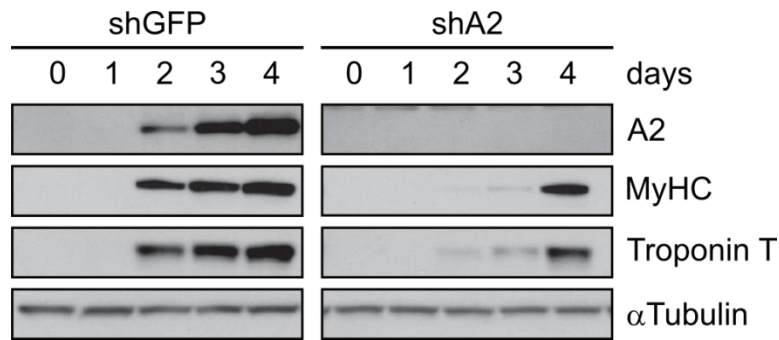


Figure 3.2. Western blots of APOBEC2 and muscle-specific proteins in knockdown cells. Western blots showing levels of APOBEC2 (A2), myosin heavy chain (MyHC), Troponin T, and alpha (α)Tubulin in total protein lysates from either shGFP or shA2 knockdown C2C12 myoblasts. α Tubulin was used as a protein loading control. Each lane represents different lengths of differentiation, day 0 to day 4, of the specified cells. I prepared this figure and one similar to it in the preprint I submitted in bioRxiv (Lorenzo et al., 2021).

The APOBEC2 knockdown cell line was initially generated to study the putative role of APOBEC2 as an RNA deaminase by comparing levels of nucleic acid substitutions in RNA transcripts from shA2 and shGFP cells. Thus, RNA-Seq was done to sequence mRNA transcripts and measure global mRNA levels from the shA2 and shGFP cells. However, from the comparison of the RNA transcripts between shA2 and shGFP, Linda Molla was unable to find evidence for the hypothesized role of APOBEC2 in RNA deamination. Nonetheless, it was noted that the decrease in APOBEC2 led to substantial gene expression changes between shA2 and shGFP cells.

I used this RNA-Seq data to find evidence instead for a molecular role of APOBEC2 in the regulation of muscle differentiation. I analyzed the RNA-Seq data that was from shA2 and shGFP cells during the early stages of differentiation – before differentiation (day 0), and 1 and 2 days after induction to differentiate (day 1 and day 2). The APOBEC2 RNA transcripts, as expected, were decreased in the shA2 C2C12 cells; APOBEC2 transcript counts in shA2 only reached up to the basal (day 0) levels of shGFP across the time points (**Figure 3.3.A**). Comparing APOBEC2 transcript levels on day 2 shows a 30-fold decrease between shA2 and shGFP (\log_2 Fold Change = -4.9, p .adjusted = 1.4×10^{43}). I then looked at the gene expression changes between shA2 and shGFP at each time point using the DESeq2 package in R (Love et al., 2014). There were around 600 differentially expressed genes before differentiation (day 0) between the shA2 and shGFP cells (MA plot day 0, **Figure 3.3.B**). This increased to about 1800 differentially expressed genes on day 1 and day 2 as the cells were differentiating into myotubes (MA plot day 1 and day 2, **Figure 3.3.B**). The small number of differentially expressed genes on day 0 likely reflects the similarity between shGFP and shA2 myoblasts before

differentiation. On the other hand, the substantially greater number of differentially expressed genes when the myoblasts were induced to differentiate (day 1 and day 2) likely represent consequential changes that may have had a bearing on the decreased levels of muscle-specific markers observed at the later time point (day 4, **Figure 3.2**).

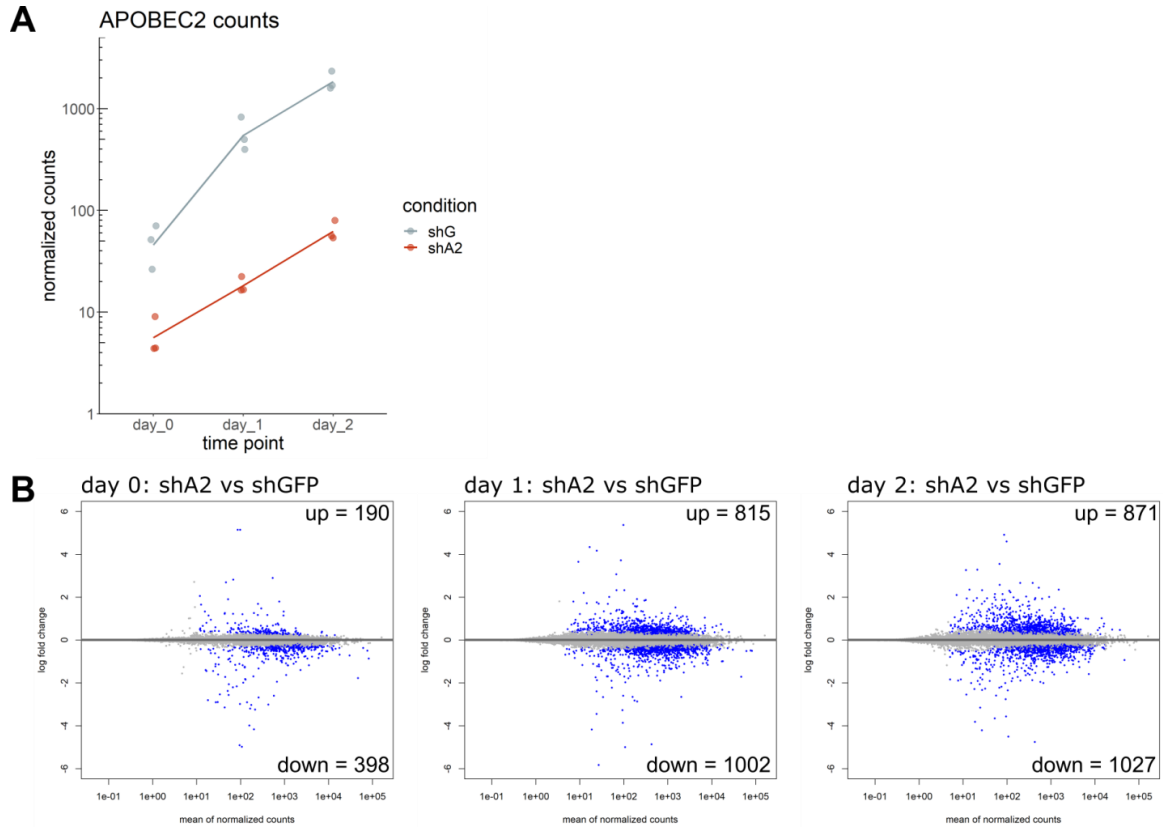


Figure 3.3. RNA-Seq differential expression analysis between APOBEC2 knockdown and control cells at early time points of myotube differentiation.

A) DESeq2 normalized counts of APOBEC2 RNA transcripts from day 0 to day 2. The y-axis (log₁₀ scale) represents the normalized counts. The line connects the mean (n = 3) of normalized counts across the specified time points. Condition: shG = shGFP and shA2 = shAPOBEC2 transduced C2C12 cells.

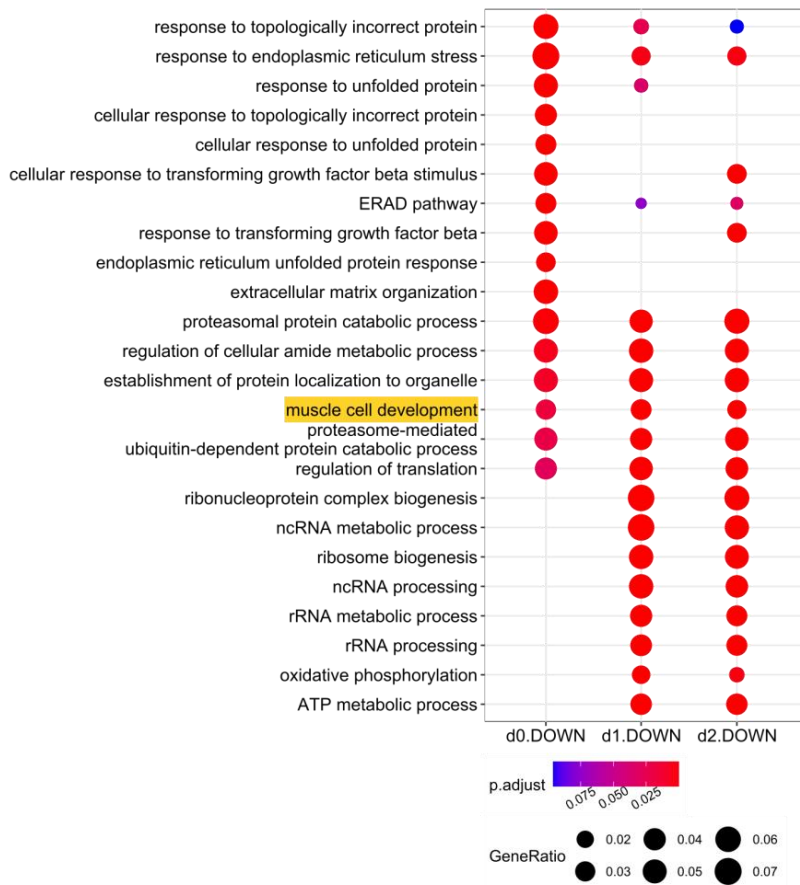
B) MA plots show the distribution of log₂ fold changes (M, y-axis) and mean of normalized counts (A, x-axis) of differentially expressed genes between the shA2 and shGFP C2C12 cells across the time points: day 0, day 1, and day 2. Plotted points highlighted in blue have p.adjusted values < 0.05, while points in gray have p.adjusted values ≥ 0.05. up and down represent number of upregulated (log₂ fold change > 0) and downregulated genes (log₂ fold change < 0) that have p.adjusted values < 0.05.

Using the lists of differentially expressed genes, I checked which genetic pathways were perturbed with the decrease in APOBEC2 expression. I selected the genes that were either downregulated (log₂ fold change < 0) or upregulated (log₂ fold change > 0) between shA2 and shGFP at each time point, and that were within the statistical significance cutoff (p.adjusted < 0.05). I used this filtered set of genes from each time point to run a gene ontology (GO) over-representation test using the clusterProfiler package in R (Wu et al., 2021). Looking at either downregulated or upregulated gene sets,

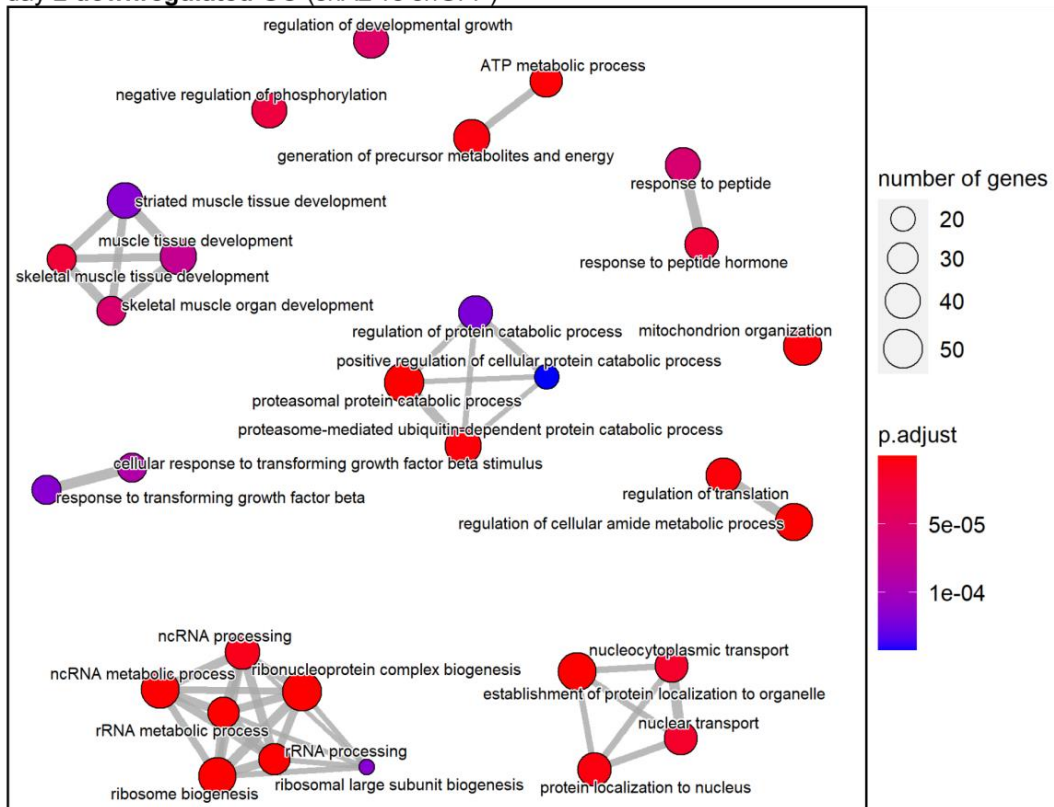
3. Results

several enriched GO terms overlap across the time points (**Figure 3.4, Figure 3.5**). For the downregulated genes, GO terms involved in protein metabolism and degradation were enriched on day 0 and day 1; while terms involved in ribosome biogenesis, including rRNA processing, were enriched on day 1 and day 2 (**Figure 3.4.A**). Interestingly, the term muscle cell development is enriched across the time points. Looking at the relationships between the GO terms enriched on day 2 through a cluster map highlighted skeletal muscle tissue development, ribosome biogenesis, and protein metabolism as affected pathways (**Figure 3.4.B**). These pathways reflect a decrease in muscle cell differentiation. Ribosome biogenesis and protein metabolic processes may also reflect the decrease in skeletal muscle differentiation since these are characteristic pathways typically increased in muscle hypertrophy and regeneration (von Walden, 2019).

A GO biological process:
downregulated GO across time points (shA2 vs shGFP)



B GO biological process:
day 2 downregulated GO (shA2 vs shGFP)



3. Results

Figure 3.4. Gene ontology analysis of downregulated genes in APOBEC2 knockdown cells during early myoblast differentiation.

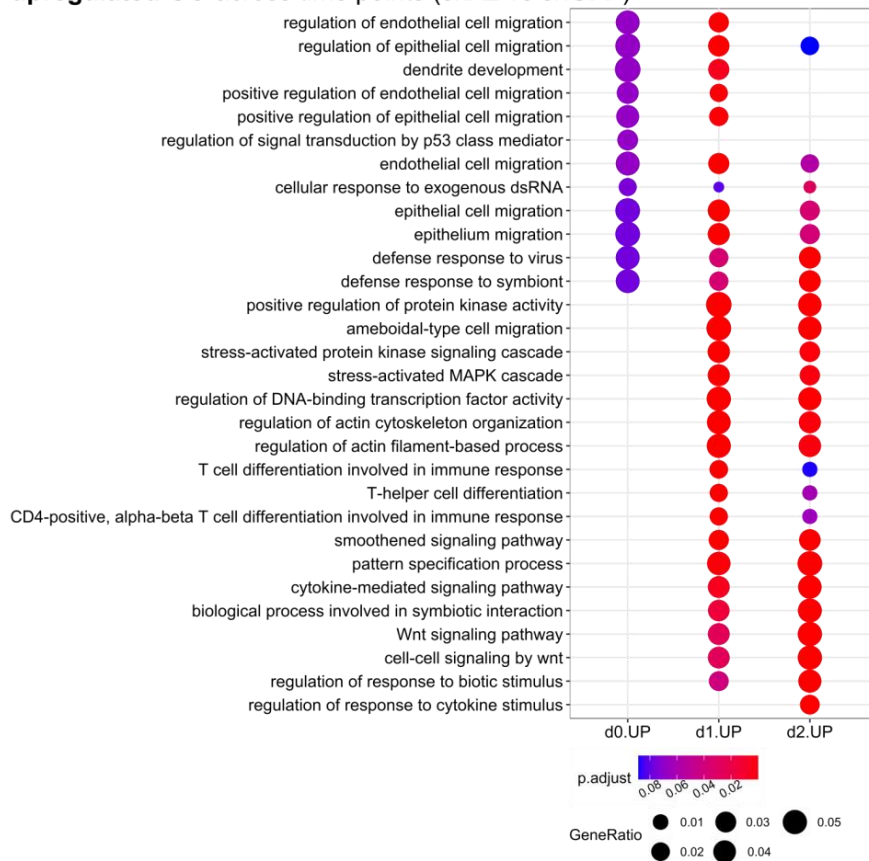
A) Dot plot representing the overlap of enriched gene ontology (GO) terms from downregulated genes across the time points. Gene lists of downregulated genes between shA2 and shGFP at each time point (day 0, day 1, and day 2: d0.DOWN, d1.DOWN, and d2.DOWN) were used as input for GO over-representation analysis. Enriched GO terms are ranked on the y-axis (top 10 of each list) by statistical significance (p.adjusted) across the lists on the x-axis. The colors of the dots correspond to p.adjusted values, where the red to blue gradient corresponds to low to high values. The sizes of the dots directly correspond to Gene Ratio. Gene Ratio corresponds to k/n where k is the number of genes from the provided list that fall within the corresponding GO term and n is the total number of genes in the provided list.

B) Cluster map of enriched GO terms from downregulated genes between shA2 and shGFP on day 2. The colors of the dots correspond to p.adjusted values, where the red to blue gradient corresponds to low to high values. The sizes of the dots directly correspond to k , the number of genes from the provided list that fall within the corresponding GO term. Edges connect overlapping GO terms; overlap is determined by genes within the terms.

On the other hand, enriched GO terms from the lists of upregulated genes across the time points are composed of terms involved in Wnt signaling and immune responses (T cell differentiation and defense response to virus) (**Figure 3.5.A**). The terms on day 0 were above the statistical significance cutoff ($p_{\text{adjusted}} < 0.05$) since the gene set of upregulated genes on day 0 was particularly small. Nevertheless, these GO terms were enriched on day 1 and day 2. The Wnt signaling and immune response GO terms were also highlighted in the cluster map of the relationships of enriched GO terms on day 2. However, a cluster of GO terms involved in kidney development and other developmental terms was particularly interesting (**Figure 3.5.B**). The enrichment of Wnt signaling GO terms for the upregulated genes may reflect the non-terminally differentiated myoblast state, wherein Wnt signaling is involved in myogenesis and muscle regeneration (von Maltzahn et al., 2012). Interestingly, GO terms involved in immune cell activation together with terms involved in organ development, such as kidney and neural tube development, were also enriched from the list of upregulated genes when APOBEC2 was knocked down. The significance of these upregulated gene ontologies was only apparent to me after I gathered more data; and so, I would discuss this further later.

3. Results

A GO biological process: upregulated GO across time points (shA2 vs shGFP)



B GO biological process: day 2 upregulated GO (shA2 vs shGFP)

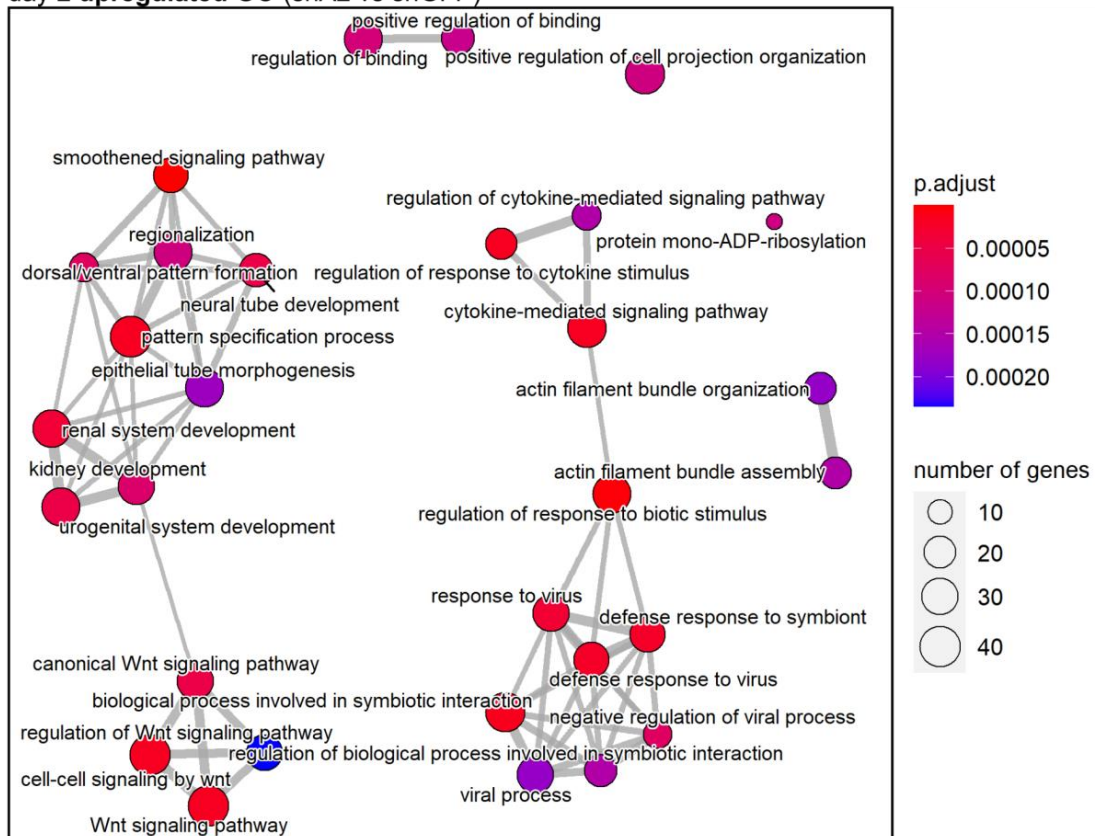


Figure 3.5. Gene ontology analysis of upregulated genes in APOBEC2 knockdown cells during early myoblast differentiation.

A) Dot plot representing the overlap of enriched gene ontology (GO) terms from upregulated genes across the time points. Gene lists of downregulated genes between shA2 and shGFP at each time point (day 0, day 1, and day 2: d0.UP, d1.UP, and d2.UP) were used as input for GO over-representation analysis. Enriched GO terms are ranked on the y-axis (top 10 of each list) by statistical significance (p.adjusted) across the lists on the x-axis. The colors of the dots correspond to p.adjusted values, where the red to blue gradient corresponds to low to high values. The sizes of the dots directly correspond to Gene Ratio. Gene Ratio corresponds to k/n where k is the number of genes from the provided list that fall within the corresponding GO term and n is the total number of genes in the provided list.

B) Cluster map of enriched GO terms from upregulated genes between shA2 and shGFP on day 2. The colors of the dots correspond to p.adjusted values, where the red to blue gradient corresponds to low to high values. The sizes of the dots directly correspond to k , the number of genes from the provided list that fall within the corresponding GO term. Edges connect overlapping GO terms; overlap is determined by genes within the terms.

Nevertheless, analysis of the RNA-Seq data from the shA2 and shGFP cells point to a decrease in muscle cell development or differentiation in the APOBEC2 knockdown cells. From the list of several hundred downregulated genes with APOBEC2 knockdown, GO enrichment analysis pointed directly to skeletal muscle differentiation and terms involved in muscle physiology: ribosome biogenesis and protein metabolism. The perturbation of these pathways at the early stages of myoblast to myotube differentiation in the knockdown cells may have led to the observed decrease in muscle-specific protein levels.

3.1.2 APOBEC2 knockdown led to a decrease in myotube formation in C2C12 myoblasts

The RNA-Seq analysis suggested a decrease in pathways involved in muscle differentiation in the shA2 C2C12 cells. I decided to confirm this directly by checking if there was a decrease in myotube formation in the shA2 cells when they are induced to differentiate. I induced the shA2 and shGFP cells to differentiate for up to 5 days, as I did before with the parental C2C12 myoblasts. Compared to the shGFP control cells, there was visibly less myoblast formation in the shA2 cells, as indicated by fewer MyHC⁺ multinucleated cells (**Figure 3.6.A,B**). On day 0 and day 2, there was a difference in the fusion index between shA2 and shGFP; however on day 5, there was significantly less myotube formation in the shA2 cells than shGFP cells, fusion index: 1.3 ± 2.0 vs 10.6 ± 7.6 (mean \pm standard deviation, Mann-Whitney test $p = 1.98 \times 10^{-6}$) (**Figure 3.6.C**). On day 2, although not statistically significant, there were visible indications that the shA2 cells had fewer myotubes and MyHC⁺ cells. Together with the RNA-Seq data, the results

3. Results

suggested that the downregulated genes involved in muscle development GO processes in shA2 at the early stages of differentiation (day 1 and day 2) influenced myoblast differentiation. This early effect likely resulted in the decrease in myotube differentiation on day 5.

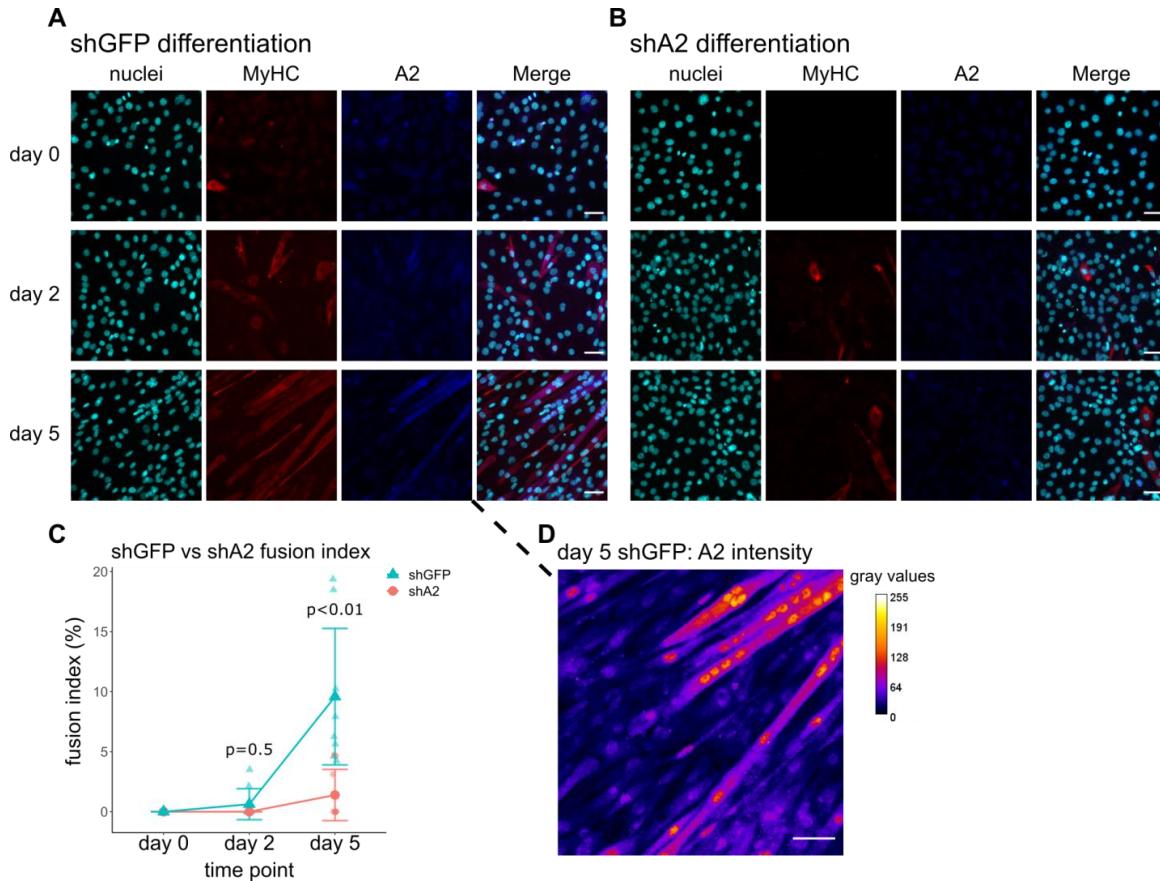


Figure 3.6. Myotube formation in differentiating APOBEC2 knockdown C2C12 myoblasts.

A,B) Immunofluorescence images of differentiating shA2 (A) and shGFP (B) C2C12 myoblasts at three time points of differentiation, day 0, day 2, and day 5. Each column represents a single channel corresponding to the nuclei, myosin heavy chain (MyHC), APOBEC2 (A2), and merged channels (Merge). Each row represents the time points. Images were taken with a widefield fluorescence microscope, processed uniformly, and representative of the groups. Scale bar: 50 μ m.

C) Measured fusion indices of differentiating shA2 and shGFP C2C12 myoblasts at three time points of differentiation, day 0, day 2, and day 5. Fusion index represents the ratio of multinucleated (nuclei \geq 2) MyHC positive cells over the total number of nuclei in each field of view. Red circles represent values for shA2; blue triangles represent values of shGFP. The corresponding line connects the mean values of the fusion index at each time point from 3 trials with 6 fields of view each. Error bars represent the standard deviation. P values were calculated from Mann-Whitney tests using JASP (Version 0.14.1).

D) Immunofluorescence image corresponding to APOBEC2 (A2) staining on shGFP control myoblasts cells differentiating for 5 days. The image corresponds to the same image in (A). A pseudocolor filter was applied to highlight the differences in APOBEC2 signal intensity in the nucleus and cytoplasm of the myotubes. The pseudocolor (Fire LUT) was applied using Fiji with gray values (signal) ranging from 0 to 255 (Schindelin et al., 2012). Scale bar: 50 μ m.

From these results and analyses, I formed the idea that APOBEC2 directly affects myoblast to myotube differentiation. The results point to a direct correlation between APOBEC2 expression and myotube formation, where the MyHC⁺ cells had visibly higher levels of APOBEC2 expression. The RNA expression data from the early time points of differentiation further support a direct role for APOBEC2 in regulating muscle differentiation. The decrease in APOBEC2 levels resulted in several hundred differentially expressed genes between the knockdown and control cells at time points prior to myotube fusion or formation. Moreover, GO terms from the list of downregulated genes with APOBEC2 knockdown were enriched for muscle developmental terms and muscle physiology-related terms, ribosome biogenesis, and protein metabolism. However, the mechanism behind APOBEC2 regulation of muscle differentiation was still unknown. I decided to pursue this further since it might reveal the molecular function of APOBEC2.

Moreover, I observed APOBEC2 was enriched in the nuclei of the differentiated myotubes and even in individual MyHC⁺ cells (**Figure 3.6.D**). I applied a pseudocolor to highlight the difference between APOBEC2 protein levels, as shown by signal intensity, in the nucleus and cytoplasm of the myotubes. APOBEC2 protein was highly enriched in the nuclei of the myotubes. This observation encouraged me to pursue the molecular function of APOBEC2 within the myotube nuclei.

3. Results

3.2 APOBEC2 and transcriptional control

The observed nuclear enrichment of APOBEC2 within the myotubes suggested a direct role of APOBEC2 in transcriptional regulation during muscle differentiation. Beyond a transcriptional role, others have also hypothesized functions for APOBEC2 within the nucleus, such as deamination of methylated cytosines leading to DNA demethylation (Guo et al., 2011; Rai et al., 2008). However, others have not found evidence for DNA demethylation activity for APOBEC2, or the other APOBEC family members (Nabel et al., 2012; Powell et al., 2013). Nevertheless, based on the conserved structure of the cytidine deaminase domain, I pursued the hypothesis that APOBEC2 interacted with nucleic acids within myotube nuclei to directly affect transcriptional control of muscle differentiation.

3.2.1 APOBEC2 occupied promoters of genes related to cell differentiation

Using APOBEC2 chromatin occupancy data left to me by Linda Molla, I sought to find evidence for a direct role of APOBEC2 in transcriptional regulation during myoblast to myotube differentiation. The data came from a chromatin immunoprecipitation experiment where endogenous APOBEC2 was pulled down and the associated DNA was sent for massive parallel sequencing (ChIP-Seq). This experiment allowed me to determine potential APOBEC2 regulated chromatin regions under the assumption that APOBEC2 was acting as a transcriptional regulator in muscle differentiation.

ChIP-Seq was performed at time points, 14 hours (h) and 34 h, in between the RNA-Seq time points, under the assumption that it was acting as a transcription factor. From the ChIP-Seq data, several hundred peaks, APOBEC2 binding sites, were called with almost two times the number of peaks at the later time point, total peaks: 14 h = 1059 vs 34 h = 2125 (**Figure 3.7.A,B**). The majority of APOBEC2 binding sites were found on promoter regions, defined as +/- 2000 base pairs (bp) from the transcription start site (TSS) (**Figure 3.7.C**). Interestingly, genome feature annotations for the later time point (34 h) were composed of more varied annotations, with only 70% of binding sites assigned to promoter regions versus almost 90% at 14 h. These binding events outside promoter regions (TSS +/- 2000 bp) may reflect binding at regulatory enhancer regions.

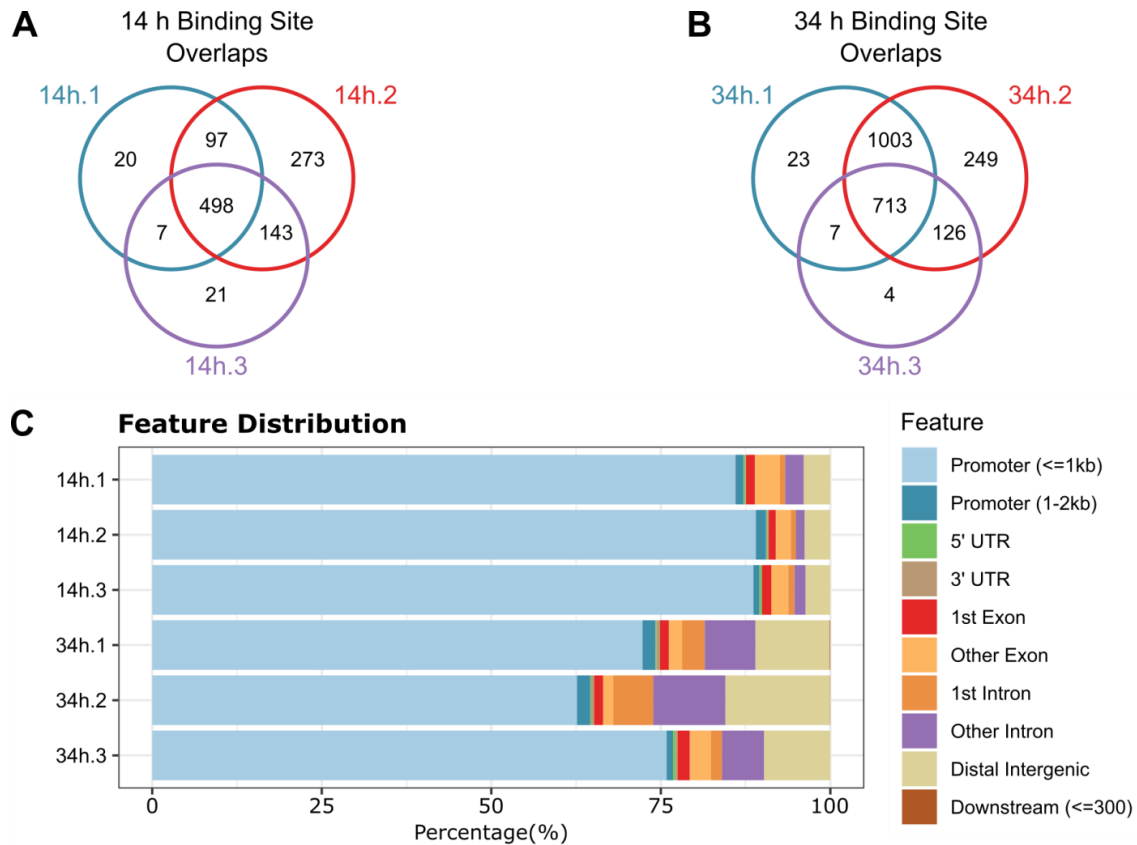


Figure 3.7. APOBEC2 ChIP-Seq binding sites at 14 and 34 hours into myotube differentiation.

A, B) The Venn diagrams depict the number of APOBEC2 binding sites and their overlaps between replicates of 14 h (14.1 - 14.3) and 34 h (34.1 - 34.3). APOBEC2 binding sites came from peak calling analysis (MACS2) from ChIP-Seq data.

C) APOBEC2 binding sites (called from the ChIP-Seq data) were annotated against mouse genomic features (UCSC mm10). Promoters were defined as +/- 2 kilobases (kb) from the transcription start site (TSS). Each bar plot represents the percentage of binding sites within the genomic features for each biological replicate. Plots were generated with the ChIPSeeker package in R.

I chose to focus my analysis on APOBEC2 ChIP-Seq peaks that increased in signal at the 34-h time point compared to the 14-h time point. I chose this strategy in defining APOBEC2 ChIP-Seq peaks since, beyond just an increase in the number of binding sites at the later time point, there was also an increase in signal at the binding sites with the 34 h samples (**Figure 3.8.A**). Moreover, I expected that biologically relevant APOBEC2 occupancy should increase with increasing APOBEC2 protein levels. Using the DiffBind package (version 3.2.4) in R, I compared the read counts, a measure of signal, at consensus peaks from the 3 replicates at the 14 h and 34 h time points. There were 969 peaks (of 2168 consensus peaks) that had significantly different signals between 14 h and 34 h ($FDR \leq 0.05$). Most (964 of 969) of the differentially bound sites had increasing signal at 34 h; conversely, sites with decreasing signal from 14 h to 34 h were

3. Results

considerably less, only 5 of 969. Moreover, more than 50% of the differentially bound APOBEC2 binding sites fall within promoter regions which supported the idea that APOBEC2 was involved in gene regulation (**Figure 3.8.B**). This set of differentially bound APOBEC2 regions, which I deemed biologically relevant, was the set I studied for the rest of the analyses.

The set of differentially bound APOBEC2 regions were annotated to the nearest genes, using the ChIPSeeker package in R with the mouse genome (mm10 UCSC) transcript database. This was done to uncover which genes were potentially regulated by APOBEC2 during muscle differentiation. From this set of genes, a gene ontology analysis for biological processes was done to group the genes into meaningful biological pathways. The gene ontology analysis was visualized with a cluster map to see clusters of interest (**Figure 3.8.C**). Two clusters of terms were particularly interesting: muscle cell differentiation cluster and myeloid cell differentiation cluster (highlighted in yellow, **Figure 3.8.C**). The enrichment of terms related to muscle cell differentiation implicated APOBEC2 directly in regulating muscle differentiation, supporting the hypothesis that APOBEC2 was acting as a transcriptional regulator. However, it was notable that terms regarding cardiac cell differentiation were enriched rather than striated skeletal muscle differentiation. This was unlike the terms that were enriched for the downregulated genes when APOBEC2 was knocked down (**Figure 3.4.A,B**). Moreover, the cluster with terms including myeloid cell differentiation and regulation of hemopoiesis caught my attention since this seemed related to upregulated genes related to immune cell activation when APOBEC2 was knocked down during differentiation (**Figure 3.5.A,B**). These were confusing results since I was expecting biological processes related to skeletal muscle differentiation given that I was using a model of skeletal muscle differentiation.

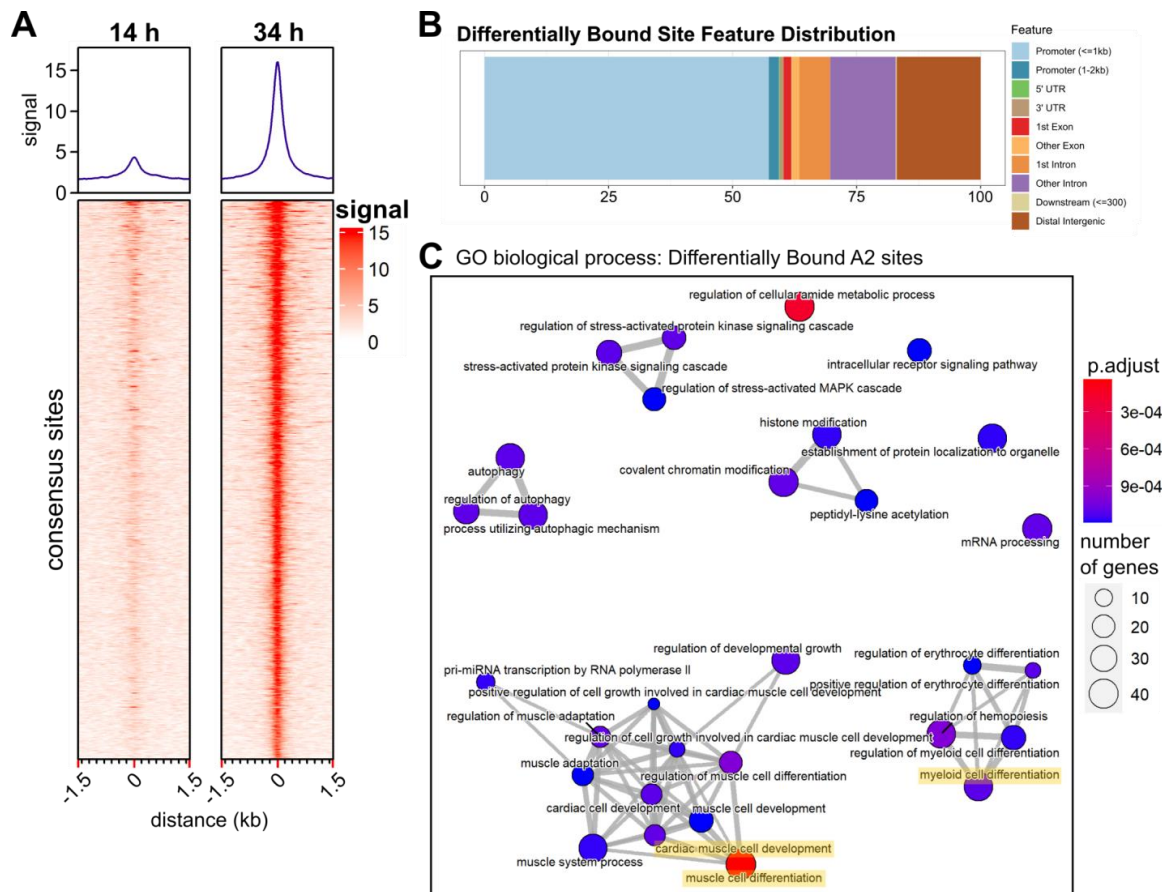


Figure 3.8. Differentially bound APOBEC2 binding sites at 14 and 34 hours into C2C12 myotube differentiation.

A) Profile of signal within APOBEC2 binding sites at 14 h and 34 h time points. For the heatmap, each line of the y-axis defines an APOBEC2 binding site. Signal was calculated from normalized read counts falling within 1.5 kilobase pairs (kb) upstream and downstream of each binding site center. A summary plot showing the average signal across all sites is present above the heatmaps. A total of 2168 consensus sites, defined as binding sites called in at least 2 samples across both time points (2 out of 6 samples). Plots were generated with the DiffBind package.

B) Differentially bound sites between 34 h and 14 h timepoints were annotated against mouse genomic features (UCSC mm10). Promoters were defined as +/- 2 kilobases (kb) from the transcription start site (TSS). The plot was generated with the ChIPSeeker package.

C) Cluster map of enriched GO terms from differentially bound genes between 14 h and 34 h ChIP-Seq time points. The colors of the dots correspond to p.adjusted values, where the red to blue gradient corresponds to low to high values. The sizes of the dots directly correspond to k, the number of genes from the provided list that fall within the corresponding GO term. Edges connect overlapping GO terms; overlap is determined by genes within the terms.

The data started to make sense when, by chance, I attended a lecture by Dr. Moritz Mall. He talked about his work on a terminal repressor found in neurons that supports neuronal differentiation through repression of alternate competing cell fates (Mall et al., 2017). Myt1l was expressed during neuron differentiation yet, as a transcriptional regulator, it was targeting other non-neuronal cell fates for repression. This led me to

3. Results

another paper that discussed how the muscle lineage-specific transcription factor Myod1 could induce spurious differentiation to non-muscle lineages without other repressive factors (Lee et al., 2020). These lines of thought guided me to the idea that APOBEC2 was functioning as a regulator of non-muscle lineages during skeletal muscle cell differentiation given that it was occupying non-skeletal muscle-related genes. Moreover, since APOBEC2 was also expressed in cardiac muscle, it may play a role in determining skeletal versus cardiac muscle cell fate.

3.2.2 APOBEC2 occupied genes changed in gene expression during myoblast differentiation

To confirm if APOBEC2 was a repressor or activator of its target genes, I first examined whether the APOBEC2 target genes (A2 occupied) were biased towards up- or downregulated genes during control (shGFP) myotube differentiation based on the global gene expression changes in the cells at the early time points of differentiation, day 0 to day 2. I assigned the respective fold change values ($\log_2\text{FoldChange}$) from the RNA-Seq differential expression comparisons from the APOBEC2 knockdown and control cells to the APOBEC2 target genes, which was based on the ChIP-Seq analysis. A shift in the mean or median fold changes of the APOBEC2 target genes to either negative or positive values would suggest whether it was either a repressor or activator.

There was no significant shift towards either upregulated (\log_2 fold change > 0) or downregulated genes (\log_2 fold change < 0) for the APOBEC2 target genes during myoblast differentiation (day 1 vs day 0 and day 2 vs day 0, occupied vs unoccupied, **Figure 3.9.A,B**). The mean or median values of the \log_2 fold changes of APOBEC2 target genes at either day 1 or day 2 of differentiation were close to 0 and not significantly different against the background, unoccupied set of genes. I was expecting a shift towards downregulated genes if APOBEC2 was repressing other lineages during myoblast differentiation.

I then checked how the knockdown of APOBEC2 affected the gene expression of the APOBEC2 target genes (**Figure 3.9.C,D**). There was a slight shift towards downregulated genes for the APOBEC2 target genes when APOBEC2 was knocked down (**Figure 3.9.D**). However, the shift in mean \log_2 fold change was very small and close to 0; and thus, may not be biologically meaningful (mean shA2 vs shGFP: -0.0025 vs 0.004, $p = 0.01$ Student t-test).

These results did not specify whether APOBEC2 was a transcriptional repressor or activator of its target genes. Nonetheless, APOBEC2 occupied genes changed in gene expression during myotube differentiation and with APOBEC2 knockdown. To uncover if APOBEC2 was involved in regulating the expression of the non-muscle lineage-related genes, I proceeded to investigate how these specific genes changed in expression during myoblast differentiation in the control and APOBEC2 knockdown cells.

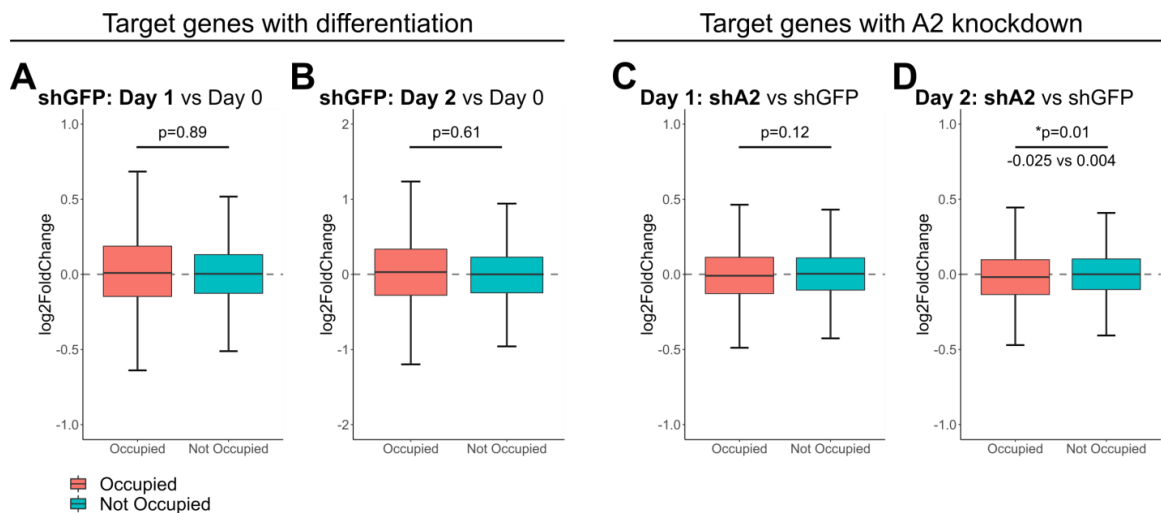


Figure 3.9. Gene expression changes of APOBEC2 occupied genes during myotube differentiation, and with APOBEC2 knockdown during differentiation.

A, B) Boxplots showing the distribution of log₂ fold changes of gene expression (RNA-Seq counts) between shGFP day 1 vs day 0 (A) and day 2 vs day 0 (B).

C, D) Boxplots showing the distribution of log₂ fold changes of gene expression (RNA-Seq counts) between APOBEC2 knockdown (shA2) and control (shGFP) cells On day 1 (C) and day 2 (D). Occupied genes represent APOBEC2 target genes taken from the DiffBind analysis in section 3.2.1. Not Occupied genes are APOBEC2 non-target genes that represent the background set of differentially expressed genes for each comparison. The horizontal line represents the median, lower and upper edges of the box represent the first and third quartiles, and whiskers represent the 1.5 x interquartile range. P values were taken from Student t-tests.

3.2.3 APOBEC2 regulated non-skeletal muscle-related genes during skeletal muscle differentiation

I examined what APOBEC2 target genes might be regulated by APOBEC2 during C2C12 myotube differentiation. From the list of APOBEC2 target genes, I selected the genes that changed in gene expression, either activated (upregulated) or repressed (downregulated), with APOBEC2 knockdown during myotube differentiation (shA2 vs shGFP). I selected the genes with gene expression changes that met the statistical cutoff, DESeq2 p-adjusted value < 0.05, at any of the time points, day 0, day 1, or day 2. A total of 189 of 969 APOBEC2 target genes met these criteria.

3. Results

I used the mean of the transformed normalized counts of the differentially expressed target genes to produce a heatmap to visualize the changes in gene expression for each of the genes (**Appendix I**). The heatmap did not show that the target genes were biased towards repressed or activated genes during myotube differentiation, or when APOBEC2 was knocked down. I decided to perform a GO enrichment analysis in the list of differentially expressed APOBEC2 target genes to determine which biological processes could be affected by APOBEC2. The GO analysis pointed to biological processes involving cell differentiation; in particular, glial cell differentiation, regulation of angiogenesis, and T cell activation (**Figure 3.10.A**). These differentiation processes were non-muscle related which supported the idea that APOBEC2 regulated these non-muscle related processes during myotube differentiation.

I decided to focus my analysis on these non-muscle-related differentiation processes. A significant portion of the APOBEC2 target genes within these biological processes were transcriptional regulators (22 genes out of 189). Using a heatmap, the gene expression changes of these target genes could be clustered into 2 main groups (**Figure 3.10.B**). One group were the genes that were activated during differentiation (shGFP day 0 to day 2) but had repressed expression with APOBEC2 knockdown. The gene expression changes in this first group suggested that these were APOBEC2 activated genes. The other group was made of genes that had lower expression in the shGFP cells (day 0 to day 2) while having higher expression with APOBEC2 knockdown. This suggested that these genes were APOBEC2 repressed genes. These APOBEC2 related gene expression changes indicated that APOBEC2, directly or indirectly, was regulating these transcriptional regulators involved in non-muscle cell fates during myotube differentiation.

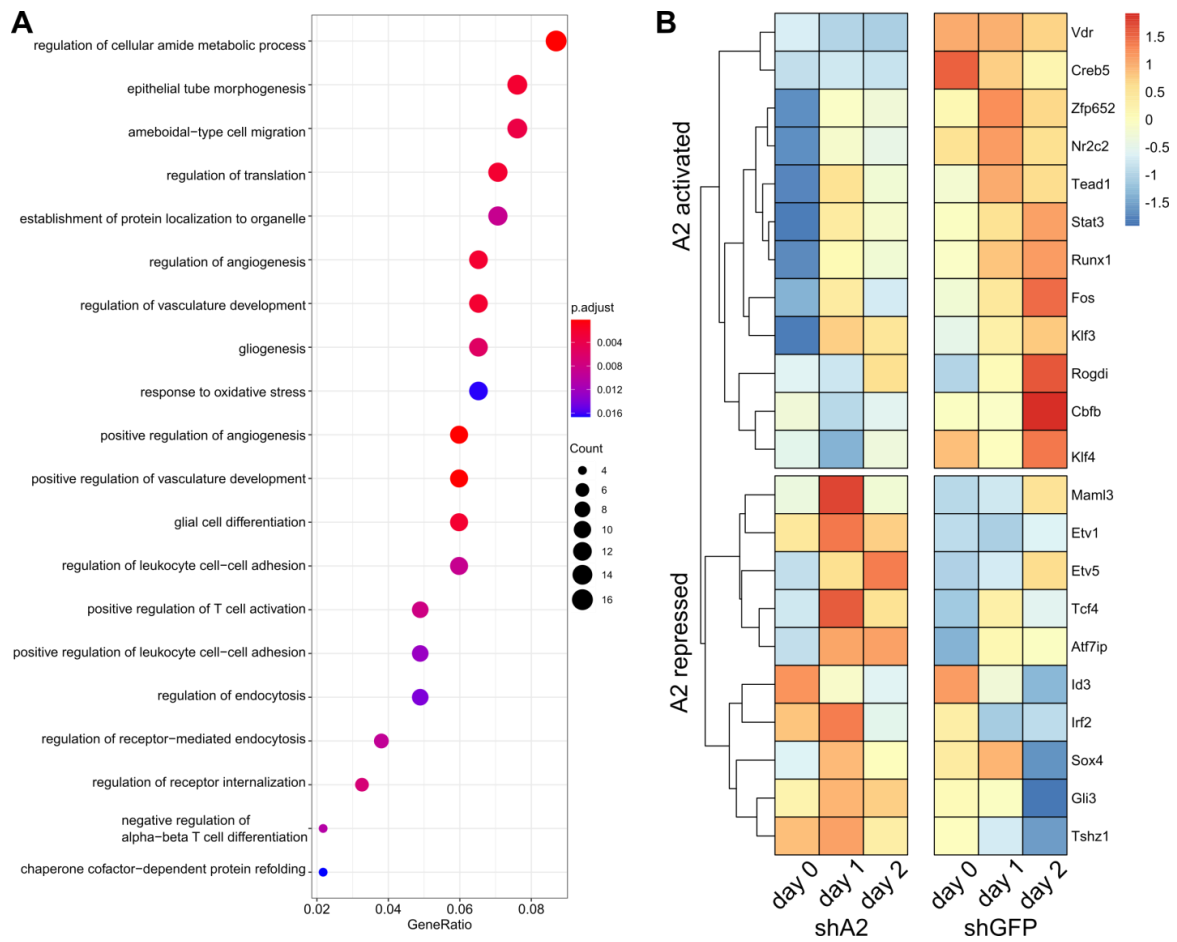


Figure 3.10. Gene ontology and heatmap of differentially expressed APOBEC2 target genes.

A) Dot plot representing enriched gene ontology (GO) terms from the list of differentially expressed APOBEC2 target genes. Top 20 GO terms ranked by p.adjusted values from GO over-representation analysis done using clusterProfiler. The colors of the dots correspond to p.adjusted values, where the red to blue gradient corresponds to low to high values. The sizes of the dots directly correspond to Gene Ratio. Gene Ratio corresponds to k/n where k is the number of genes from the provided list that fall within the corresponding GO term and n is the total number of genes in the provided list.

B) Heatmap of a curated list of transcriptional regulators from the gene lists of the tissue differentiation/development-related GO terms from (A). Gene expression values were taken from the RNA-Seq experiment from APOBEC2 knockdown (shA2) and control shGFP (shG) C2C12 myoblasts at corresponding time points of differentiation, day 0 to day 2. Heatmap colors correspond to mean normalized read counts scaled by row (z-score). Heatmap was prepared using the pheatmap package in R (Kolde, 2019).

This was curious since why would APOBEC2 regulate, even activate, several transcriptional regulators related to non-muscle cell fates. These transcription factors and cofactors, such as Stat3, Runx1, Gli3, Sox4, and Id3, however, have pleiotropic transcriptional roles in several cell fates. Moreover, as I mentioned earlier, even pioneer transcription factors such as Myod1 have promiscuous targets that are offset by transcriptional repressors or other epigenetic marks to ensure proper fate determination

3. Results

(Lee et al., 2020). Perhaps APOBEC2 acted as a repressor on the repressed non-muscle transcription factors, while it also acted as an activator of pleiotropic transcription factors.

The decrease in myotube formation with APOBEC2 knockdown might be due to the disruption of the expression of the transcriptional regulators involved in cell differentiation. The loss might have caused repression of genes necessary for myotube formation, or even the activation of spurious non-muscle transcriptional programs. The upregulation of genes involved in kidney, neural tube, and immune cell differentiation seen in the APOBEC2 knockdown cells reflected this (**Figure 3.5**). I next embarked on determining the molecular mechanism behind the role of APOBEC2 as a transcriptional regulator during muscle differentiation.

3.3 APOBEC2 and corepressor complexes

The prior results showed that APOBEC2 was necessary for myoblast to myotube differentiation. More specifically, there were indications that APOBEC2 was acting as a transcriptional regulator of genes involved in biological pathways necessary for proper muscle differentiation. Thus, I aimed to determine how APOBEC2 could act as a transcriptional regulator.

I sought to identify which proteins APOBEC2 interacted with within the nuclear compartment of the cell, especially since APOBEC2 was enriched in the nucleus of the myotubes (**Figure 3.6.D**). It turned out that since APOBEC2 was long considered a dead deaminase, it had been used by other groups working on the APOBEC family as a negative control in their experiments. One such experiment was a protein-protein interactome experiment where APOBEC2 was fused to a biotin ligase to determine which proteins are within proximity of APOBEC2 in living cells – a technique called BioID. The biotinylated proteins were extracted and enriched from the living cells and ran through mass spectrometry for identification and quantification. The APOBEC2 BioID data from our collaborator Dr. Javier Di Noia showed that APOBEC2 interacted with several cytoplasmic and nuclear proteins. However, among several clusters, it showed that APOBEC2 interacted with a cluster of proteins that centered on histone deacetylase 1 (HDAC1) (**Figure 3.11**). More interestingly, the cluster of proteins had several members that belong to HDAC1 corepressor complexes involved in transcriptional regulation (Kelly & Cowley, 2013). Though the data came from biotin ligase-tagged APOBEC2 overexpressed in mouse B cells and human HEK293T cells, it gave me a good initial candidate protein to verify in muscle cells.

3. Results

BioID: APOBEC2-HDAC1 co-repressor complex

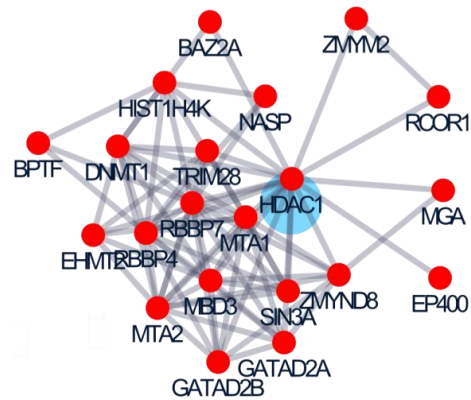


Figure 3.11. APOBEC2 protein interactome from BioID experiment performed in HEK293T cells and human B cells. HDAC1-co-repressor complex cluster. The red dots represent the proteins within close taggable proximity of the APOBEC2 biotin-ligase fusion protein. The edges represent the connection between the proteins based on known protein-protein interactions. This figure was provided by the group of Javier Di Noia who performed the experiment.

3.3.1 APOBEC2 was found in the chromatin fraction of cells

To determine whether APOBEC2 interacted with HDAC1 corepressor complexes within muscle cells at the physiological protein levels, I performed co-immunoprecipitation (co-IP) experiments using endogenous APOBEC2 in differentiated muscle cells. I first had to obtain a nuclear protein fraction enriched for APOBEC2 from differentiated C2C12 cells. This required several iterations of protocol optimization since APOBEC2 was found in the cytoplasm, and both the soluble and insoluble nuclear fractions (**Figure 3.12.A**). Alpha-tubulin was used as a marker for the cytoplasmic fraction; while the transcription factor specificity protein 1 (SP1) was used as a marker for the nuclear fraction. Additionally, histone H3 was used as a marker for chromatin. Confirming the presence of APOBEC2 within the insoluble nuclear fraction together with histones supported the role of APOBEC2 in chromatin. I used a protocol with a high salt nuclear lysis buffer (800 mM NaCl) and nuclease-treatment to separate chromatin-bound proteins from genomic DNA (specifics of the protocol could be found in the methods section). This allowed me to have better yields of nuclear APOBEC2 together with other proteins found in the soluble and insoluble (chromatin) nuclear fractions (**Figure 3.12.B**). This also showed that APOBEC2 and HDAC1 could both be found in the nuclear fraction. Thus, I proceeded with the co-IP experiment.

3. Results

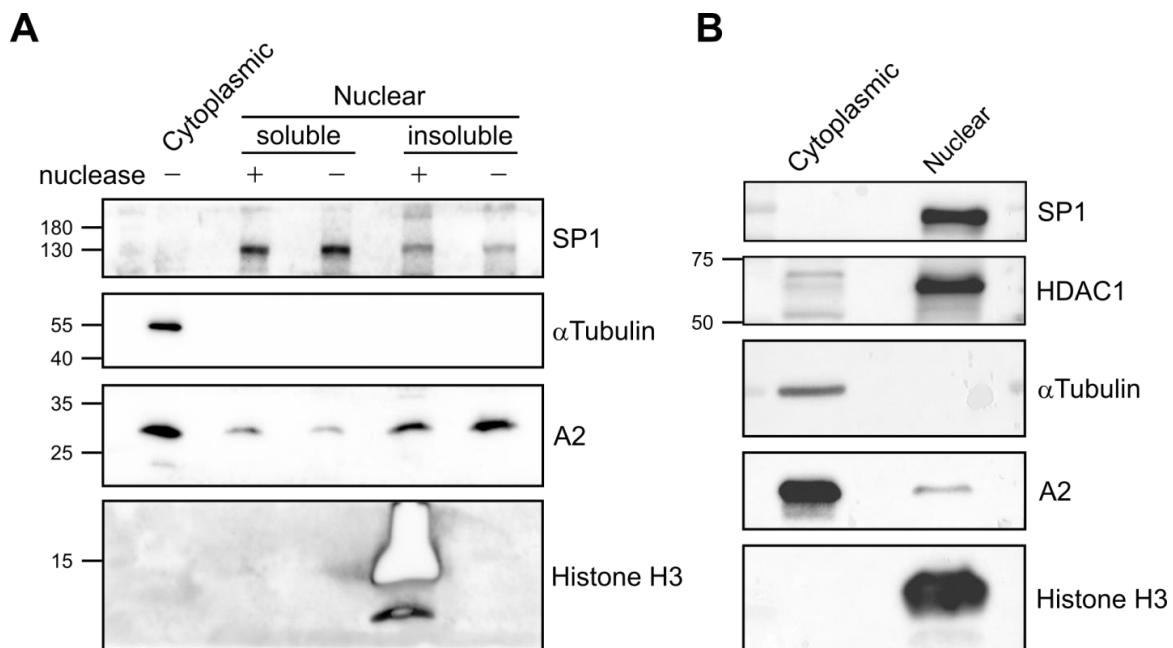


Figure 3.12. APOBEC2 protein expression within differentiated C2C12 myotube protein lysate fractions.

A) Western blot from protein lysates prepared using the Thermo NE-PER kit. Fractions were either treated (+) or not treated (-) with nuclease (benzonase). Blots were stained for specificity protein 1 (SP1), alpha (α)Tubulin, APOBEC2 (A2), and histone H3. Histone H3 was overexposed but did not affect the interpretation.

B) Western blot from protein lysates prepared using the high salt nuclear lysis protocol. Fractions were from the optimized protocol to separate the cytoplasmic and total (soluble and insoluble) nuclear fractions. Blots were stained for specificity protein 1 (SP1), histone deacetylase 1 (HDAC1), alpha (α)Tubulin, APOBEC2 (A2), and histone H3.

3.3.2 APOBEC2 interacted directly with histone deacetylase HDAC1

Using the nuclear fraction, I immunoprecipitated APOBEC2 using an APOBEC2-specific antibody bound to magnetic beads. I then eluted APOBEC2 together with its potential interactors. To optimize the salt concentration of the co-IP buffer, I tested a range of salt concentrations since protein interactions are affected by ionic strength, where higher salt concentrations often destabilize protein-protein interactions. I chose salt concentrations from 75 to 150 mM NaCl to be within the physiological salt concentration of approximately 150 mM salt. From these experiments, I confirmed that APOBEC2 directly interacted with HDAC1 (**Figure 3.13.A,B**). The interaction was specific since no discernable signal could be found in the isotype control IgG IPs, which served as a negative control. Furthermore, the interaction between APOBEC2 and HDAC1 decreased with increasing salt concentration; the signal for HDAC1 co-IP was discernable up to 100

mM NaCl (**Figure 3.13.A**). I settled on 75 mM NaCl where I could get a strong signal for HDAC1 co-IP and reproducible results (**Figure 3.13.B**).

I also tested the complementary co-IP where I instead immunoprecipitated HDAC1 with an anti-HDAC1 antibody (**Figure 3.13.C**). With this configuration, I was unable to co-IP APOBEC2. This may be due to the much greater abundance of HDAC1 than APOBEC2 within the nuclear fraction, where there might have been only a small fraction of total HDAC1 bound to APOBEC2. Another explanation could be that the anti-HDAC1 antibody interfered with the interaction surfaces between APOBEC2 and HDAC1.

I further validated the co-IP by using C2C12 cells exogenously expressing Flag-tagged APOBEC2 so I could use a Flag-specific antibody for immunoprecipitation (**Figure 3.13.D**). I used total nuclear protein lysates from differentiated myoblasts either expressing the Flag-tagged APOBEC2 (Flag-A2) or just the pMX-GFP vector (GFP). Input from Flag-A2 stained for anti-A2 showed 2 bands that correspond to Flag-A2 and endogenous APOBEC2. Using the anti-Flag antibody for IP, the heavier Flag-A2 was immunoprecipitated specifically from only the Flag-A2 input. Importantly, APOBEC2 and HDAC1 co-IP was successful using the anti-Flag antibody. Using a different antibody further confirmed the interaction between A2 and HDAC1.

3. Results

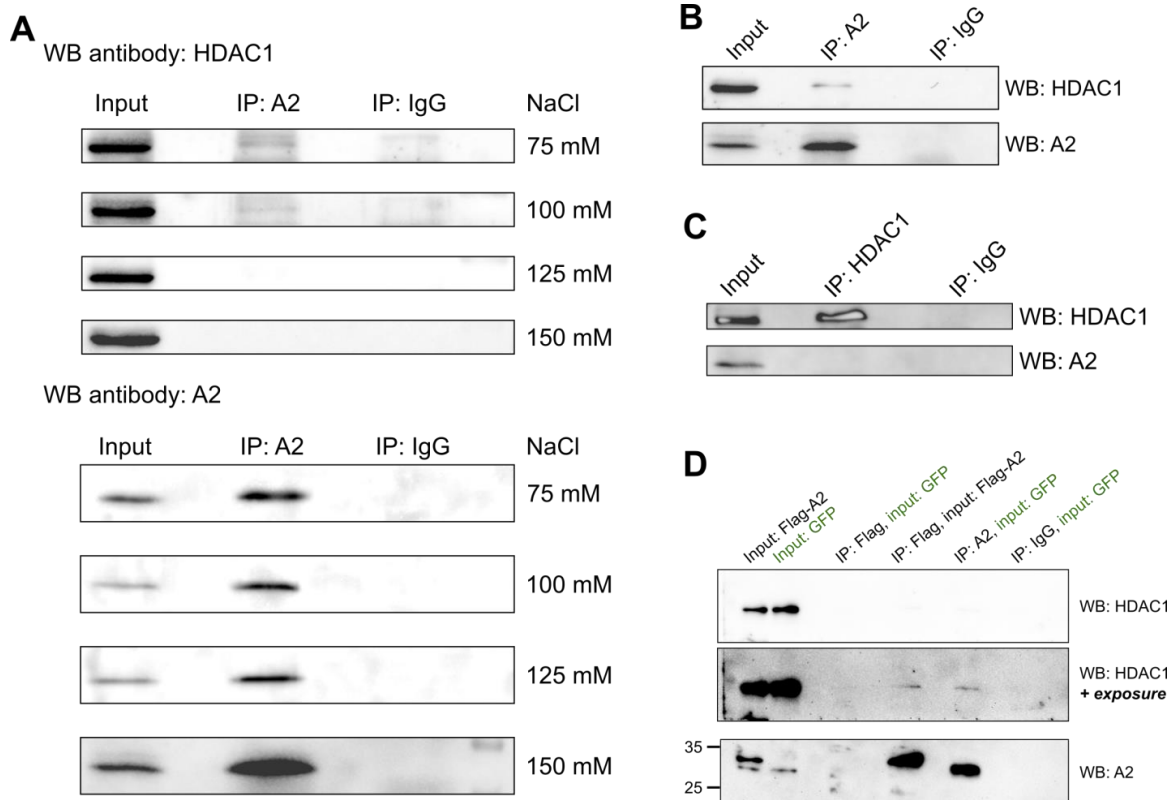


Figure 3.13. Co-immunoprecipitation experiments of APOBEC2 with HDAC1 in the nuclear fraction of differentiated C2C12 myotubes.

A) Western blots (WB) showing co-immunoprecipitation (co-IP) experiments performed with total nuclear protein lysate from differentiated C2C12 myotubes. Co-IPs were performed with 75 to 150 mM NaCl in the co-IP buffer (see methods). Upper 4 blots used anti-HDAC1 for WB; lower 4 blots used anti-APOBEC2 (A2) for WB. Input represents 5% (12.5 μ g) of the total nuclear protein lysate (250 μ g) used for the experiment. Each immunoprecipitation (IP) was done with either anti-APOBEC2 (A2) or an isotype control IgG. Blots for 150 mM NaCl were taken from a separate image but does not affect the interpretation of the results.

B) WB from a co-IP experiment using anti-A2 antibody with the total nuclear fraction from differentiated C2C12 myotubes as input. The experiment was performed at 75 mM NaCl with similar conditions to (A). Labels are similar to (A). APOBEC2 and HDAC1 blots are representative of >3 trials.

C) WB from a co-IP experiment using anti-HDAC1 antibody with the total nuclear fraction from differentiated C2C12 myotubes as input. Experiment conditions were similar to (A) and (B). Labels are similar to (B). HDAC1 WB signal was oversaturated but does not affect the interpretation.

D) WB from a co-IP experiment using anti-HDAC1 antibody with the total nuclear fraction from differentiated C2C12 myotubes expressing amino(N)-terminal Flag-tagged APOBEC2 as input. Experiment conditions were similar to (A) and (B). Two different inputs were used for comparison. Input: Flag-A2 was from C2C12 cells expressing amino(N)-terminal Flag-tagged A2, while Input: GFP was from C2C12 cells expressing the empty vector, pMX-GFP.

The results of the co-IP experiments revealed the direct interaction between APOBEC2 and HDAC1, a transcriptional corepressor. HDAC1 represses the transcription of genes through the deacetylation of histone tails in nucleosomes (Park & Kim, 2020). HDAC1 does not have inherent DNA binding activity but acts together with other proteins within corepressor complexes to target specific genes (Kelly & Cowley, 2013). I aimed to

identify other proteins within the C2C12 muscle cell context that APOBEC2 interacted with to pinpoint the HDAC1 corepressor complex APOBEC2 interacts with in differentiated C2C12 myotubes. A collaborator, Jana Ridani from Dr. Javier DiNoia's group performed the BioID in the differentiated C2C12 cells. This experiment confirmed the interaction between APOBEC2 and HDAC1 but, importantly, identified a specific corepressor complex – CHD4-HDAC1 corepressor complex. This was interesting since CHD4 was demonstrated to play a role in skeletal versus heart muscle cell fate determination (Gómez-Del Arco et al., 2016). These findings placed APOBEC2 directly with a transcription regulatory complex. Since APOBEC2 was found at both activated and repressed genes during myotube differentiation, APOBEC2 interaction with HDAC1 might regulate or specify HDAC1 corepressor activity at APOBEC2 target genes. This molecular interaction provided an epigenetic mechanism for the observed transcriptional changes in the APOBEC2 knockdown C2C12 myoblasts.

3.3.3 *APOBEC2 amino-terminal loss disrupted nuclear localization*

I next sought to uncover how APOBEC2 interacts with HDAC1. Unique among the APOBEC family, APOBEC2 contains an unstructured, 40 amino acid long amino(N)-terminal domain (Krzysiak et al., 2012). This domain is negatively charged and might play a role in protein-protein interactions. I decided to test this by removing this N-terminal domain to produce a truncated variant of APOBEC2, del(1-41)A2. This truncated variant of APOBEC2 was known to fold properly and form the conserved cytidine deaminase domain based on published crystal and solution structures of this truncated APOBEC2 (Krzysiak et al., 2012; Prochnow et al., 2007).

To test whether removal of this N-terminal domain would affect the ability of APOBEC2 to interact with HDAC1 through a co-IP, I overexpressed a carboxy(C)-terminal hemagglutinin (HA)-tagged del(1-41)A2 in C2C12 myoblasts and induced them to differentiate into myotubes. Even though I overexpressed the del(1-41)A2, it was not detectable within the total nuclear protein fraction (**Figure 3.14**). The absence of del(1-41)A2 in the nuclear fraction prevented me from directly testing its interaction with the nuclear HDAC1. For this experiment, I used C-terminal HA-tagged APOBEC2 and del(1-41)A2, since I was planning to use it for co-IP experiments where I could use either the Flag-tag or HA-tag in testing different interactions. The HA-tag did not cause the change in fraction localization for del(1-41)A2 since HA-tagged APOBEC2 was still found within the total nuclear protein fraction.

3. Results

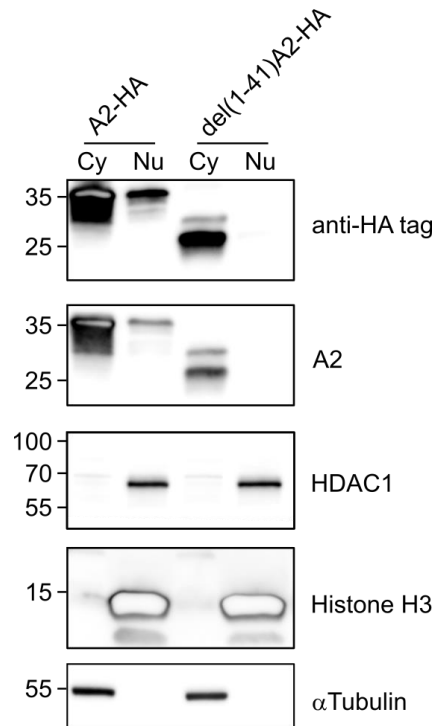


Figure 3.14. N-terminal truncated APOBEC2, del(1-41)A2, localization in cytoplasmic or nuclear fractions from differentiated C2C12 myotubes. Western blot showing localization of proteins within cytoplasmic (Cy) or total nuclear protein (Nu) fractions from differentiated C2C12 myotubes. Carboxy(C)-terminal hemagglutinin (HA)-tagged APOBEC2 and del(1-41)A2 were overexpressed in C2C12 myoblasts through retroviral transduction with the pMX-GFP vector. Blots were stained for the HA-tag, APOBEC2 (A2), histone deacetylase 1 (HDAC1), histone H3, and alpha (α)Tubulin. HA: three repeating HA tags.

This work however produced a variant of APOBEC2 unable to enter the nucleus and, thus, would be incapable of regulating gene transcription directly within the nucleus. To test whether the deleted N-terminal domain was necessary for HDAC1 or other protein-protein interactions, a different approach would be needed where I could test the affinity of the isolated proteins for one another. Furthermore, figuring out how del(1-41) APOBEC2 failed to enter the nucleus would require finding out how APOBEC2 was enriched within nuclei of differentiated myotubes. I decided however to use del(1-41)A2 instead as a control for my next experiments.

So far, the results demonstrated that APOBEC2 was necessary for myotube differentiation. The loss of APOBEC2 in the C2C12 myotube differentiation model led to a decrease in transcription factors, especially ones that are directly involved in muscle differentiation. Furthermore, APOBEC2 directly interacts with the transcriptional corepressor HDAC1. Together these results suggested that APOBEC2 acts as a transcription regulator during myotube differentiation. I next aimed to test if APOBEC2 expression could restore myotube differentiation in the APOBEC2 knockdown cells.

3.4 Restoring APOBEC2 in knockdown C2C12 cells

When induced to differentiate, the APOBEC2 knockdown C2C12 myoblasts formed fewer myotubes. This was also reflected in a decrease in muscle markers (MyHC and troponin T, **Figure 3.2**). To complement the A2 knockdown experiment, I reintroduced A2 protein in the knockdown C2C12 myoblasts and tested whether A2 alone would restore differentiation.

3.4.1 Expression of APOBEC2 increased muscle marker expression in APOBEC2 knockdown cells

To reintroduce APOBEC2 in the knockdown C2C12 myoblasts, I employed a retroviral system to overexpress APOBEC2 in the cells. I used an A2 cDNA sequence with several silent mutations on the shRNA target sequence. These silent mutations would disrupt complementarity between the shRNA and APOBEC2 cDNA rendering the mRNA resistant to shRNA-mediated transcript decay. This shRNA-resistant APOBEC2 cDNA (A2r) sequence was inserted into the pMX vector that contained a GFP reporter. The GFP reporter was not targeted by the shRNA sequence within the control shGFP myoblasts. After several iterations to get high transduction and expression of APOBEC2, I settled on a time course for the experiment (**Figure 3.15.A**). Three days before switching to the differentiation medium, I transduced the pMX vectors into the knockdown C2C12 myoblasts (shA2 and shGFP) allowing sufficient time for expression of the reintroduced APOBEC2. I transduced the pMX-A2r vector and an empty pMX vector as a control. Transduction efficiency was similar between the 2 cell lines and 2 vectors as seen in the GFP reporter expression (**Figure 3.15.B**). After 4 days of differentiation and expression of A2r, I could not see a visible increase in myotube formation, the elongated fused cells in the phase-contrast images. There was still a reduction in myotube formation for the shA2 cells. I proceeded to check protein levels of muscle markers to determine if there were indications of restoration of myotube differentiation.

3. Results

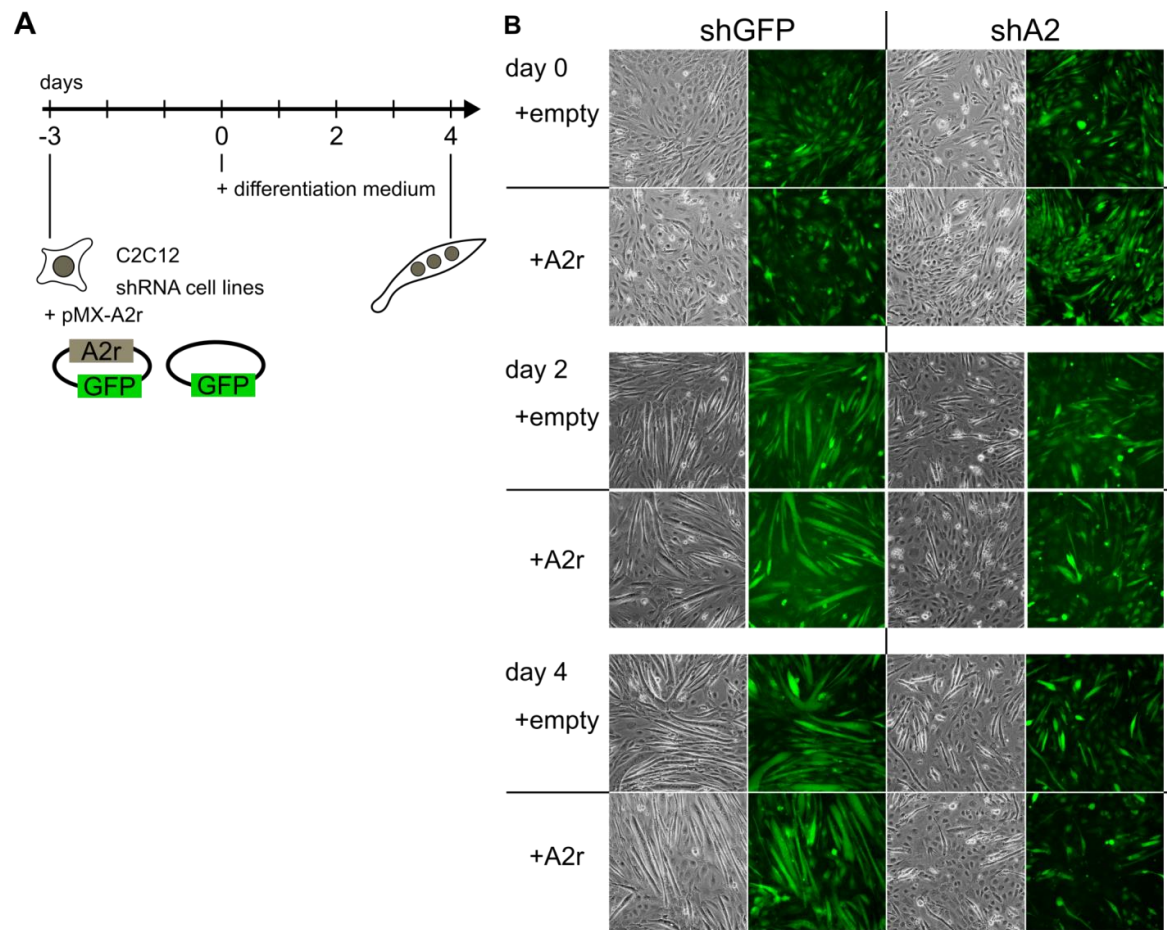


Figure 3.15. APOBEC2 exogenous expression in APOBEC2 knockdown C2C12 cell lines.

A) Time course diagram of APOBEC2 reintroduction into APOBEC2 knockdown C2C12 cell line. On day -3, the knockdown cells were transduced with the vector expressing APOBEC2 cDNA resistant to APOBEC2 shRNA (pMX-A2r) or an empty vector. On day 0, the differentiation medium was added to the cells. On day 4, the experiments were terminated with the formation of the fused myotubes.

B) Side-by-side phase contrast and fluorescence micrographs represent transduction efficiencies in the different cell lines. Green cells represent cells expressing GFP, which was included in the pMX vector. Time points are based on the time course diagram in (A). Magnification 100X.

I obtained total protein lysates from two time points, day 0 and day 4 from when the cells were induced to differentiate. Through Western blots, I checked muscle marker protein levels, MyHC, and troponin T (**Figure 3.16**). There was an increase in MyHC On day 0 when A2r was expressed in both the shGFP and shA2 cells. There was also an increase in MyHC On day 4 though it was not as visibly obvious. There were also no differences in troponin T levels when A2r was expressed. Regardless, it was clear that APOBEC2 expression resulted in increased MyHC expression, a marker of myotube differentiation. This result complemented the results of the knockdown experiment and

established a direct role for APOBEC2 in myotube differentiation. Next, I tested the del(1-41)A2 variant of APOBEC2 that had limited nuclear localization.

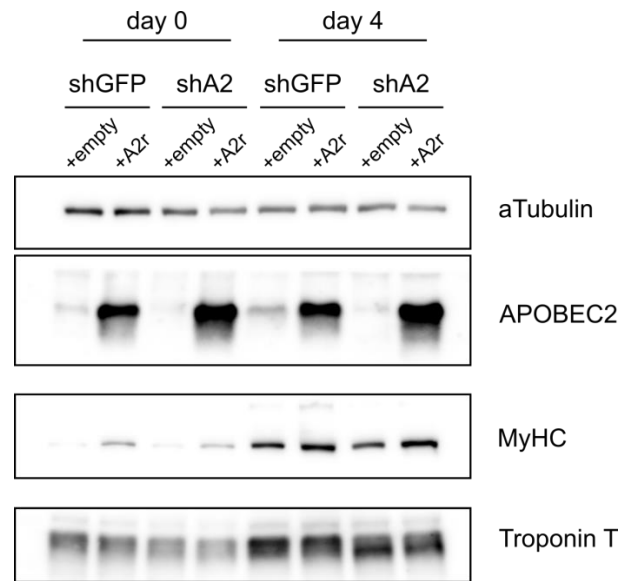


Figure 3.16. Exogenous expression of APOBEC2 in knockdown cells affects muscle marker protein levels. Western blots showing levels of APOBEC2 (A2), myosin heavy chain (MyHC), Troponin T, and alpha (α)Tubulin in total protein lysates from APOBEC2 knockdown (shA2) and control (shGFP) C2C12 cells. The cells were expressing either APOBEC2 cDNA resistant to APOBEC2 shRNA (+A2r) or the empty control vector (+empty). α Tubulin was used as a protein loading control. The lysates were collected from cells On day 0 and day 4 based on the time course in Figure 3.15.A. This blot is representative of three independent trials.

3.4.2 Variants of APOBEC2 failed to restore myotube differentiation

I produced the pMX-del(1-41)A2r vector to introduce truncated APOBEC2 (resistant to shA2) into the C2C12 myoblasts. For this experiment, I chose time point day 0, right before inducing myotube differentiation, since this gave a more visually distinguishable result than day 4. The results however were not as distinguishable as I expected, so I had to rely on quantifying the normalized MyHC signal with α Tubulin as the reference protein across 3 independent trials (**Figure 3.17.A**). For the shA2 cells, there were visible and quantifiable differences with the treatments (**Figure 3.17.B**). Comparing the MyHC signal across the shA2 cells, there was almost 70% more MyHC protein in the shA2 cells expressing A2r than the empty vector (normalized signal = 0.36 ± 0.03 vs 0.21 ± 0.01 , mean \pm standard deviation, $p < 0.001$). However, for the del(1-41)A2 expressing shA2 cells (shA2 + dA2r), there were almost equal levels of MyHC to the empty vector control (0.24 ± 0.01 vs 0.21 ± 0.01 , $p = 0.273$). The differences in MyHC expression in the shGFP cells were not as apparent compared to shA2 cells. There was still a statistically significant difference between the shGFP cells expressing A2r than the empty vector, but

3. Results

there was still no difference between cells expressing dA2r and empty control (**Figure 3.17.C**). The differences in MyHC expression could not have been due to APOBEC2 and variant protein expression levels since these were similar.

From these results, it was clear that wildtype APOBEC2 increased MyHC protein levels in the APOBEC2 knockdown cells while the del(1-41)A2 variant was unable to increase MyHC protein levels. Given that del(1-41)A2 was not found in the total nuclear fraction, this result indicates that the role of APOBEC2 in myotube differentiation requires nuclear localization. This supported the hypothesis that APOBEC2 functioned as a transcription regulator during myotube differentiation.

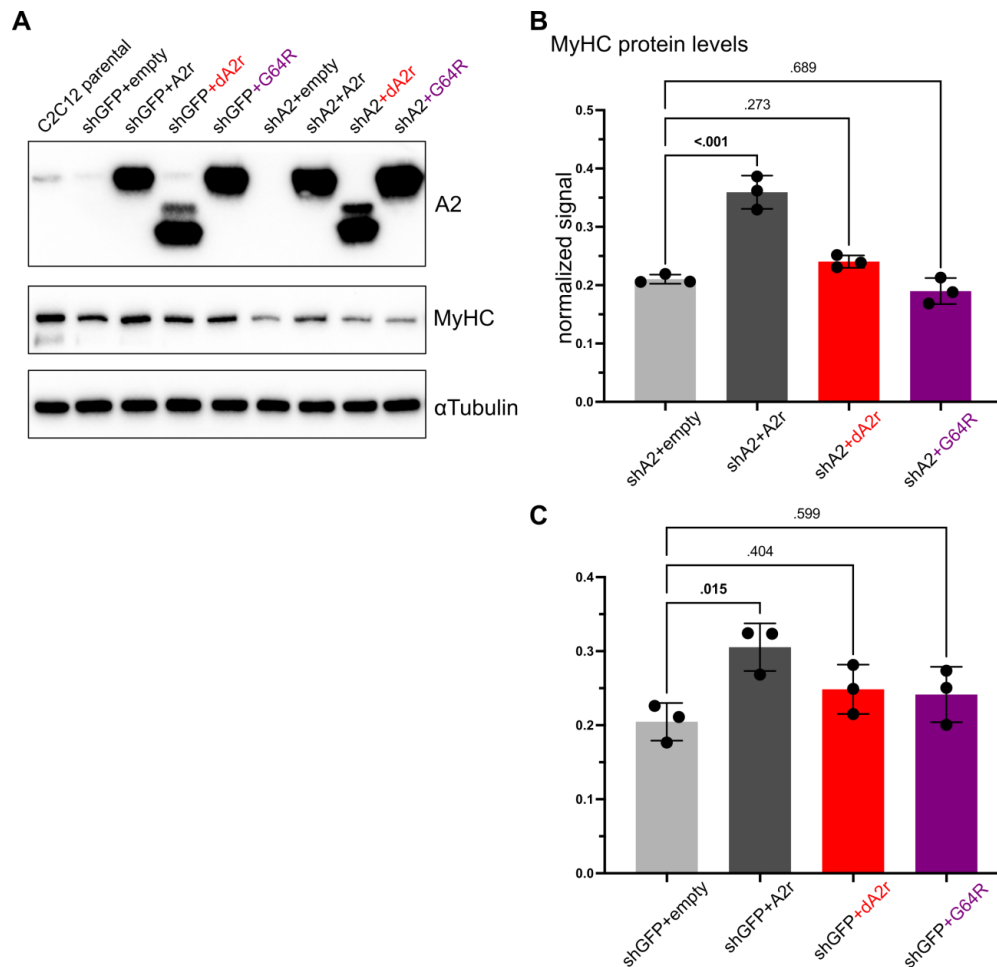


Figure 3.17. Exogenous expression of APOBEC2 and variants in C2C12 myoblasts.

A) Western blots showing levels of APOBEC2 (A2), myosin heavy chain (MyHC), and alpha (α)Tubulin in total protein lysates from APOBEC2 knockdown (shA2) and control (shGFP) C2C12 cells. The cells were expressing either empty control vector (+empty) or APOBEC2 cDNA resistant to APOBEC2shRNA – wildtype mouse APOBEC2 (+A2r), del(1-41)A2 (+dA2r), or the APOBEC2 G64R variant (+G64R). α Tubulin was used as a protein loading control. The lysates were collected from differentiated myotubes On day 0 based on the time course presented in **Figure 3.15**. This blot is representative of three independent trials.

B,C) Bar plots of MyHC protein levels in shA2 (B) and shGFP (C) cells expressing APOBEC2 and variants. MyHC protein levels were represented as the normalized signal, where MyHC signal intensity was normalized to α Tubulin signal intensity and then the variances across the three replicates were reduced by normalizing to the total MyHC signal in each corresponding replicate (Degasperi et al., 2014). The height of the bar plots represents the mean of the normalized signal from 3 independent replicates, and error bars represent the standard deviation. Bar plots and statistics were prepared using Prism 9; statistical analysis: one-way ANOVA with multiple pairwise comparisons. Values above each pairwise comparison represent adjusted P values (Bonferroni correction).

In these experiments, I added another potentially consequential variant of APOBEC2, the Gly64Arg (G64R) variant. This mutation was described to us by Dr. Fowzan Alkuraya, a geneticist from Riyadh, who read about our findings on APOBEC2 from the preprint I wrote (**Appendix II**). We had a remote meeting where he described a homozygous mutation found in APOBEC2 in two young patients that presented a lethal

3. Results

myopathy, APOBEC2: NM_006789.3: c.190G>A; (p.Gly64Arg) variant. The consanguineous parents of the patients were heterozygous for the mutation. I took the opportunity to include this G64R variant to evaluate whether this mutation inactivated APOBEC2.

I decided to introduce the human mutation into the mouse APOBEC2 cDNA since mouse and human APOBEC2 were highly conserved (90%). In particular, the mutation was found in loop 1 of APOBEC2, which is identical in both humans and mice (**Figure 3.18**). From the Western blots, it was clear that the G64R was unable to increase the levels of MyHC unlike the wildtype APOBEC2 (**Figure 3.17.A,B**). It had almost equal protein levels of MyHC to the empty control (0.19 ± 0.02 vs 0.21 ± 0.01 , $p = 0.689$). This was an exciting finding that we communicated back to Dr. Alkuraya as the lethal myopathy might be traced back to this mutation of APOBEC2. This G64R mutation was found on a loop that, based on the published structures of other APOBECs, contacts the DNA substrate or ligand (Maiti et al., 2018; Shi et al., 2017). These findings prompted me to explore the ability of APOBEC2 and this variant to interact with DNA ligands.



Figure 3.18. Protein sequence alignment of mouse and human APOBEC2. Protein sequences were retrieved from Uniprot, mouse: Q9WB35, and human: Q9Y235. Loop 1 is between helix 1, marked by the orange bar, and strand 1, marked by the purple bar. Secondary structures are based on APOBEC2 structure, PDB: 2RPZ. The G64 residue found on loop 1 is marked by a red triangle. Alignment was done using Clustal Omega, sequence similarity = 90.6% (203/224 residues).

The results here show that APOBEC2 expression directly causes an increase in MyHC protein levels, a marker of myotube differentiation. Importantly, the truncated APOBEC2 variant, del(1-41)A2 which does not enter the nucleus, was unable to increase MyHC levels even if it was expressed at comparable levels to wildtype APOBEC2. This strongly suggested that APOBEC2 plays a role within the nucleus, potentially as a

transcriptional regulator during myotube differentiation. Moreover, the results indicated that the APOBEC2 G64R mutation, in an APOBEC family conserved DNA-interacting loop, was also unable to increase MyHC protein levels.

I next aimed to demonstrate how exogenous APOBEC2 expression affected gene transcription of APOBEC2 target genes as this would validate the results of the RNA-Seq. And this would further support the hypothesized molecular function of APOBEC2 as a transcriptional regulator.

3. Results

3.5 Exogenous APOBEC2 expression in a fibroblast cell line

To evaluate the effect of APOBEC2 expression on gene transcription, I followed a similar experimental design as I discussed above when I restored APOBEC2 expression in the knockdown cell line. However, even with numerous repeated attempts and experiment adjustments, I was unable to show consistent effects on the target genes in C2C12 cells (**Appendix III**). Instead, I decided to test the effect of APOBEC2 expression in a non-muscle cell line, NIH3T3 mouse fibroblasts, in case transcriptional effects were more pronounced in a different epigenetic context. Perhaps, other transcriptional regulators in the muscle context were confounding the effects of APOBEC2 expression.

3.5.1 Expression of APOBEC2 in NIH3T3 fibroblasts led to downregulation of APOBEC2 occupied genes

I limited my initial experiments to NIH3T3 fibroblasts expressing wildtype APOBEC2. I produced NIH3T3 cell lines stably expressing APOBEC2 or the empty vector, pMX-GFP, for comparison. I picked genes that were occupied by APOBEC2 based on the ChIP-Seq from the C2C12 cells and expressed in NIH3T3 cells (**Figure 3.19.A**). I chose targets that were changing in gene expression with differentiation and with APOBEC2 knockdown (**Figure 3.19.B**). I chose this initial set of genes since they were transcription factors that increased in expression with knockdown of APOBEC2, which suggests they were repressed. Moreover, they had different modes of APOBEC2 occupancy, where APOBEC2 occupancy for Id3 was increasing at 34 h while the others had unchanging signal. I extracted total RNA and measured transcript abundance using reverse transcription-quantitative PCR (RT-qPCR). I measured a statistically significant but modest decrease in the expression in Id3 and Kazald1 with APOBEC2 expression (**Figure 3.19.C**). The reduction in transcript levels with APOBEC2 expression in NIH3T3 cells was promising. It suggested that APOBEC2 may directly regulate the expression of these genes.

To further develop this experiment, I then included the APOBEC2 variants for comparison. I also sought to determine whether the inability of the G64R variant to increase MyHC levels in the C2C12 cells was also due to a difference in cellular localization, similar to N-terminal truncated APOBEC2, del(1-41)A2. Thus, I checked the localization of exogenously expressed APOBEC2 using microscopy.

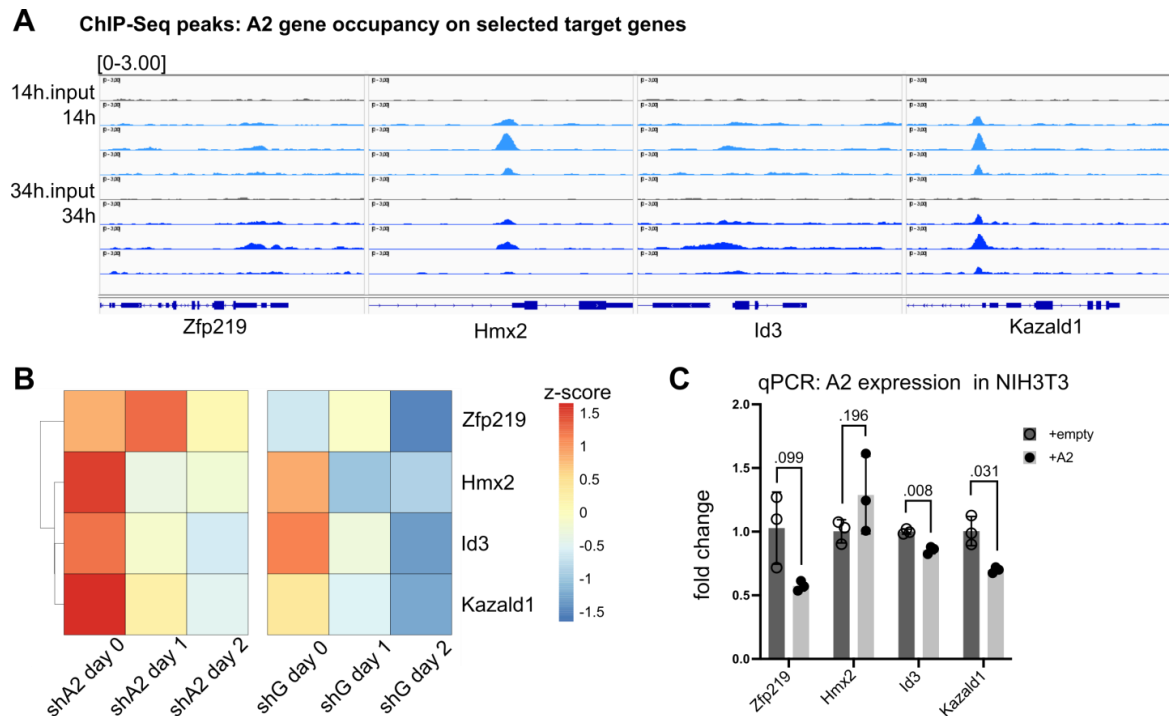


Figure 3.19. Exogenous expression of APOBEC2 in NIH3T3 fibroblasts affects APOBEC2 target genes.

A) Normalized ChIP-Seq peaks on selected APOBEC2 target genes, y-axis: 0 - 3.00. 14h.input and 34h.input (in gray) refer to respective input control ChIP-Seq experiments. Each row represents triplicates from each respective experiment time point, 14h (in light blue) and 34h (in blue). The image was prepared using the genome browser software, IGV (Robinson et al., 2011).

B) Heatmap was prepared from the RNA-Seq experiment from APOBEC2 knockdown (shA2) and control shGFP (shG) C2C12 myoblasts at corresponding time points of differentiation, day 0 to day 2. Heatmap colors correspond to mean normalized read counts scaled by row (z-score). Heatmap was prepared using the pheatmap package in R (Kolde, 2019).

C) RT-qPCR (qPCR) was done on selected target genes from total RNA extracted from NIH3T3 cells expressing APOBEC2. NIH3T3 cells were transduced with pMX-A2 (+A2) or pMX-empty (+empty) retroviral vectors. Bars represent mean fold change (normalized expression, $\Delta\Delta Cq$) between +empty and +A2 cells with Hprt and Tbp as reference housekeeping genes, calculated in BioRad CFX Maestro software. Error bars represent standard deviation. Bar plots and statistics were prepared using Prism 9 software; statistical analysis: multiple t-tests. Values above each pairwise comparison represent adjusted P values (Holm-Sidak method).

3.5.2 Expressed APOBEC2 had limited nuclear localization

I fused an N-terminal Flag-tag to APOBEC2 and its variants so I could use an anti-Flag antibody to specifically detect cellular localization of the exogenously expressed proteins. I included a variant of APOBEC2 with a mutation directly on the catalytic cytidine deaminase domain, the E100A variant. I detected Flag-tagged APOBEC2 expression using immunofluorescence microscopy in the NIH3T3 fibroblasts (**Figure 3.20**). Exogenous Flag-A2 could be found both in the cytoplasm and the nucleus of the cells (shown in green). I measured the nuclear and cytoplasmic Flag signal, mean

3. Results

fluorescence intensity, in each of the samples and quantified the number of cells with a nuclear enrichment of the Flag-tagged protein (**Table 1**). I considered cells with Flag-tag nuclear to cytoplasmic mean fluorescence intensity ratio greater than 1.1 to have nuclear enrichment.

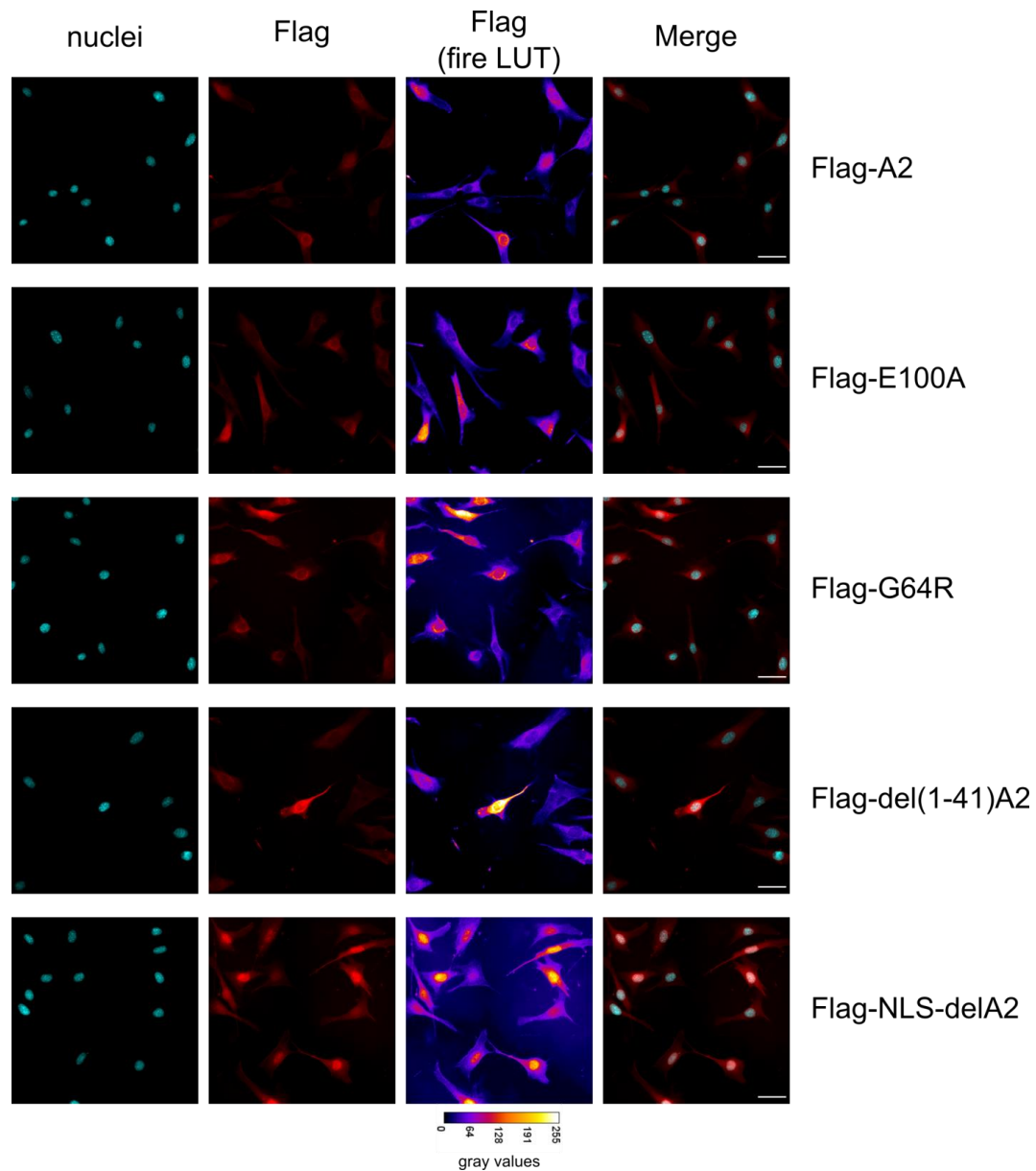


Figure 3.20. Exogenous APOBEC2 cytoplasmic and nuclear localization in NIH3T3 cells. Immunofluorescence images of NIH3T3 cells expressing exogenous Flag-tagged APOBEC2. Cells were transduced with retroviral vectors containing cDNA for Flag-A2, Flag-E100A, Flag-G64R, Flag-del(1-41)A2, or Flag-NLS-dA2. Each column represents a single channel corresponding to the nuclei, Flag-tag (Flag), Flag with pseudocolor (Fire LUT), and merged channels (Merge). Images were taken with a widefield fluorescence microscope, processed uniformly, and representative of the groups. Scale bar: 50 μ m. Scale for gray values, 0-255, of fire LUT is shown.

Flag-A2 and Flag-E100A showed similar numbers of cells with nuclear enrichment, 12% and 11% respectively; this meant the E100A variant did not affect nuclear localization (**Table 1**). As I expected based on the prior result that del(1-41)A2 was absent in the total nuclear protein fraction, there were fewer cells with a nuclear enrichment of del(1-41)A2 than wildtype APOBEC2, 6% vs 12%. However, this difference was not statistically significant (chi-squared test $p = 0.349$). Unexpectedly, the G64R variant had more cells with nuclear enrichment than wildtype APOBEC2, 37% vs 12% (chi-squared test $p < 0.001$).

Table 1. Counts of cells with APOBEC2 nuclear enrichment

Nuclear Enrichment (N/C > 1.1)	Flag-tagged exogenous protein				
	A2	E100A	G64R	del(1-41)A2	NLS-delA2
Yes	8 (12%)	7 (11%)	25 (37%)	3 (6%)	49 (82%)
No	57 (88%)	56 (89%)	42 (63%)	45 (94%)	11 (18%)
Total	65	63	67	48	60
p-value (Pairwise chi-squared test vs APOBEC2)¹	-	0.83	<0.001	0.349	<0.001

¹chi-squared tests using JASP (version 0.14.1)

The numbers of cells with APOBEC2 nuclear enrichment show that the variants affect the localization of APOBEC2. However, the nuclear enrichment of exogenous APOBEC2 was not as strong as what was observed in the differentiated myotubes (**Figure 3.6.D**). This weaker nuclear localization of exogenously expressed A2, which was emphasized further when I added an NLS signal to the del(1-41)A2 variant, could have resulted in the smaller difference in nuclear localization between APOBEC2 and del(1-41)A2. This was done to see if nuclear localization of del(1-41)A2 could be restored or increased. The NLS-del(1-41)A2 (NLS-dA2) showed stronger nuclear enrichment as expected, 82% (vs 12% wildtype A2, $p < 0.001$). The addition of the NLS showed A2 nuclear localization akin to what I observed in the differentiated myotubes.

Perhaps, the failure to detect gene expression changes in the C2C12 cells when I expressed APOBEC2 was due to insufficient nuclear enrichment. The modest effects on gene expression in the NIH3T3 cells could also be due to the limited number of cells

3. Results

showing strong APOBEC2 nuclear enrichment. Perhaps if I added an NLS to the exogenous APOBEC2, it would have shown stronger effects in terms of gene expression changes on APOBEC2 target genes, and potentially increased the myotube differentiation in the APOBEC2 knockdown cells.

The finding that the G64R variant was more readily enriched in the nucleus compared to wildtype APOBEC2 was puzzling. I was expecting that, similar to del(1-41)A2, this variant would have impaired nuclear entry to explain the limited MyHC expression when it was expressed in the APOBEC2 knockdown cells. It was difficult to imagine how a mutation on a presumably DNA binding loop, loop 1, could cause increased nuclear enrichment.

These findings suggested that more needed to be known about the control of APOBEC2 nuclear localization to clearly explain the changes in nuclear localization of the APOBEC2 variants. Nevertheless, the results so far suggested that APOBEC2 acted within the nucleus. This was highlighted by APOBEC2 nuclear enrichment in differentiated myotubes. It was confirmed when del(1-41)A2, which had impaired nuclear enrichment, was unable to increase MyHC protein levels similar to wildtype APOBEC2. Furthermore, these results together with the ChIP-Seq data suggested that APOBEC2 was a critical transcriptional regulation during myotube differentiation. I next aimed to determine whether APOBEC2 regulated transcription indirectly as a transcriptional coregulator, or as a transcription factor that can directly interact with DNA.

3.6 APOBEC2 nucleic-acid ligand interaction

Analysis of the ChIP-Seq data indicated that APOBEC2 interacted with chromatin of myoblasts at genes related to myotube differentiation. Furthermore, the co-IP showed that APOBEC2 directly interacted with the transcriptional corepressor, HDAC1. However, these results did not directly confirm if APOBEC2 directly interacted with the genomic DNA, a hallmark of a transcription factor. I decided to determine if APOBEC2 had an affinity for single-stranded DNA (ssDNA) similar to the enzymatically active members of the APOBEC family. AICDA interacts with ssDNA at immunoglobulin gene loci during B cell activation; perhaps APOBEC2 targets muscle differentiation-related genes with a similar molecular mechanism during myotube differentiation.

3.6.1 SP/KLF GC-rich motifs were enriched in APOBEC2 occupied genomic regions

As input for the DNA motif discovery tool called STREME (from the MEME suite web interface), I used the set of A2 occupied genomic regions, with 401 bp centered at the peak I defined earlier using the differential binding analysis. Using this tool, I searched for enriched motifs within the APOBEC2 occupied genomic regions with the standard settings (Bailey, 2021). The top enriched motifs were then compared to annotated mouse transcription factor binding motifs with the Tomtom tool using the HOCOMOCO mouse (v11 CORE) transcription factor database (Gupta et al., 2007). The top 3 enriched motifs from the APOBEC2 ChIP-Seq data were most similar to SP/KLF, Fos/Jun, and NFY transcription factor motifs (**Figure 3.21**). The complete results of the STREME analysis could be found in **Appendix IV**. Though not among the top 3 motifs, several other GC-rich motifs that were similar to SP/KLF transcription factor motifs were also enriched. These results suggested that, at these APOBEC2 binding sites, APOBEC2 might interact with SP/KLF or other transcription factors to affect binding at their cognate motifs.

I chose the top enriched motif, the GC-rich SP/KLF motif, for the subsequent experiment to test whether APOBEC2 directly interacted with DNA. I also chose this motif for the test ligand since it contained several cytidines that could be potential substrates for APOBEC activity. I also included a poly-A motif that was among the enriched motifs found in the APOBEC2 occupied genomic regions (**Figure 3.21**). This motif however was not centrally enriched (**Appendix IV**), which meant that it was not

advantage of requiring small quantities of protein and ligand, microliter quantities at nanomolar concentrations of protein.

I labeled APOBEC2 and the variants I have described so far, del(1-41), G64R, and E100A, with a fluorescent Cy5 dye. I biased labeling towards the amino-terminal residue to minimize disruption of APOBEC2 protein by increasing the pH of the labeling reaction (see Materials and Methods section 5.13). I also made sure that labeling was at approximately one Cy5 dye molecule per protein molecule to avoid other unintended disruptions to protein structure or ligand affinity. After confirming that APOBEC2 and the variants had similar labeling efficiencies and fluorescent signals, I performed the affinity measurements.

Following a standard MST protocol, I mixed 50 nM of Cy5-labeled APOBEC2 protein with a range of DNA ligand concentrations ranging from 6 nM to 50,000 nM. After a 30-minute incubation at room temperature, I performed the MST measurements. I compared the affinities of the wildtype APOBEC2 protein and two different variants, G64R and E100A. I was unable to get reliable measurements for the del(1-41)A2 variant due to the tendency of the purified protein to precipitate. APOBEC2 and G64R showed shifts in fluorescence from the bound and unbound states, ΔF_{norm} , with either the SP/KLF motif and the polyA motif ligands (**Figure 3.22.A,C**). However, the E100A variant did not show a shift in fluorescence, which indicated that it could not bind either of the motifs.

The G64R variant showed a stronger affinity than wildtype APOBEC2 for both the SP/KLF motif and poly-A motif DNA ligands (**Figure 3.22.B,D**). In particular, the G64R variant had 3-fold higher affinity than wildtype for both ligands, SP/KLF motif: dissociation constant (K_d) = 248.16 ± 47.924 nM vs 926.48 ± 243.48 nM; and poly-A motif: K_d = 6691.6 ± 1898.2 nM vs K_d = 20213 ± 6575.4 nM. Both wildtype APOBEC2 and the G64R variant showed 20-fold less affinity for the poly-A motif.

3. Results

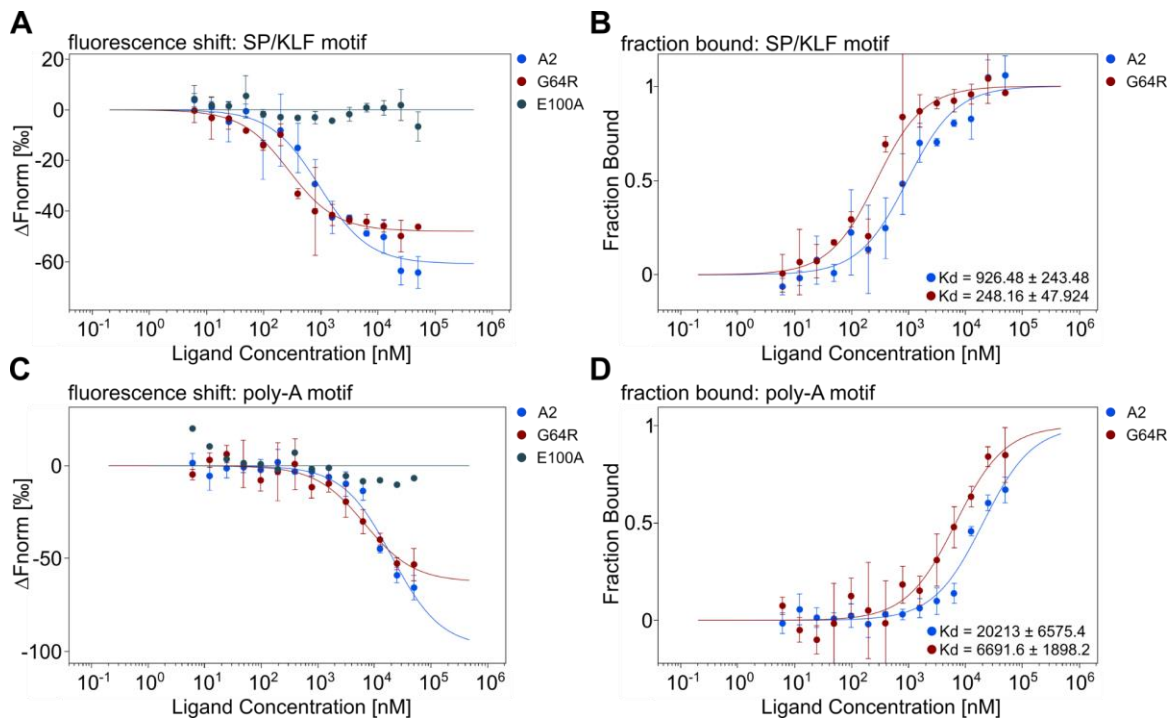


Figure 3.22. APOBEC2 and single-stranded DNA ligand affinity measurements.

A,C) Fluorescence shift plots of Cy5-APOBEC2 and variants, G64R and E100A, with either the (A) SP/KLF motif or (C) the poly-A motif. The points represent the mean normalized fluorescence shift (ΔF_{Norm}) from 3 independent runs, the error bars represent the standard deviation. The curve was fit using the Kd model of the MO.Affinity Analysis v2.1 software (NanoTemper Technologies).

B,D) The derived fraction bound plots of the (B) SP/KLF motif and (D) with Cy5-APOBEC2 and the G64R variant from the respective fluorescence shift plots. The corresponding affinity measurements, Kd, of the interactions were calculated using MO.Affinity Analysis v2.1 software (NanoTemper Technologies).

These results indicate that APOBEC2 can directly interact with DNA ligands. The binding was specific as shown by a stronger preference for the SP/KLF motif, the top enriched motif found in the APOBEC2 ChIP-Seq binding sites. Also, directly altering the zinc-binding catalytic domain of APOBEC2 with the E100A mutation abrogated DNA interaction. Furthermore, the G64R variant found in the patients with fatal muscular dystrophy had a much stronger affinity for the DNA ligands. Thus, this increase in affinity may have physiological consequences. Furthermore, the increased DNA affinity may explain the increase in nuclear retention observed for this variant (**Table 1**).

I have shown that APOBEC2 can directly interact with specific DNA sequences. Thus, from my results, I have provided evidence that the unknown molecular function of APOBEC2 is transcriptional regulation. The results showed that APOBEC2 is crucial to proper myotube differentiation. The ChIP-Seq in the differentiating myoblasts showed that APOBEC2 occupies and potentially regulates the transcription of genes related to muscle

cell differentiation. Moreover, I showed that APOBEC2 directly interacts with the transcriptional corepressor HDAC1, which provides a direct molecular mechanism for transcriptional regulation. There was also evidence for the regulation of APOBEC2 nuclear localization. From all these experiments, I could to a certain degree of scientific certainty propose that APOBEC2 is a transcription factor.

4. Discussion

4. Discussion

There have been several attempts at defining the molecular function of APOBEC2. Given that it belongs to a distinguished family of cytidine deaminases, it was expected that its intact deaminase domain should function in modifying cytidines to uridines. However, this has not been unequivocally demonstrated. Thus, it was deemed a non-functional deaminase with roles far removed from its potential to bind nucleic acid ligands (Powell et al., 2014).

The strict structural conservation of APOBEC2 across evolutionary millennia, from the founding of the vertebrate lineage, suggested it had a vital purpose. This was reflected by the works of others that demonstrated its loss resulted in defects in skeletal and cardiac muscle (Etard et al., 2010; Sato et al., 2010). Thus, I embarked on this work to reveal its essential purpose.

In this work, I focused on revealing the role of APOBEC2 within myotube nuclei. More specifically, I hypothesized that APOBEC2, even if it may have lost its catalytic function, has retained its ability to interact with nucleic acid ligands; and through this interaction, it might act as a transcriptional regulator.

4.1 On the role of APOBEC2 in transcriptional regulation

The first contribution of this work was determining that APOBEC2 was directly involved in the transcriptional regulation of myotube differentiation. Using the C2C12 myotube differentiation model, I demonstrated that APOBEC2 loss disrupts differentiation. In particular, early loss of APOBEC2 in C2C12 myotube differentiation resulted in gene expression changes that decreased myotube formation. Moreover, APOBEC2 was found within the myotube nuclei where it interacted with the chromatin.

Within chromatin, APOBEC2 occupied transcriptional regulatory regions of several hundred genes, including those involved in biological processes relevant to muscle differentiation. Interestingly, APOBEC2 also occupied transcriptional regulators involved in cell differentiation of other lineages, such as microglial cells and immune cells like T cells. These occupied APOBEC2 target genes changed in expression with APOBEC2 loss.

To demonstrate the link between APOBEC2 function and myotube differentiation, I reintroduced APOBEC2 in the knockdown cells and observed an increase in MyHC protein levels. Additionally, the truncated variant of APOBEC2, del(1-41)A2, which was unable to enter the nucleus, did not increase MyHC protein levels. This demonstrated that APOBEC2 function was linked to nuclear localization.

I then attempted to demonstrate the role of APOBEC2 as a transcriptional regulator of its target genes. One such target gene was *Id3*; a transcriptional repressor that inhibited muscle differentiation (Chen et al., 1997). I hypothesized that APOBEC2 promoted muscle differentiation indirectly through repression of *Id3* expression. I could demonstrate repression of *Id3* in NIH3T3 fibroblasts though the effects were modest even with an overexpression system. Moreover, it was inconsistent in the C2C12 myoblast model where I could have tested the hypothesis in the right tissue context. I then realized that there was a limited nuclear entry of the exogenous APOBEC2. Stronger experimental evidence of APOBEC2 acting as a transcriptional regulator would require strong nuclear enrichment akin to that seen in the differentiated myotubes.

Uncovering the factors controlling APOBEC2 nuclear localization was beyond the scope of this work. However, I demonstrated that removal of the N-terminal flexible region of APOBEC2 affected its nuclear localization. Moreover, this truncated APOBEC2 was unable to increase MyHC levels supporting the hypothesis that APOBEC2 was acting as a transcriptional regulator within the nucleus. APOBEC2 nuclear localization was believed to be a passive process, without nuclear import or export signals, when expressed in HEK293T kidney cells and Ramos B cells (Patenaude et al., 2009). However, my results demonstrated that the N-terminal region was required for nuclear localization contradicting that nuclear entry was a passive process. Perhaps, however, the N-terminal region was necessary for nuclear retention of APOBEC2. The truncation might have disrupted APOBEC2 intermolecular interactions necessary for nuclear retention.

One potential mechanism for APOBEC2 nuclear retention might be through this intrinsically disordered N-terminal region. Such intrinsically disordered regions, which have also been described in transcription factors, are believed to affect the ability of these proteins to form phase-separated nuclear condensates that serve a role in gene regulation (Boija et al., 2018). APOBEC2 could potentially interact with these transcriptional condensates. Further studies on the disordered N-terminal domain of APOBEC2 may uncover its role in intermolecular interactions and the mechanism for APOBEC2 nuclear localization.

Nevertheless, another piece of evidence that APOBEC2 acts as a transcriptional regulator was the finding that it interacted directly with the transcriptional corepressor HDAC1. I showed that within the nuclei of differentiated C2C12 myotubes, endogenous APOBEC2 interacted with HDAC1. This provided a direct molecular mechanism for how APOBEC2 could regulate the transcription of its target genes. In C2C12 myotube

4. Discussion

differentiation, MYOD1 similarly interacted with HDAC1 at repressed target genes that are activated once HDAC1 is displaced during differentiation (Mal & Harter, 2003). Similarly, APOBEC2 target genes could be repressed through the interaction with HDAC1.

Further experiments to determine the genes occupied by APOBEC2 together with HDAC1 during muscle differentiation could define the set of repressed APOBEC2 target genes. From my analysis, I could infer that the genes that were occupied by APOBEC2 and repressed during C2C12 differentiation were repressed through histone deacetylation by HDAC1. Presumably, APOBEC2 target genes which were instead activated lacked HDAC1 interaction. Moreover, HDAC1-mediated transcriptional repression may only be one of the mechanisms APOBEC2 uses for transcriptional control.

Additionally, expanding the time points of the ChIP-Seq beyond the early time points in this experiment may provide more APOBEC2 target genes. I would expect that the current set of APOBEC2 genes I defined here would increase in ChIP-Seq signal with increased APOBEC2 protein levels. Moreover, new target genes may emerge, providing further insights into the biological pathways regulated by APOBEC2. Currently, the data suggest that APOBEC2 regulates non-skeletal muscle biological processes. APOBEC2 did not directly target pathways regulated by the skeletal muscle myogenic regulatory factors, MYOD1, MYF5, MYOG, and MRF4 (Bentzinger et al., 2012). Instead, APOBEC2 targeted the genes of transcription factors associated with multiple cell lineages, such as Stat3, Runx1, Gli3, and Sox4.

These transcription factors were particularly interesting since they were among the genes that were within the enriched GO terms from the list of upregulated genes with APOBEC2 knockdown. In particular, they belong to GO terms related to T cell and glial cell differentiation. Interestingly, these transcription factors have also been linked to skeletal muscle differentiation. Stat3 and Runx1 play roles in muscle regeneration through interactions with MYOD1 (Tierney et al., 2014; Umansky et al., 2015). Gli3 regulates both muscle development and regeneration (Renault et al., 2013). But, the role of these transcription factors in a broad range of differentiation signaling pathways makes them interesting targets for APOBEC2. Further investigation on the role of APOBEC2 in the regulation of these transcription factors may generalize the role of APOBEC2 in broader non-muscle tissue contexts.

4.2 On the ability of APOBEC2 to directly interact with DNA as a transcription factor

The second contribution of my work was revealing that APOBEC2 retains the ability to interact with a DNA ligand. Initially, APOBEC2 was thought to bind ApoB mRNA and deaminate free cytidines (Anant et al., 2001; Liao et al., 1999). However, these functions were later attributed to contaminating bacterial deaminases from recombinant protein preparation (Mikl et al., 2005). Moreover, there has been no detectable deaminase activity for APOBEC2 (Harris et al., 2002; Mikl et al., 2005). Despite this lack of enzymatic activity, my work has shown that APOBEC2 retains the ability to bind to a specific single-stranded DNA ligand. Binding was specific to a ligand containing an SP/KLF motif that was enriched in APOBEC2 chromatin binding sites. The affinity of APOBEC2 for a poly-A motif found adjacent to APOBEC2 binding sites, a potentially non-specific target, was much weaker. This function may have been missed by others since I used recombinant APOBEC2 produced in eukaryotic insect cells instead of a prokaryotic *E. coli* production system. Eukaryotic post-translational protein modifications might be critical to APOBEC2 nucleic acid interaction.

The calculated dissociation constant (K_d), a measure of affinity, of APOBEC2 for the SP/KLF motif was approximately 0.9 μM. This was well within the range of calculated affinities of other APOBEC members for their cognate ligands. For APOBEC1, the estimated K_d from gel shift assays was 0.6 μM for ssDNA (Wolfe et al., 2020). For AID, the estimated K_d from gel shift assays ranged from 1.5 to 7 μM for several linear ssDNA motifs but were much stronger for the G-quadruplex (G4) structured motifs, 0.1 to 0.2 μM (Qiao et al., 2017). For the catalytic C-terminal cytidine deaminase domain of APOBEC3G, using a similar MST method, the estimated K_d was much weaker at 55 μM (Maiti et al., 2018). This was notable since this measurement comes from an optimized sequence of the APOBEC3G domain that showed stronger deaminase activity and stronger substrate binding.

Moreover, I show that a mutation (E100A) on the catalytic glutamate residue of APOBEC2 abolishes ligand binding. Similarly, a homologous mutation (E259A) of the catalytic residue in APOBEC3G also decreases its affinity for ssDNA (Maiti et al., 2018). It was unlikely that the E100A mutation on APOBEC2 disrupted the folding of the cytidine deaminase domain since crystal structures of homologous mutations in AID (E58A), APOBEC3A (E72A), and APOBEC3G (E259A) show intact deaminase domains

4. Discussion

(Kouno et al., 2017; Maiti et al., 2018; Qiao et al., 2017; Shi et al., 2017). However, this mutation may have disrupted the interaction with the ssDNA ligand since the co-crystal structure of AID with its deoxy-cytidine substrate suggests direct interaction between the cytidine substrate and catalytic E residue (Qiao et al., 2017). The loss of ligand interaction of APOBEC2 E100A suggests that this residue interacts with the ssDNA substrate and that it is crucial to the interaction.

Furthermore, I have shown that the APOBEC2 G64R variant, which was found in two patients afflicted with a fatal myopathy, increased affinity for the ssDNA ligand. This variant was also ineffective in increasing MyHC protein levels when overexpressed in the myoblasts suggesting that it was a consequential mutation to APOBEC2 function. This mutation is found on a residue in loop 1, which has been shown to directly interact with ssDNA ligand in the APOBEC3A and APOBEC3G structures (Kouno et al., 2017; Maiti et al., 2018; Shi et al., 2017) (**Figure 4.1**). Furthermore, a mutation on the homologous residue in AID (G23S) impairs specifically only the somatic hypermutation activity of AID (Wei et al., 2011). As seen in the co-crystal structures, loop 1, together with loops 3, 5, and 7, forms the DNA ligand-binding groove in the APOBEC co-crystal structures. Changes in the length or the residues of these loops are believed to affect DNA motif specificity (Krishnan et al., 2018). The increase in affinity for the ssDNA ligand I observed for the G64R variant may have been caused by either, or both, a change in the shape of the loop caused by the larger R residue, or a change in the charge of the loop caused by the positive charge of the R residue. The change in ligand affinity for the G64R variant may have had consequential effects on ligand specificity affecting APOBEC2 function.

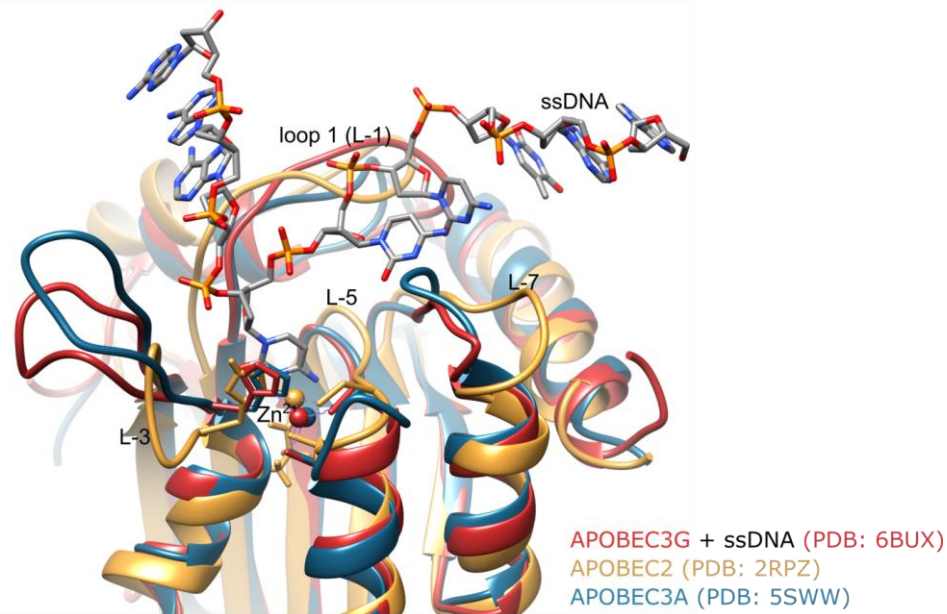


Figure 4.1 APOBEC2 structure alignment with APOBEC3A and APOBEC3G bound to ssDNA substrate. APOBEC3G (mouse) C-terminal domain in complex with ssDNA, 5'-dAATCCCAAA-3', structure derived from X-ray crystal structure is depicted in a red ribbon structure (PDB: 6BUX, (Maiti et al., 2018)). APOBEC2 (mouse) structure, without the first 45 residues, derived from solution nuclear magnetic resonance (NMR) spectroscopy is depicted in a yellow ribbon structure (PDB: 2RPZ, model 13). APOBEC3A (human) complexed with ssDNA derived from the X-ray crystal structure is depicted in a blue ribbon structure (PDB: 5SWW, (Shi et al., 2017)). Structures were aligned using MatchMaker in UCSF Chimera with the APOBEC3G structure as a reference structure. The ssDNA ligand from the APOBEC3G co-crystal structure (in gray) is depicted as a stick model of atoms and bonds. The Zn²⁺ ion is depicted as a colored ball (color corresponds to respective ribbon structure).

Given that APOBEC2 has identical folding to the other APOBECs with ligand co-crystal structures, and that mutations on homologous DNA interacting residues affect APOBEC2 DNA ligand affinity, I hypothesize that an APOBEC2 co-crystal structure with an ssDNA ligand would show similar ligand-binding mechanisms to the other APOBECs.

The ability of APOBEC2 to interact with a double-stranded (ds) DNA ligand through a gel shift assay has been investigated (Powell et al., 2014). However, there was no evidence for direct APOBEC2 dsDNA binding. Instead, APOBEC2 cooperatively enhanced the binding of transcription factor Pou6f2 to its cognate motif. Studying the interaction of APOBEC2 with a dsDNA ligand using MST might provide more insights on APOBEC2 chromatin interaction.

During transcription initiation, APOBEC2 might interact with transient ssDNA structures caused by the loading of the transcriptional complex. Such ssDNA conformations of chromatin may arise during promoter melting during the transition of RNA polymerase II (Pol II) complex to an open state (Boyaci et al., 2019; Plaschka et al., 2016). Promoter-proximal pausing of the Pol II complex may also lead to the formation of ssDNA structural conformations in chromatin (Szlachta et al., 2018). Notably, these

4. Discussion

transient ssDNA structures regulate gene transcription during naïve B lymphocyte activation (Kouzine et al., 2013). G4 structured ssDNA, similar to that preferred by AID, could arise in G-rich sequences of promoter regions of genes during transcription (Agrawal et al., 2014; Du et al., 2008; Maizels, 2006). Given that APOBEC2 binds a GC-rich SP/KLF motif, perhaps it targets C-rich complementary sequences of G4 quadruplex structures. Though not as well studied as G4 quadruplex structures, C-rich sequences are known to form secondary structures called i-motifs in vitro and in vivo (Zeraati et al., 2018). APOBEC2 potentially regulates the transcription of its target genes through direct interactions with these transient ssDNA conformations.

Given that I have shown that APOBEC2 binds ssDNA, it would be interesting to test whether it has deaminase activity on the specific motifs found at its genomic binding sites. The apparent lack of deaminase activity for APOBEC2 has been linked to an APOBEC2-specific difference at loop 5. Other members of the APOBEC family have a conserved SWS amino acid sequence at loop 5 whereas APOBEC2 has an SSS sequence. Conversion of SWS to SSS in APOBEC1 greatly reduces its cytidine deaminase activity when tested in a bacterial mutator assay (Harris et al., 2002). However, conversion of SSS to SWS in APOBEC2 does not grant deaminase activity (Powell et al., 2014). This suggests that though this conserved SWS motif seems vital to APOBEC1 deaminase function, and likely other AID/APOBEC family members, loss of detectable deaminase activity for APOBEC2 may not be solely due to this. Interestingly, in the co-crystal structure of the active C-terminal deaminase domain of APOBEC3G with ssDNA substrate, the W residue of the SWS motif has no molecular interactions with the substrate (Maiti et al., 2018). It was also initially thought that loop 1 obscured the active site of APOBEC2; however, the flexibility of loop 1 allows different conformations that open the active site (Krzysiak et al., 2012). Furthermore, co-factors interacting with APOBEC2, possibly through the unstructured N-terminal region, may cause structural shifts allowing access to the active site. There has still been no confirmed molecular mechanism for the loss of APOBEC2 enzymatic activity, nor has there been demonstrable deaminase activity for APOBEC2.

I propose that during APOBEC2 evolution it lost its cytidine deaminase activity when it was co-opted as a transcription factor. Perhaps APOBEC2 lost its ability to deaminate its DNA ligand to avoid detrimental mutations on the genome. And instead, APOBEC2 retained its ability to interact with DNA to recruit transcriptional coregulators, such as HDAC1, as a transcription factor during myotube differentiation (**Figure 4.2**).

Further studies on the targets, coregulators, and cofactors of APOBEC2 would help elucidate the mechanism of APOBEC2 gene targeting.

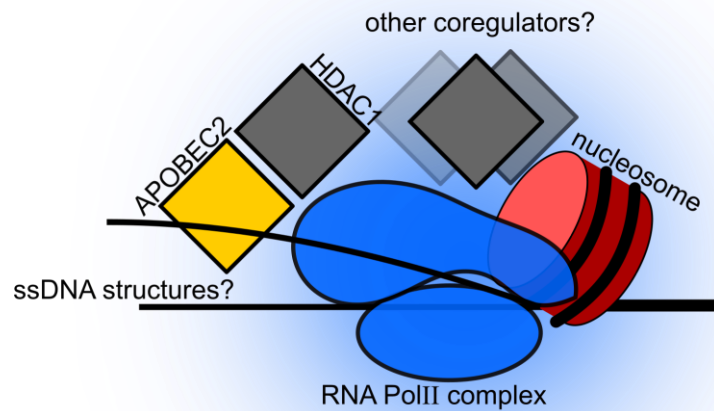


Figure 4.2. Proposed model of APOBEC2 transcriptional complex interaction. In the transcriptional control of muscle differentiation, APOBEC2 directly interacts with ssDNA structures at promoter regions produced during RNA polymerase II (Pol II) complex pausing. APOBEC2 then recruits transcriptional coregulators such as the corepressor HDAC1. The effect of APOBEC2 on transcription would depend on the recruited coregulatory complexes.

4. Discussion

4.3 On the role of APOBEC2 in muscle and beyond

Through this work, I have shown that APOBEC2 is crucial in skeletal myotube differentiation, where it plays a role as a transcriptional regulator of biological processes necessary for myotube formation. Reduction of APOBEC2 levels led to hundreds of differentially expressed genes early in myotube differentiation that ultimately led to the decreased formation of myotubes. However, APOBEC2 only targeted a fraction of these differentially expressed genes, most of which fall within biological processes involved in cell differentiation and development. Additionally, these biological processes and genes were only tangentially related to muscle differentiation.

If APOBEC2 was suppressing these non-skeletal muscle differentiation programs, then perhaps there would be cell-specific markers of non-muscle cells expressed in A2 knockdown C2C12 myoblasts. I found upregulated genes that fall in non-skeletal muscle-related biological processes, such as kidney development and neural tube development. However, I could not find genes that would suggest that the skeletal myoblasts were no longer skeletal myoblasts. Perhaps an additional approach to define the chromatin state, through global chromatin accessibility assays such as Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq), may improve the definition of the disrupted cell state of the APOBEC2 knockdown myoblasts. Nevertheless, my analysis suggested that APOBEC2 was repressing several non-muscle-related biological processes.

I hypothesize that APOBEC2 acts as a regulator that prevents spurious cell fates during muscle differentiation, similar to roles played by MYT1L in repressing promiscuous MYOD1 activity (Lee et al., 2020). As a regulator of spurious cell fates, APOBEC2 does not act as a myogenic regulatory factor crucial to myogenesis. APOBEC2 knockout animals still develop muscle albeit defective myopathic muscle. Moreover, isolated myoblasts from APOBEC2 knockout mice can form myotubes in vitro (Ohtsubo et al., 2017). Instead, perhaps the transcriptional function of APOBEC2 plays a role in adult muscle development and regeneration.

Continued work on the APOBEC2 knockout mice traced the myopathy to mitochondrial dysfunction (Sato et al., 2017). Further work, although counterintuitive, on the APOBEC2 knockout myoblasts from the same animals suggest that APOBEC2 negatively regulates muscle differentiation (Ohtsubo et al., 2017). In contrast, my work and the work of others have demonstrated that decreasing APOBEC2 levels lead to decreased muscle differentiation (Carrió & Suelves, 2015; Vonica et al., 2011). It is also

difficult to imagine a model where APOBEC2 is both detrimental and necessary for healthy skeletal muscle differentiation. Ohtsubo et al (2017) did however propose a model where APOBEC2 depletion leads to premature satellite cell activation, which are postnatal skeletal muscle stem cells (Ohtsubo et al., 2017). Perhaps, in the context of satellite cell activation during postnatal muscle development and regeneration, APOBEC2 acts as a transcription factor that suppresses non-skeletal muscle differentiation programs including satellite cell activation. Thus, when APOBEC2 is depleted, spurious genetic programs deplete satellite cells, through uninhibited activation, and disrupt healthy skeletal muscle differentiation.

Furthermore, such a model also explains the APOBEC2-dependent Müller glial cell reprogramming and dedifferentiation during zebrafish retinal regeneration (Powell et al., 2012, 2013). Similarly, in zebrafish, APC knockout embryos upregulated APOBEC2 expression that was initially thought to be directly involved in the observed promoter hypomethylation of intestinal lineage specifying transcription factors (Rai et al., 2010). Initially, these phenotypes were thought to be due to the role of APOBEC2 in DNA demethylation. However, it was eventually demonstrated that APOBEC2 does not demethylate DNA (Nabel et al., 2012; Powell et al., 2013). I speculate instead that defects in retinal regeneration or intestinal cell differentiation observed in the zebrafish were due to the transcriptional role of APOBEC2 in regulating cell differentiation processes. This is also supported by the observation that differentially expressed APOBEC2 target genes in the C2C12 skeletal myoblasts were related to broad cell differentiation programs. Furthermore, the genes upregulated with APOBEC2 knockdown were related to non-skeletal muscle biological processes, such as T cell activation. Perhaps APOBEC2 has transcriptional regulatory roles beyond just the skeletal muscle cell lineage.

It is believed that the APOBEC family emerged when the ancestral AID cytidine deaminase was co-opted by a primordial B lymphocyte to generate immune diversity in an adaptive immune system (Rogozin et al., 2007). Since APOBEC2 is as ancient and conserved as AID, perhaps they shared similar fates. In my work, I have shown evidence that APOBEC2 acts as a transcription factor in muscle differentiation by regulating both muscle-related and non-muscle-related cell differentiation. One biological pathway that kept emerging in my analyses was the regulation of T cell activation. Perhaps APOBEC2 has unexplored ancestral roles in transcriptional regulation of T cells and other immune cells. Even though APOBEC2 is most abundant in skeletal and heart muscle, APOBEC2 could also be detected in several immune cell lineages (**Figure 4.3**).

4. Discussion

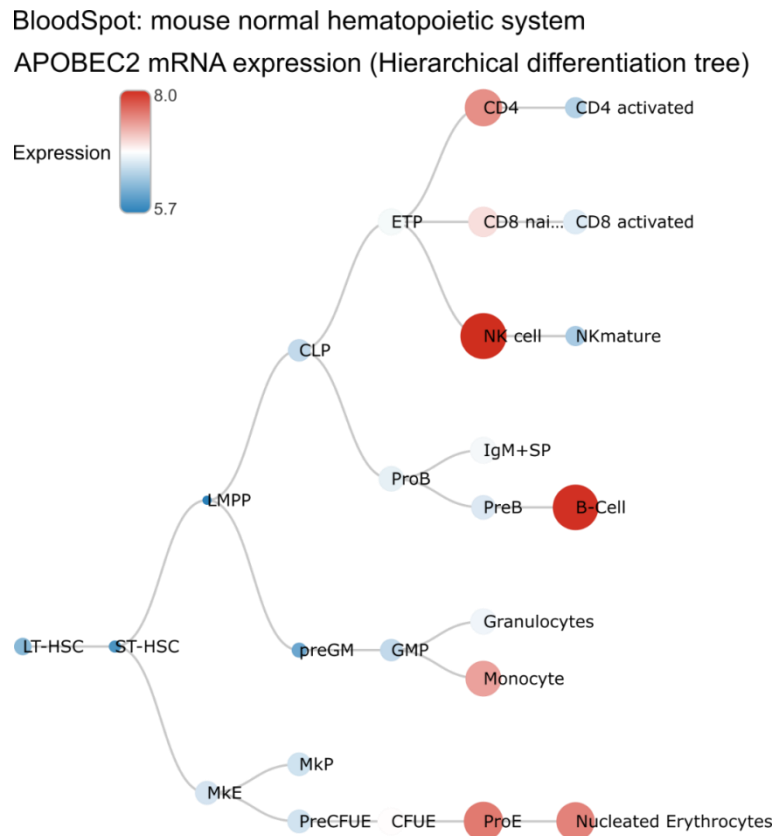


Figure 4.3. APOBEC2 expression in blood cell lineages. Hierarchical hematopoietic system differentiation tree depicting APOBEC2 expression levels at specific cell populations. The tree and APOBEC2 expression values (normalized expression from microarray experiments) were taken from the BloodSpot web interface (Bagger et al., 2016). Scale and blue to red colors represent the log₂ normalized expression values from lowest to highest. The hierarchical differentiation tree is rooted in long-term hematopoietic stem cells (LT-HSC). ST-HSC, short-term HSC; LMPP, lymphoid-primed multipotent progenitor; CLP, common lymphoid progenitor; ETP, early T cell precursor; CD4/CD8, CD4⁺ or CD8⁺ T cells; NK, natural killer cells; ProB/PreB, Pro-B, and pre-B cells; IgM+SP, IgM⁺ B cells; MKE, megakaryocyte-erythroid progenitors; MkP, megakaryocyte progenitors; PreCFUE/CFUE, pre-erythroid colony-forming unit; and, ProE, proerythroblast.

APOBEC2 is most highly expressed in both skeletal and heart muscle (Liao et al., 1999). I limited my work to skeletal myoblasts, but perhaps APOBEC2 also plays a role in cardiomyocyte physiology. In the zebrafish model, the knockdown of APOBEC2 led to defects in both skeletal and heart muscle function (Etard et al., 2010). In contrast, the mouse APOBEC2 knockout model only showed a defect in skeletal muscle tissue (Sato et al., 2010). Perhaps this difference could be explained if APOBEC2 function is critical only during muscle regeneration. Zebrafish can regenerate cardiomyocytes, whereas mammalian hearts are unable to do so (Beffagna, 2019).

Generalizing the role of APOBEC2 beyond muscle would require further studies in other tissue contexts. Examining the role of APOBEC2 in regeneration and repair across

different tissues and organisms may also provide deeper insight into the mechanism of APOBEC2 transcriptional control.

In summary, results from my work have contributed to unraveling the unknown essential molecular function of APOBEC2. I have provided evidence that APOBEC2 acts as a transcription factor in myoblast to myotube differentiation. I have not shown if it has deaminase activity, but, I have shown that it interacts with ssDNA, a characteristic of other members of the APOBEC family. This mechanism of transcriptional control is unique to the APOBEC family and expands the functional repertoire of these prolific cytidine deaminases.

5. Materials and Methods

5. Materials and Methods

Table 2. List of reagents and materials

REAGENTS AND MATERIALS	CATALOG NUMBER	SOURCE
Benzonase ≥ 250 units/μL, $\geq 99\%$	E8263	Merck/Millipore
Bovine serum albumin, heat shock fraction, pH 7, $\geq 98\%$	A7906	Merck/Sigma-Aldrich
Corning® BioCoat™ Collagen I 22 mm Round #1 German Glass Coverslip	354089	Corning
Cy5 Mono NHS Ester	GEPA15101	Merck/Sigma-Aldrich
Dimethyl sulfoxide	D2650	Merck/Sigma-Aldrich
Dulbecco's Modified Eagle's Medium-high glucose	D6429	Merck/Sigma-Aldrich
Dulbecco's Phosphate Buffered Saline (1X), DBPS without Calcium and Magnesium	BE17-512F	Lonza
Dynabeads® M-280 Sheep anti-Mouse IgG	11201D	Thermo Fisher Scientific/Invitrogen
Dynabeads® M-280 Sheep anti-Rabbit IgG	12203D	Thermo Fisher Scientific/Invitrogen
ECL™ Start Western Blotting Detection Reagent	GERPN3243	Merck/GE
Fetal Bovine Serum, qualified, Brazil	10270106	Thermo Fisher Scientific/Gibco
Halt™ Protease Inhibitor Cocktail, EDTA-free (100X)	78437	Thermo Fisher Scientific
Horse Serum, New Zealand origin	16050122	Thermo Fisher Scientific/Gibco
IGEPAL® CA-630 (NP-40)	56741	Merck/Sigma-Aldrich
iTaq Universal SYBR Green Supermix	1725121	Bio-Rad
Lipofectamine™ 2000 Transfection Reagent	11668027	Thermo Fisher Scientific/Invitrogen
Magnesium chloride hexahydrate, $\geq 99\%$, p.a., ACS	2189.1	Roth
Millex-HA Syringe Filter Unit, 0.45 μm, mixed cellulose esters, 33 mm, ethylene oxide sterilized	SLHA033SS	Merck/Millipore
Mini-PROTEAN® TGX™ Precast Protein Gels, 4-20%	4561096	Bio-Rad
NEBuilder® HiFi DNA Assembly Master Mix	E2621	New England BioLabs
Opti-MEM™ I Reduced Serum Medium	31985062	Thermo Fisher Scientific/Gibco
Penicillin-Streptomycin	P4333	Merck/Sigma-Aldrich
Pierce™ 16% Formaldehyde (w/v), Methanol-free	28908	Thermo Fisher Scientific
Polybrene Infection / Transfection Reagent	TR-1003-G	Merck/Sigma-Aldrich
ProtoScript® First Strand cDNA Synthesis Kit	E6300	New England BioLabs
Puromycin Dihydrochloride	A1113803	Thermo Fisher Scientific/Gibco
Q5® Site-Directed Mutagenesis Kit	E0554	New England BioLabs
Qubit RNA BR assay kit	Q10211	Thermo Fisher Scientific/Invitrogen
RNeasy mini kit	74106	Qiagen
Sodium chloride $\geq 99.5\%$ analytical reagent grade	10735921	Thermo Fisher Scientific
Sodium dodecyl sulfate, $\geq 99.0\%$ (GC), dust-free pellets	75746	Merck/Sigma-Aldrich
standard glass capillaries for microscale thermophoresis	MO-KO22	NanoTemper Technologies
SuperSignal™ West Femto Maximum Sensitivity Substrate	34095	Thermo Fisher Scientific

5. Materials and Methods

Triton® X 100	3051.3	Roth
Trizma Base (Tris bae)	93350	Merck/Sigma-Aldrich
Trypsin-EDTA solution	T3924	Merck/Sigma-Aldrich
TURBO DNA-free kit	AM1907	Thermo Fisher Scientific/Invitrogen
TWEEN 20, MP Biomedicals	MP1TWEEN20 1	Thermo Fisher Scientific

5. Materials and Methods

Table 3. List of antibodies

ANTIBODIES	CATALOG NUMBER	SOURCE
anti-Histone H3, HRP-conjugated	ab21054	abcam
goat anti-mouse IgG (H+L), HRP-conjugated	170-6516	Bio-Rad
goat anti-rabbit IgG (H+L), HRP-conjugated	170-6515	Bio-Rad
mouse anti-alpha Tubulin (DM1A)	ab7291	abcam
mouse anti-APOBEC2 clone 15e11	produced in lab	produced in lab
mouse anti-Flag (M2)	F1804	Merck/Sigma-Aldrich
mouse anti-Myosin heavy chain	MF20	Developmental Studies Hybridoma Bank (DSHB)
mouse anti-TroponinT (JLT-12)	T6277	Merck/Sigma-Aldrich
rabbit anti-APOBEC2	HPA017957	Merck/Sigma-Aldrich
rabbit anti-HDAC1 antibody	ab7028	abcam
rabbit anti-Sp1	07-645	Merck/Sigma-Aldrich

Table 4. List of plasmids

PLASMIDS	CATALOG NUMBER	SOURCE
pMXs-IRES-GFP-SV40-Puro (pMX)		Linda Molla
pCL-Eco retrovirus packaging vector	12371	Addgene
pLKO.1-APOBEC2-shRNA	TRCN0000112019	The Broad Institute's Mission TRC-1 mouse library
pLKO.1 puro GFP siRNA	12273	Addgene
psPAX2	12260	Addgene
pMD2.G	12371	Addgene
pFastBac HT A	10584027	ThermoFischer Scientific/Gibco
<i>PLASMIDS GENERATED IN THIS WORK</i>		
pFastBac HT(A) - mouse APOBEC2 (A2)		Linda Molla
pFastBac HT(A) -A2-E100A		Jose Paulo Lorenzo
pFastBac HT(A) -del(1-41)A2		Jose Paulo Lorenzo
pFastBac HT(A)-A2-rescue- G64R		Jose Paulo Lorenzo
pMx-A2-3xHA		Jose Paulo Lorenzo
pMx-del(1-40)A2-3xHA		Jose Paulo Lorenzo
pMx-A2-rescue		Linda Molla
pMx-Flag-A2-rescue		Linda Molla
pMx-Flag-NLS-A2-rescue		Jose Paulo Lorenzo
pMx-del(1-40)A2-rescue		Jose Paulo Lorenzo
pMx-Flag-del(1-40)A2-rescue		Jose Paulo Lorenzo
pMx-Flag-NLS-del(1-40)A2- rescue		Jose Paulo Lorenzo
pMx-A2-rescue-G64R		Jose Paulo Lorenzo
pMx-Flag-A2-rescue-G64R		Jose Paulo Lorenzo
pMx-Flag-NLS-A2-rescue- G64R		Jose Paulo Lorenzo
pMx-A2-E100A-rescue		Linda Molla
pMx-Flag-A2-E100A-rescue		Linda Molla
pMx-Flag-NLS-A2-E100A- rescue		Jose Paulo Lorenzo

5. Materials and Methods

Table 5. List of primers and oligonucleotides

PRIMERS/ OLIGONUCLEOTIDES	Experiment	Sequence (5'-3')
qPCR_Tbp_F	qPCR	GCTCTGGAATTGTACCGCAGC
qPCR_Tbp_R	qPCR	ATCAACGCAGTTGTCCGTGG
qPCR_Hprt_F	qPCR	TCAGTCAACGGGGGACATAAA
qPCR_Hprt_R	qPCR	GGGGCTGTACTGCTTAACCAG
qPCR_Id3_F	qPCR	GACAAGAGGAGCTTTTGCCAC
qPCR_Id3_R	qPCR	ATCCATGCCCTCAGGCTTC
qPCR_Kazald1_F	qPCR	GCTTCTGCAACCCTCACAGT
qPCR_Kazald1_R	qPCR	TCAGGAAACAAACTCCGTGATTG
qPCR_Hmx2_F	qPCR	CCAAGCCCTAAGCACCATAACC
qPCR_Hmx2_R	qPCR	TCCTTAAAGTCCGGGTGAGAAG
qPCR_Zfp219_F	qPCR	GCTACTCCAACGGACCAGG
qPCR_Zfp219_R	qPCR	CTGCACGAGTCTCAGACCAAC
SP/KLF motif	MST	GGC GCG GCC CCG CCC CCT CCT CC
poly-A motif	MST	TCT CAA GAA AAA AAA AAA AAG AC

Table 6. List of equipment

EQUIPMENT	CATALOG NUMBER	SOURCE
2-Gel Tetra and Blotting Module	1660827ED U	Bio-Rad
Centrifuge 5427 R	5409000010	Eppendorf
Centrifuge 5920 R	5948000018	Eppendorf
CFX Connect Real-Time PCR Detection System	1855201	Bio-Rad
CFX Connect Real-Time System	1855201	Bio-Rad
Chemidoc	17001401	Bio-Rad
Eclipse Ts2 Inverted Microscope + DS-Fi3 microscope camera	Ts2-FL + DS-Fi3	Nikon
In-VitroCell Direct Heat CO2 Incubator	NU-5820	Nuaire
Mini-PROTEAN Tetra Cell	1658004ED U	Bio-Rad
Monolith NT.115	NT115	NanoTemper Technologies
Multiskan™ FC Microplate Photometer	51119000	Thermo Fisher Scientific
NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer	ND-ONEC- W	Thermo Fisher Scientific
Pipetboy acu 2	155 019	Integra Biosciences
PIPETMAN Classic Starter Kit, 4 Pipette Kit, P2, P20, P200, P1000	F167380	Gilson
Safe 2020 Class II Biological Safety Cabinet	51026637	Thermo Fisher Scientific
T100 Thermal Cycler	1861096	Bio-Rad
Thermo Scientific™ Multifuge™ X3R	75004515	Thermo Fisher Scientific

5. Materials and Methods

5.1 Cell culture

C2C12 myoblasts (CRL-1772) were obtained from the American Type Culture Collection (ATCC). Cells were maintained in *proliferation medium*: DMEM, high glucose (Merck/Sigma-Aldrich) supplemented with 10% fetal bovine serum (ThermoFischer Scientific/Gibco) and 1% penicillin-streptomycin (Merck/Sigma-Aldrich). Cells were regularly passaged before reaching full confluency. For passage, cells were washed with 1X DPBS (Lonza) then detached with Trypsin-EDTA solution (Merck/Sigma-Aldrich) for 8 to 10 minutes until cells were fully separated. Trypsin-EDTA was inactivated by adding 2X volume proliferation medium before pelleting cells at 350 xg for 5 minutes. Cells were seeded at a high density 16 to 20 h prior to inducing myoblast to myotube differentiation, $\sim 2.6 \times 10^4$ cells/cm². Cells were induced to differentiate by replacing proliferation medium with *differentiation medium*: DMEM, high glucose supplemented with 2% horse serum (ThermoFischer Scientific/Gibco) and 1% penicillin-streptomycin.

NIH3T3 (CRL-1658) and 293T (CRL-3216) cells were obtained from the ATCC. Both NIH3T3 and 293T cells were maintained in proliferation medium, similar to C2C12 myoblasts. Passaging followed the same protocol of C2C12 myoblasts.

5.2 Production of APOBEC2 knockdown C2C12 myoblasts

C2C12 myoblasts were infected with lentiviruses carrying shRNA, targeting either mouse APOBEC2 or GFP, as a non-targeting control. C2C12 myoblasts were seeded at low density for transductions, ~ 2800 cells/cm². The shRNA pLKO.1-puro constructs were obtained from The Broad Institute's Mission TRC-1 mouse library. The plasmids used were pLKO.1-APOBEC2-shRNA (TRC Clone ID: TRCN0000112019) and pLKO.1 puro GFP siRNA (Addgene, # 12273) (Orimo et al., 2005). The following sequences were targeted: APOBEC2 shRNA - GCTACCAGTCAACTTCTTCAA and GFP shRNA - GCAAGCTGACCCTGAAGTTCA.

Virions were produced by co-transfection of pLKO.1-puro shRNA containing construct, packaging plasmid psPAX2 (Addgene, #12260) and envelope plasmid pMD2.g (Addgene, #12259) in 293T cells (CRL-3216, ATCC). Transfections were done using Lipofectamine 2000 transfection reagent (ThermoFisher Scientific/Invitrogen) following the recommended protocol. Supernatants with lentiviral particles were collected at 24 and 48 hours (h) after transfection. The supernatants were filtered through a 0.45 μ m filter

(Merck/Millipore) and before adding to C2C12 myoblasts. For APOBEC2 constitutive knockdown, C2C12 cells were infected with pLKO.1 containing lentiviruses in growth media containing 8 µg/mL polybrene. A day after transduction, the medium was replaced with proliferation medium. Puromycin selection was done 48 h post-transduction by replacing the medium with proliferation medium containing 4 µg/ml puromycin (ThermoFischer Scientific/Gibco). The selection was done for 48 h to select for cells with stable shRNA expression. Stable cells were maintained for a couple of passages in the selection medium to ensure integration. Knockdown of APOBEC2 in cell lines was then confirmed by detection of APOBEC2 protein expression.

5.3 Molecular cloning of APOBEC2 overexpression constructs

Original pMX-APOBEC2 and pMX-APOBEC2 rescue constructs were provided by Linda Molla. The backbone construct, pMX, was derived from the original pMXs-IRES-GFP (pMXs-IG) with a puromycin selection marker driven by an SV40 promoter (Kitamura et al., 2003). The pMXs-APOBEC2 rescue construct was designed to escape knockdown by APOBEC2 shRNA by introducing silent mutations in the shRNA recognition sequence. These constructs were used as a template for making the pMXs-del(1-40)APOBEC2 and the point mutation constructs: pMXs-APOBEC2-G64R, pMXs-APOBEC2-E100A, and pMXs-APOBEC2-T120N. N-terminal truncation was done through PCR with primers designed to remove the first 40 amino acids. To introduce the point mutations, site-directed mutagenesis was done using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs, Inc.). Insertion of a Flag-tag or nuclear localization signal (NLS) was done by inserting short oligonucleotides containing the desired sequences with the NEBuilder HiFi DNA Assembly kit (New England Biolabs, Inc.). The following sequences were used: Flag-GG linker: GAC TAT AAA GAC GAT GAC GAC AAG-GGA GGA and FLAG-NLS-GG linker GAC TAT AAA GAC GAT GAC GAC AAG-CCA AAG AAG AAG CGT AAG GTA-GGA GGA. APOBEC2 transgene constructs were subcloned into the pMX backbone to avoid unintended mutations in the vector backbone sequences.

5. Materials and Methods

5.4 Retrovirus production and transduction of cells

Virions were produced by co-transfection of pMXs constructs with pCL-Eco retrovirus packaging vector (Addgene) in 293T cells. Transfections were done using Lipofectamine 2000 transfection reagent (ThermoFisher Scientific/Invitrogen) following the recommended protocol. For a T-25 flask with 293T cells at ~70% confluency, 20 μ L Lipofectamine 2000 was diluted in 500 μ L OptiMEM (ThermoFischer Scientific/Gibco) and incubated at room temperature for 5 minutes. In a separate tube, 8 μ g of pMXs construct + 8 μ g of pCL-Eco were diluted in 500 μ L OptiMEM. The Lipofectamine solution was then added and the mixture was incubated for 20 minutes at room temperature. The transfection mixture was then added to the 293T cells, 1 mL transfection mixture plus 4 mL proliferation medium. The cell medium was changed after 6 h. After 48 h, the supernatant containing the virions was collected and debris was separated by pelleting at 1000 \times g for 2 minutes. The cleared supernatant was then filtered with a syringe filter, 0.45 μ m filter (Merck/Millipore). Aliquots of the virions were flash-frozen in liquid nitrogen and stored at -80°C.

For transduction, cells were plated 16-20 h before transduction at, ~2800 cells/cm². Virions were thawed quickly in a water bath set at 37°C. For a 12-well plate, each well was changed to received 500 μ L proliferation medium + 0.4 μ L polybrene. Then, 250 μ L virion aliquot + 250 μ L proliferation medium + 0.4 μ L polybrene was added to each well. Medium change was done 24 h later. GFP expression is visible by 48 h post-transduction. Puromycin selection was done 48 h post-transduction by replacing the medium with proliferation medium containing 4 μ g/ml puromycin. Stable cell lines were propagated and stored in liquid nitrogen.

5.5 Microscopy – fusion index and APOBEC2 localization

Cells were seeded on plain coverslips or collagen-coated coverslips (BD Biosciences) for C2C12 cells. For differentiation, C2C12 cells were seeded at ~2.6x10⁴ cells/cm² a day before induction of differentiation. Otherwise, cells were seeded at around 1.3x10⁴ cells/cm² to avoid cell crowding during microscopy. At the experimental time points, cells were washed with cold PBS and fixed with paraformaldehyde (4%) in PBS for 10 minutes at 4°C. This was followed by two washes with PBS, 5 minutes each at room temperature. The coverslips with the cells were then submerged in

blocking/permeabilization buffer (0.5% BSA, 1% gelatin, 5% normal goat or donkey serum, and 0.1% Triton X-100 in PBS) for 1 hour at room temperature. This was followed by an overnight stain with primary antibodies diluted in *blocking/permeabilization buffer* in a humidified chamber at 4°C. After staining, the coverslips were washed thrice with PBS, 5 minutes each at room temperature. Coverslips were then incubated with fluorescent secondary antibodies diluted in *blocking/permeabilization buffer* for 1 hour at room temperature. Afterward, coverslips were washed thrice with PBS for 5 minutes at room temperature. Immunofluorescence staining of C2C12 and NIH3T3 cells was carried out with primary antibodies: MyHC MF20 (DSHB) and FLAG M2 (Sigma). Nuclei were counterstained and coverslips were mounted with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories, H-1200). Images were taken using widefield Zeiss Cell Observer (DKFZ Light Microscopy Facility) and image processing and analysis were done through Fiji/ImageJ (Schindelin et al., 2012). Fiji macros for analysis and quantification were designed with the guidance of Dr. Damir Kronic from the DKFZ Light Microscopy Facility. For calculating the nuclear enrichment of APOBEC2, nuclear masks and dilated nuclear masks were used to calculate the ratio of nuclear to cytoplasmic mean fluorescence intensity (Kelley & Paschal, 2019).

5.6 Analysis of RNA-Seq data from APOBEC2 knockdown cells

The transcript read counts were obtained through Salmon (Patro et al., 2017) from the processed paired-end reads aligned to the mm10 genome. The libraries for sequencing were prepared with the TruSeq Stranded mRNA Sample Prep kit (Illumina) and sequenced on the HiSeq2500 (Illumina), 50 bp paired-read sequencing. Library preparation, sequencing, read processing, and alignment were originally performed by Linda Molla and the Rockefeller University Genomics Resource Center. The RNA-Seq data is accessible through the Gene Expression Omnibus: GSE117730.

For differential expression analysis, the read counts were processed as input to the DESeq2 package (Love et al., 2014). Biological triplicates were present for each condition (shGFP and shA2) across the three time points: day 0, day 1, and day 2 of differentiation. A grouped design with condition and time point was used for the DESeq2 analysis. Differentially expressed genes were obtained from each time point through contrasts of corresponding groups; shA2.day2 versus shGFP.day2. MA plots for each contrast were prepared using the DESeq2 built-in function, plotMA. Gene ontology (GO) enrichment

5. Materials and Methods

was done through the clusterProfiler package (Wu et al., 2021). GO over-representation analysis was done through the enrichGO tool with arguments: `OrgDb = org.Mm.eg.db`, `ont = "BP"` and `pAdjustMethod = "fdr"`. Dot plots of GO terms and GO cluster maps were made using the clusterProfiler built-in `compareCluster` and `emapplot` functions respectively.

5.7 Analysis of ChIP-Seq data from differentiating C2C12 myoblasts

The ChIP-Seq data used in this thesis was from work by Linda Molla. The following protocol was adapted from a preprint written from this work (Lorenzo et al., 2021). For chromatin immunoprecipitation, C2C12 cells were plated at ~70% confluence 12 hours prior to inducing differentiation. After inducing to differentiate, cells were processed at 14-hour and 34-hour time points. Directly on the plate, cells were fixed with 1% PFA in PBS for 10 minutes at room temperature. Glycine (125 mM) was added to the cells to stop fixing. Cells were washed twice with PBS containing protease inhibitor cocktail (PIC) (Roche, 11836170001). Cells were lysed with cold *Farnham lysis buffer* (5mM PIPES pH 8.0, 0.5% NP-40, 85mM KCl, 1mM EDTA, PIC) and incubated rotating for 15 minutes at 4°C. Lysates were scraped off the plates, pelleted, and resuspended in *LB2 buffer* (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. Lysates were centrifuged and the pellets were resuspended in *LB3 buffer* (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 with a Covaris ultrasonicator model S220 for 15 minutes with the following settings: 140W peak power, 5% duty, 200 cycles per burst. Triton X-100 was added to the samples. Samples were clarified by centrifugation at 20,000 xg for 10 minutes at 4°C. The supernatant is the soluble chromatin extract. The soluble fragmented chromatin from $\sim 2.5 \times 10^7$ was used for each IP. For each IP 100ul Dynabeads (Thermofisher anti-rabbit M280) were mixed with 10ul polyclonal rabbit-APOBEC2 antibodies (from Dr. Alin Vonica MD, Ph.D.) 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 for 30 minutes at 65°C while shaking. 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).

The ChIP-Seq library preparation, sequencing, and analysis were performed by Linda Molla and the Bioinformatics Resource Center at Rockefeller University. Briefly here, the ChIP-Seq included biological triplicates for each time point, differentiation at 14-hour and 34-hour. ChIP-Seq libraries were prepared using NEBNext Ultra DNA Library Prep Kit. Libraries were sequenced with 75 base pair single-read sequencing on the NextSeq 500 (Illumina). Single-end reads were aligned to the mm10 genome. Reads mapping to more than one genomic location were filtered before peak calling using Model-based Analysis of ChIP-Seq (MACS2) tool with duplicate filtering applied and corresponding input DNA sample as a control (Feng et al., 2011). The ChIP-Seq data is accessible through the Gene Expression Omnibus: GSE117729.

For the Differential Binding analysis done in this thesis, the DiffBind package was used (Ross-Innes et al., 2012). The DiffBind tool takes ChIP-Seq peaks and read alignment files (.bam) as input for the analysis. APOBEC2 ChIP-Seq peaks called by the MACS2 analysis were used. For the aligned reads, the ChIP-Seq fasta files were aligned to the mm10 (UCSC) genome using bwa (Li & Durbin, 2009). Before running the DiffBind analysis, the APOBEC2 ChIP-Seq peaks were filtered using the mm10 blacklist and “graylist” derived from the corresponding input sample. For the DiffBind analysis, the peaks at 34-hour and 14-hour time points were contrasted. The resulting differentially bound peaks were annotated to the mm10 (UCSC) genome using the ChIPseeker package (Yu et al., 2015). Venn diagrams and overlap of peaks were done using built-in tools in DiffBind. Bar plots on the annotation of the peaks were done using the plotAnnoBar tool in ChIPseeker. Gene ontology enrichment (over-representation) analysis was done using the enrichGO tool in clusterProfiler. GO over-representation analysis was done through the enrichGO tool with arguments: `OrgDb = org.Mm.eg.db`, `ont = “BP”` and `pAdjustMethod = “fdr”`.

5. Materials and Methods

5.8 Extracting cytoplasmic and nuclear protein fractions from C2C12 cells

C2C12 cells were plated at high density for differentiation, $\sim 2.6 \times 10^5$ cells/cm² in a 15-cm tissue culture dish. On the following day, cells were induced to differentiate by switching from *proliferation medium* to *differentiation medium*. Cells were allowed to differentiate for up to 4 days, changing *differentiation medium* every 2 days. On day 4, once C2C12 myoblasts have sufficiently differentiated into myotubes, cells were detached with a trypsin-EDTA solution for 10 minutes. Cells were collected by centrifugation (350 xg for 5 minutes) and washed once with DPBS. Cells were pelleted (350 xg for 5 minutes) and the DPBS wash was aspirated. Cell pellets were kept on ice at all proceeding steps.

Cell pellets were resuspended in 10 volumes (based on packed cell volume) of *Tween 20 lysis buffer*: 0.5% Tween 20, 50 mM Tris-Cl pH 7.5, 2 mM EDTA, and 1X protease and phosphatase inhibitor cocktail in sterile deionized water. Typically for one 15-cm dish, you obtain about 100 μ L of cells so you would resuspend the pellet in 1000 μ L of *Tween 20 lysis buffer*. Resuspended cells were vortexed at the highest settings for 15 seconds before incubating on ice for 10 minutes. This step was done twice, and cell membrane lysis and separation of nuclei were confirmed under the microscope.

Nuclei were pelleted by centrifuging at 6000 xg for 1 minute at 4°C. The supernatant containing the cytoplasmic protein fraction was taken, flash-frozen in liquid nitrogen, and stored at -80°C. Nuclei were washed once in ice-cold DPBS to remove traces of the cytoplasmic protein fraction. Once the DPBS wash was completely removed, the nuclei were resuspended in 2.5 volumes of *high salt nuclear lysis buffer*: 800 mM NaCl, 50 mM Tris-Cl pH 7.5, 1% NP40 (Igepal CA-630), and 1X protease and phosphatase inhibitor cocktail in sterile deionized water. Resuspended nuclei were vortexed at the highest settings for 15 seconds before incubating on ice for 10 minutes. This step was done twice, and the lysis of nuclei was confirmed under the microscope.

Nuclear protein lysates were diluted in 1 volume of 50 mM Tris-Cl pH 7.5 and 1X protease and phosphatase inhibitor cocktail in sterile deionized water before adding 100 U/mL of Benzonase and 2 mM MgCl₂. Nuclear protein lysates with Benzonase were incubated for 60 minutes with constant rotation at 4°C. Nuclear protein lysates were then centrifuged at 16000 xg for 10 minutes at 4°C to separate cell debris. The supernatant contained the nuclear protein lysate was taken, flash-frozen in liquid nitrogen, and stored at -80°C.

The separation of the nuclear and cytoplasmic protein fraction was assessed by Western blotting. SP1 and histone H3 were used for “nuclear and chromatin only” protein markers while α -tubulin was used as a “cytoplasmic only” protein marker. See methods for details of Western blotting and a list of antibodies.

5.9 Co-immunoprecipitation

Magnetic beads, Dynabeads M-280 Sheep anti-mouse or anti-rabbit IgG (Thermo), were coupled with antibodies against target protein for immunoprecipitation. A volume of 25 μ L magnetic beads was pre-washed in DPBS with 0.5% bovine serum albumin (BSA). The beads were then incubated with 2 to 4 μ g of antibody in 0.5% BSA overnight with constant turning at 4°C. Antibodies used were: rabbit anti-APOBEC2 (Sigma), rabbit anti-SP1 (Merck), or anti-Flag M2 (Sigma). The respective mouse or rabbit IgG isotype controls were used as negative controls for immunoprecipitation.

The separated nuclear protein fraction was pre-cleared on 25 μ L of pre-washed magnetic beads. For each co-IP, an input of 250 μ g of nuclear protein fraction was diluted in 50 mM Tris pH 7.5 and 1X protease and phosphatase inhibitor cocktail. For the co-IP buffer, a final salt concentration ranging from 50 to 150 mM NaCl was used. Trace amounts of NP40, <0.5%, carried over from the protein fraction preparation may be present in the final buffer. The pre-cleared nuclear lysate was then added to the 25 μ L of the beads coupled with the respective antibodies. Additionally, 25 μ L of the input (10%) was set aside for Western blotting. Immunoprecipitation was done overnight with constant rotation at 4°C. On the following day, the lysate with the antibody-coupled magnetic beads was placed on the magnetic rack, and 25 μ L of the supernatant (flow-through) was taken for Western blotting. Beads were washed thrice in co-IP wash buffer: 50 mM Tris-Cl pH 7.5, 50 to 150 mM NaCl, 0.1% NP40, and 1X protease and phosphatase inhibitor cocktail in sterile deionized water. The washed beads were directly boiled, at 95°C for 5 minutes, in 36 μ L of 1X Laemmli buffer, *4X Laemmli buffer*: 250 mM Tris base, 8% SDS, 40% glycerol, 20% beta-mercaptoethanol, and 0.2% bromophenol blue. The supernatant of the boiled beads was used as the co-IP eluate for Western blotting.

5. Materials and Methods

5.10 Western blotting

For Western blotting (or immunoblotting) experiments, whole protein lysates or the co-IP fractions (input, flowthrough, and eluate) were used. Whole protein lysates were prepared from cells using RIPA (radioimmunoprecipitation assay) lysis buffer: 50 mM Tris pH 8.0, 150 mM NaCl, 1.0% Triton X-100, 0.5% sodium deoxycholate, and 0.1% SDS in sterile deionized water. Cell pellets were treated with RIPA buffer for 60 minutes with constant rotation at 4°C. Lysates were then collected and snap-frozen in liquid nitrogen then stored at -80°C. After thawing the lysates on ice, the lysates were cleared by centrifugation at 16000 xg for 15 minutes.

The same amount of total protein, 10 to 30 µg, was loaded into each well; for the co-IP experiments, 10% input and all the flowthrough or eluate fractions were loaded into each well. Before loading, samples were boiled, at 95°C for 5 minutes, in *IX Laemmli buffer*. Samples were run on a 4-20% polyacrylamide gel (Mini-PROTEAN® TGX™ Precast Protein Gels, Bio-Rad) in *TG buffer* (25 mM Tris and 192 mM Glycine) with and 0.1% SDS. After the run, proteins were blotted on nitrocellulose membranes using a wet transfer system (Bio-Rad) in cold *TG buffer* with 20% methanol. Blots were immersed in *blocking buffer: TBS-T buffer* (20 mM Tris-Cl pH 7.6, 150 mM NaCl, and 0.1% Tween 20) with 5% bovine serum albumin (BSA). Blocking was done for 60 minutes with constant agitation at room temperature. After blocking, blots were immersed in primary antibody diluted in *blocking buffer* overnight with constant agitation at 4°C. After antibody incubation, blots were washed thrice in *TBS-T buffer*. Washed blots were then incubated in respective HRP-conjugated anti-mouse or anti-rabbit secondary antibodies diluted in *blocking buffer* for 60 minutes with constant agitation at room temperature. Blots were then washed thrice in *TBS-T buffer* before developing with HRP-substrate (SuperSignal™ West Femto Maximum Sensitivity Substrate, Thermo). Blots were imaged with a chemiluminescent detection system (ChemiDoc, Bio-Rad). The details of the antibodies used are in **Table 3**.

5.11 Reverse transcription-quantitative PCR (RT-qPCR)

Total RNA was extracted from respective cells using the RNeasy mini kit (Qiagen). The total RNA was DNase treated using the TURBO DNA-free kit (Thermo Invitrogen) following the standard protocol. A total of 750 ng of RNA from each sample

was converted to cDNA using ProtoScript® First Strand cDNA Synthesis Kit (New England Biolabs) following the standard protocol with the provided Random Primer Mix.

The qPCR reaction was done using 15 ng of the cDNA in each reaction with the iTaq Universal SYBR Green Supermix (BioRad) following the standard protocol. The reactions were performed using a thermocycler, the CFX Connect Real-Time System (BioRad). Reactions were done in technical triplicates for each biological triplicate. Primers for target transcripts were taken from the mouse PrimerBank database and validated independently (Wang et al., 2012). The Hprt (hypoxanthine guanine phosphoribosyl transferase) transcript was used as a housekeeping gene for calculating ΔCq . The fold change was calculated by $2^{-\Delta\Delta Cq}$ method, ΔCq control – ΔCq treated. Cq values and fold change were calculated using the CFX Maestro software (BioRad). Appropriate statistics were done using the Prism 9 software.

5. Materials and Methods

5.12 Recombinant mouse APOBEC2 production

His₆-tagged mouse APOBEC2 and mutants were cloned into pFastBacHT(A) (Thermo) for recombinant protein production in insect cells. Recombinant protein production was carried out by the EMBL Protein Expression and Purification Core Facility (PEPCore). Briefly, these pFastBac-APOBEC2 constructs were used for transposition into E.coli DH10EMBacY cells (Geneva Biotech). The isolated bacmid DNA was used for generating recombinant baculoviruses. The following protocol for mouse protein production was adapted from the preprint written from this work and was provided by Dr. Kim Remans of PEPCore and (Lorenzo et al., 2021). For the mouse APOBEC2 protein production, 5 mL of baculovirus was used to infect 1 L of Sf21 cells at a density of 1×10^6 cells/ml. After 72 h, the cells were harvested by centrifugation (30 min, 600 x g, 4°C) and resuspended in lysis buffer (20 mM Tris pH 8.0, 800 mM NaCl, 20 mM imidazole, and 5 mM β-mercaptoethanol) supplemented with Benzonase (Merck), 10 mM MgCl₂ and cComplete EDTA-free protease inhibitors (Roche). The cells were lysed using a Dounce homogenizer and the lysate was cleared by centrifugation (30 min, 20000 x g, 4°C). The cleared lysate was loaded onto a 5 mL Ni-NTA column (Macherey-Nagel). After washing with a buffer consisting of 20 mM Tris pH 8.0, 300 mM NaCl, 20 mM imidazole, and 5 mM β-mercaptoethanol, the Ni-NTA column was eluted using a gradient of up to 300 mM imidazole. The elution fractions containing mouse APOBEC2 were pooled and dialyzed overnight at 4°C to ion exchange buffer (20 mM Tris pH 8.0, 100 mM NaCl, 1 mM DTT). The dialyzed sample was loaded onto a 5 mL HiTrap Heparin HP column (Cytiva) coupled to a 5 mL HiTrap Q HP anion exchange column (Cytiva). After washing, the HiTrap Heparin HP column and the HiTrap Q HP column were eluted separately in a gradient ranging from 100 mM to 1M NaCl. Finally, the mouse APOBEC2 protein eluted from the HiTrap Q HP column was subjected to a size exclusion chromatography (SEC) step using a HiLoad 16/600 Superdex 75 pg column (Cytiva) pre-equilibrated with SEC buffer (20 mM HEPES pH 7.5, 150 mM NaCl, and 0.5 mM TCEP). When removal of the His₆-tag was required, His₆-tagged TEV protease (produced in-house) was added to the purified mouse APOBEC2 protein. After the overnight TEV cleavage step at 4°C, Ni-NTA beads (Qiagen) were added to the sample and incubated for 1h at 4°C. After centrifugation (1 min, 100 x g, 4°C), untagged mouse APOBEC2 was recovered from the flow-through of the Ni-NTA beads. Recombinant mouse APOBEC2 proteins were aliquoted, flash-frozen with liquid N₂, and stored at -80°C.

5.13 Microscale Thermophoresis (MST)

Purified recombinant mouse APOBEC2 and mutant proteins were labeled using Cy5 Mono NHS Ester (GEPA15101, Sigma-Aldrich). Recombinant proteins were diluted to 5 mg/mL in 20 mM HEPES pH 7.5, 150 mM NaCl and 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP). Cy5 Mono NHS Ester dye was diluted in DMSO at 5 mg/mL. Labeling reactions with a 3:1 dye to protein molar ratio were performed in 20 mM HEPES, pH 7.5, 150 mM NaCl, and 0.5 mM TCEP. The reactions were incubated overnight at 4°C with constant agitation. After incubation, reactions were deactivated using quencher buffer (ab102884, Abcam). The remaining dye was washed away by concentrating protein using Vivaspin® 500 centrifugal concentrators (Sartorius) into MST buffer (10 mM HEPES pH 7.5, 50 mM NaCl, 5 mM MgCl₂). The degree of labeling ranged from 1.2 to 1.5 dye:protein. The following oligonucleotides (oligos) were purchased from Sigma-Aldrich: SP/KLF motif F: GGC GCG GCC CCG CCC CCT CCT CC and A-tract motif: TCT CAA GAA AAA AAA AAA AAG AC.

The oligos were diluted in nuclease-free water. For annealing oligos, equimolar quantities of complementary oligos were mixed in MST buffer (10 mM HEPES, 50 mM NaCl, 5 mM MgCl₂). Annealing was done by incubating at 95°C for 5 minutes on a heat block. The heat block was then turned off and allowed to slowly cool to room temperature. Oligos were then stored at 4°C.

For the MST experiment, oligonucleotides were diluted in MST buffer supplemented with 0.05% Tween-20 for the final reaction. The Cy5-labeled APOBEC2 or mutant was held constant at 50 nM while the oligonucleotides were titrated (1:1) between 6 to 50,000 nM. Reactions were incubated for 30 min before loading into standard glass capillaries (MO-KO22, NanoTemper Technologies). MST measurements were performed using a Monolith NT.115 (NanoTemper Technologies) at 85% LED power and 40% MST power. MO.Affinity Analysis v2.1.3454 (NanoTemper Technologies) was used for curve fitting and calculating K_d values with Thermophoresis + T Jump settings. The data represent means from two to three independent experiments.

5. Materials and Methods

6. References

6. References

- Agrawal, P., Lin, C., Mathad, R. I., Carver, M., & Yang, D. (2014). The Major G-Quadruplex Formed in the Human BCL-2 Proximal Promoter Adopts a Parallel Structure with a 13-nt Loop in K⁺ Solution. *Journal of the American Chemical Society*, *136*(5), 1750–1753. <http://doi.org/10.1021/ja4118945>
- 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>
- Bailey, T. L. (2021). STREME: accurate and versatile sequence motif discovery. *Bioinformatics*, (March), 1–7. <http://doi.org/10.1093/bioinformatics/btab203>
- Beffagna, G. (2019). Zebrafish as a Smart Model to Understand Regeneration After Heart Injury: How Fish Could Help Humans. *Frontiers in Cardiovascular Medicine*, *6*(August), 1–8. <http://doi.org/10.3389/fcvm.2019.00107>
- Bentzinger, C. F., Wang, Y. X., & Rudnicki, M. A. (2012). Building muscle: molecular regulation of myogenesis. *Cold Spring Harbor Perspectives in Biology*, *4*(2). <http://doi.org/10.1101/cshperspect.a008342>
- 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>
- 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>
- Boyaci, H., Chen, J., Jansen, R., Darst, S. A., & Campbell, E. A. (2019). Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature*, *565*(7739), 382–385. <http://doi.org/10.1038/s41586-018-0840-5>
- Carrió, E., & Suelves, M. (2015). DNA methylation dynamics in muscle development and disease. *Frontiers in Aging Neuroscience*, *7*(MAR), 19. <http://doi.org/10.3389/fnagi.2015.00019>
- Chen, B., Han, B. H., Sun, X. H., & Lim, R. W. (1997). Inhibition of muscle-specific gene expression by Id3: Requirement of the C-terminal region of the protein for stable expression and function. *Nucleic Acids Research*, *25*(2), 423–430. <http://doi.org/10.1093/nar/25.2.423>
- 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>
- Conticello, S. G., Thomas, C. J. F., Petersen-Mahrt, S. K., & Neuberger, M. S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Molecular Biology and Evolution*, *22*(2), 367–77. <http://doi.org/10.1093/molbev/msi026>
- Degasperi, A., Birtwistle, M. R., Volinsky, N., Rauch, J., Kolch, W., & Kholodenko, B. N. (2014). Evaluating Strategies to Normalise Biological Replicates of Western Blot Data. *PLoS ONE*, *9*(1), e87293. <http://doi.org/10.1371/journal.pone.0087293>
- Du, Z., Zhao, Y., & Li, N. (2008). Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Research*. <http://doi.org/10.1101/gr.6905408.2>
- Etard, C., Roostalu, U., & Strähle, U. (2010). Lack of Apobec2-related proteins causes a

- dystrophic muscle phenotype in zebrafish embryos. *The Journal of Cell Biology*, *189*(3), 527–39. <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>
- Fossat, N., Tourle, K., Radziewicz, T., Barratt, K., Liebhold, D., Studdert, J. B., ... Tam, P. P. L. (2014). C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO Reports*, *15*(8), 903–910. <http://doi.org/10.15252/EMBR.201438450>
- Gómez-Del Arco, P., Perdiguero, E., Yunes-Leites, P. S., Acín-Pérez, R., Zeini, M., Garcia-Gomez, A., ... Redondo, J. M. (2016). The Chromatin Remodeling Complex Chd4/NuRD Controls Striated Muscle Identity and Metabolic Homeostasis. *Cell Metabolism*, *23*(5), 881–92. <http://doi.org/10.1016/j.cmet.2016.04.008>
- 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>
- Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., & Noble, W. (2007). Quantifying similarity between motifs. *Genome Biology*, *8*(2), R24. <http://doi.org/10.1186/gb-2007-8-2-r24>
- 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., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., ... Malim, M. H. (2003). DNA deamination mediates innate immunity to retroviral infection. *Cell*, *113*(6), 803–809. [http://doi.org/10.1016/S0092-8674\(03\)00423-9](http://doi.org/10.1016/S0092-8674(03)00423-9)
- 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](http://doi.org/10.1016/S1097-2765(02)00742-6)
- Hayward, J. A., Tachedjian, M., Cui, J., Cheng, A. Z., Johnson, A., Baker, M. L., ... Tachedjian, G. (2018). Differential Evolution of Antiretroviral Restriction Factors in Pteropid Bats as Revealed by APOBEC3 Gene Complexity. *Molecular Biology and Evolution*, *35*(7), 1626–1637. <http://doi.org/10.1093/molbev/msy048>
- Holland, S. J., Berghuis, L. M., King, J. J., Iyer, L. M., Sikora, K., Fifield, H., ... Boehm, T. (2018). Expansions, diversification, and interindividual copy number variations of AID/APOBEC family cytidine deaminase genes in lampreys. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), E3211–E3220. <http://doi.org/10.1073/pnas.1720871115>
- Iyer, L. M., Zhang, D., Rogozin, I. B., & Aravind, L. (2011). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Research*, *39*(22), 9473–9497. <http://doi.org/10.1093/nar/gkr691>
- Kelley, J. B., & Paschal, B. M. (2019). Fluorescence-based quantification of nucleocytoplasmic transport. *Methods*, *157*(12), 106–114. <http://doi.org/10.1016/j.ymeth.2018.11.002>
- Kelly, R. D. W., & Cowley, S. M. (2013). The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. *Biochemical Society*

6. References

- Transactions*, 41(3), 741–749. <http://doi.org/10.1042/BST20130010>
- Kitamura, T., Koshino, Y., Shibata, F., Oki, T., Nakajima, H., Nosaka, T., & Kumagai, H. (2003). Retrovirus-mediated gene transfer and expression cloning: Powerful tools in functional genomics. *Experimental Hematology*, 31(11), 1007–1014. [http://doi.org/10.1016/S0301-472X\(03\)00260-1](http://doi.org/10.1016/S0301-472X(03)00260-1)
- Kolde, R. (2019). Package ‘pheatmap’: pretty heat map, 1–8.
- 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>
- Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Bandle, R., Nelson, S., ... Casellas, R. (2013). Global Regulation of Promoter Melting in Naive Lymphocytes. *Cell*, 153(5), 988–999. <http://doi.org/10.1016/j.cell.2013.04.033>
- Krishnan, A., Iyer, L. M., Holland, S. J., Boehm, T., & Aravind, L. (2018). Diversification of AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 115(14), E3201–E3210. <http://doi.org/10.1073/pnas.1720897115>
- 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>
- 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>
- Lellek, H., Kirsten, R., Diehl, I., Apostel, F., Buck, F., & Greeve, J. (2000). Purification and Molecular Cloning of a Novel Essential Component of the Apolipoprotein B mRNA Editing Enzyme-Complex *. *Journal of Biological Chemistry*, 275(26), 19848–19856. <http://doi.org/10.1074/JBC.M001786200>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <http://doi.org/10.1093/BIOINFORMATICS/BTP324>
- 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>
- Liu, M. C., Liao, W. Y., Buckley, K. M., Yang, S. Y., Rast, J. P., & Fugmann, S. D. (2018). AID/APOBEC-like cytidine deaminases are ancient innate immune mediators in invertebrates. *Nature Communications*, 9(1), 1–11. <http://doi.org/10.1038/s41467-018-04273-x>
- 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>
- Lorenzo, J. P., Molla, L., Ibarra, I. L., Ruf, S., Ridani, J., Subramani, P. G., ... Papavasiliou, F. N. (2021). APOBEC2 is a Transcriptional Repressor required for proper Myoblast Differentiation. *BioRxiv*, 2020.07.29.223594. <http://doi.org/10.1101/2020.07.29.223594>
- 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>
- Maiti, A., Myint, W., Kanai, T., Delviks-Frankenberry, K., Sierra Rodriguez, C., Pathak, V. K., ... Matsuo, H. (2018). Crystal structure of the catalytic domain of HIV-1 restriction factor APOBEC3G in complex with ssDNA. *Nature Communications*, 9(1), 1–11. <http://doi.org/10.1038/s41467-018-04872-8>
- Maizels, N. (2006). Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nature Structural & Molecular Biology*, 13(12), 1055–1059. <http://doi.org/10.1038/nsmb1171>
- Mal, A., & Harter, M. L. (2003). MyoD is functionally linked to the silencing of a muscle-specific regulatory gene prior to skeletal myogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 100(4), 1735–1739. <http://doi.org/10.1073/pnas.0437843100>
- 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>
- Matsumoto, T., Marusawa, H., Endo, Y., Ueda, Y., Matsumoto, Y., & Chiba, T. (2006). Expression of APOBEC2 is transcriptionally regulated by NF- κ B in human hepatocytes. *FEBS Letters*, 580(3), 731–735. <http://doi.org/10.1016/j.febslet.2005.12.081>
- McMahon, D. K., Anderson, P. A., Nassar, R., Bunting, J. B., Saba, Z., Oakeley, A. E., & Malouf, N. N. (1994). C2C12 cells: biophysical, biochemical, and immunocytochemical properties. *American Journal of Physiology-Cell Physiology*, 266(6), C1795–C1802. <http://doi.org/10.1152/ajpcell.1994.266.6.C1795>
- Mehta, A., Kinter, M. T., Sherman, N. E., & Driscoll, D. M. (2000). Molecular Cloning of Apobec-1 Complementation Factor, a Novel RNA-Binding Protein Involved in the Editing of Apolipoprotein B mRNA. *Molecular and Cellular Biology*, 20(5), 1846–1854. <http://doi.org/10.1128/MCB.20.5.1846-1854.2000>
- 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>
- Münk, C., Willemsen, A., & Bravo, I. G. (2012). An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC Evolutionary Biology*, 12(1). <http://doi.org/10.1186/1471-2148-12-71>
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, 102(5), 553–63. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11007474>
- Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., & Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *The Journal of Biological Chemistry*, 274(26), 18470–6. <http://doi.org/10.1074/jbc.274.26.18470>
- 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>
- Ohtsubo, H., Sato, Y., Suzuki, T., Mizunoya, W., Nakamura, M., Tatsumi, R., & Ikeuchi, Y. (2017). APOBEC2 negatively regulates myoblast differentiation in muscle regeneration. *The International Journal of Biochemistry & Cell Biology*, 85, 91–101. <http://doi.org/10.1016/j.biocel.2017.02.005>

6. References

- 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>
- Park, S.-Y., & Kim, J.-S. (2020). A short guide to histone deacetylases including recent progress on class II enzymes. *Experimental & Molecular Medicine*, *52*(2), 204–212. <http://doi.org/10.1038/s12276-020-0382-4>
- 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 & Molecular Biology*, *16*(5), 517–27. <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>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. <http://doi.org/10.1002/jcc.20084>
- Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., & Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. *Nature*, *533*(7603), 353–358. <http://doi.org/10.1038/nature17990>
- 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., Elsaedi, 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>
- Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F., & Chen, X. S. (2007). The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature*, *445*(7126), 447–451. <http://doi.org/10.1038/nature05492>
- Qiao, Q., Wang, L., Meng, F. L., Hwang, J. K., Alt, F. W., & Wu, H. (2017). AID Recognizes Structured DNA for Class Switch Recombination. *Molecular Cell*, *67*(3), 361–373.e4. <http://doi.org/10.1016/j.molcel.2017.06.034>
- 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>
- Rai, K., Sarkar, S., Broadbent, T. J., Voas, M., Grossmann, K. F., Nadauld, L. D., ... Jones, D. A. (2010). DNA Demethylase Activity Maintains Intestinal Cells in an Undifferentiated State Following Loss of APC. *Cell*, *142*(6), 930–942. <http://doi.org/10.1016/J.CELL.2010.08.030>
- Ratcliff, J., & Simmonds, P. (2021). Potential APOBEC-mediated RNA editing of the genomes of SARS-CoV-2 and other coronaviruses and its impact on their longer term

- evolution. *Virology*, 556(October 2020), 62–72.
<http://doi.org/10.1016/j.virol.2020.12.018>
- 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>
- Renault, M. A., Vandierdonck, S., Chapouly, C., Yu, Y., Qin, G., Metras, A., ... Gadeau, A. P. (2013). Gli3 regulation of myogenesis is necessary for ischemia-induced angiogenesis. *Circulation Research*, 113(10), 1148–1158.
<http://doi.org/10.1161/CIRCRESAHA.113.301546>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24–26. <http://doi.org/10.1038/nbt.1754>
- Rogozin, I. B., Basu, M. K., Jordan, I. K., Pavlov, Y. I., & Koonin, E. V. (2005). APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle (Georgetown, Tex.)*, 4(9), 1281–5. <http://doi.org/10.4161/cc.4.9.1994>
- Rogozin, I. B., Iyer, L. M., Liang, L., Glazko, G. V., Liston, V. G., Pavlov, Y. I., ... Pancer, Z. (2007). Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nature Immunology*, 8(6), 647–56. <http://doi.org/10.1038/ni1463>
- Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., ... Carroll, J. S. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 2011 481:7381, 481(7381), 389–393.
<http://doi.org/10.1038/nature10730>
- Rubio, M. A. T., Pastar, I., Gaston, K. W., Ragone, F. L., Janzen, C. J., Cross, G. A. M., ... Alfonzo, J. D. (2007). An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19), 7821–6.
<http://doi.org/10.1073/pnas.0702394104>
- 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), 578–594.
<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., Ohtsubo, H., Nihei, N., Kaneko, T., Sato, Y., Adachi, S., ... Yoshizawa, F. (2017). Apobec2 deficiency causes mitochondrial defects and mitophagy in skeletal muscle. *The FASEB Journal*, fj.201700493R. <http://doi.org/10.1096/fj.201700493R>
- 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. *The Journal of Biological Chemistry*, 285(10), 7111–8.
<http://doi.org/10.1074/jbc.M109.052977>
- Sawyer, S. L., Emerman, M., & Malik, H. S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biology*, 2(9), E275.
<http://doi.org/10.1371/journal.pbio.0020275>
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676–82. <http://doi.org/10.1038/nmeth.2019>

6. References

- 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>
- Szlachta, K., Thys, R. G., Atkin, N. D., Pierce, L. C. T., Bekiranov, S., & Wang, Y. H. (2018). Alternative DNA secondary structure formation affects RNA polymerase II promoter-proximal pausing in human. *Genome Biology*, 19(1). <http://doi.org/10.1186/s13059-018-1463-8>
- Teng, B., Burant, C. F., & Davidson, N. O. (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science (New York, N.Y.)*, 260(5115), 1816–9. <http://doi.org/10.1126/science.8511591>
- Tierney, M. T., Aydogdu, T., Sala, D., Malecova, B., Gatto, S., Puri, P. L., ... Sacco, A. (2014). STAT3 signaling controls satellite cell expansion and skeletal muscle repair. *Nature Medicine*, 20(10), 1182–1186. <http://doi.org/10.1038/nm.3656>
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220). <http://doi.org/10.1126/SCIENCE.1260419>
- Umansky, K. B., Gruenbaum-Cohen, Y., Tsoory, M., Feldmesser, E., Goldenberg, D., Brenner, O., & Groner, Y. (2015). Runx1 Transcription Factor Is Required for Myoblasts Proliferation during Muscle Regeneration. *PLoS Genetics*, 11(8), 1–31. <http://doi.org/10.1371/journal.pgen.1005457>
- von Maltzahn, J., Chang, N. C., Bentzinger, C. F., & Rudnicki, M. A. (2012). Wnt signaling in myogenesis. *Trends in Cell Biology*, 22(11), 602–609. <http://doi.org/10.1016/j.tcb.2012.07.008>
- von Walden, F. (2019). Ribosome biogenesis in skeletal muscle: coordination of transcription and translation. *Journal of Applied Physiology*, 127(2), 591–598. <http://doi.org/10.1152/jappphysiol.00963.2018>
- 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>
- Wang, X., Spandidos, A., Wang, H., & Seed, B. (2012). PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Research*, 40(D1), D1144–D1149. <http://doi.org/10.1093/nar/gkr1013>
- Wei, M., Shinkura, R., Doi, Y., Maruya, M., Fagarasan, S., & Honjo, T. (2011). Mice carrying a knock-in mutation of Aicda resulting in a defect in somatic hypermutation have impaired gut homeostasis and compromised mucosal defense. *Nature Immunology*, 12(3), 264–70. <http://doi.org/10.1038/ni.1991>
- Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D., & Duhr, S. (2010). Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications*, 1(1), 100. <http://doi.org/10.1038/ncomms1093>
- Wolfe, A. D., Li, S., Goedderz, C., & Chen, X. S. (2020). The structure of APOBEC1 and insights into its RNA and DNA substrate selectivity. *NAR Cancer*, 2(4), 36–38. <http://doi.org/10.1093/narcan/zcaa027>
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., ... Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, 2(3), 100141. <http://doi.org/10.1016/J.XINN.2021.100141>
- Yamanaka, S., Poksay, K. S., Arnold, K. S., & Innerarity, T. L. (1997). A novel translational repressor mRNA is edited extensively in livers containing tumors caused by the transgene expression of the apoB mRNA-editing enzyme. *Genes &*

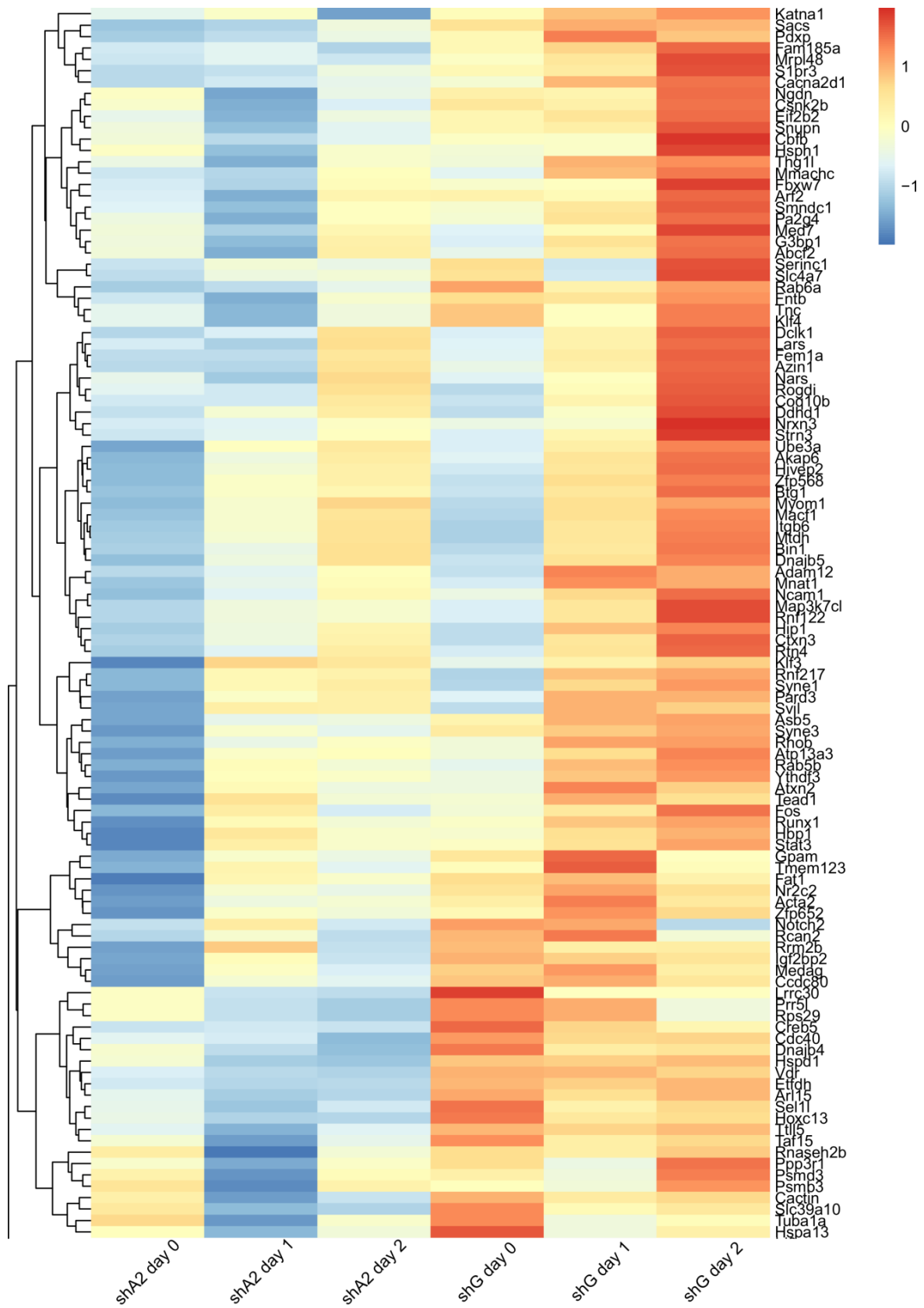
- Development*, 11(3), 321–333. <http://doi.org/10.1101/GAD.11.3.321>
- Yu, G., Wang, L.-G., & He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 31(14), 2382–2383. <http://doi.org/10.1093/bioinformatics/btv145>
- Zeraati, M., Langley, D. B., Schofield, P., Moye, A. L., Rouet, R., Hughes, W. E., ... Christ, D. (2018). I-motif DNA structures are formed in the nuclei of human cells. *Nature Chemistry*, 10(6), 631–637. <http://doi.org/10.1038/s41557-018-0046-3>
- Zhou, P., Yang, X., Wang, X., Hu, B., Zhang, L., Zhang, W., ... Shi, Z. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(January). <http://doi.org/10.1038/s41586-020-2012-7>

6. References

7. Appendix

7. Appendix

7.1 Appendix I



(Continued to the following page)

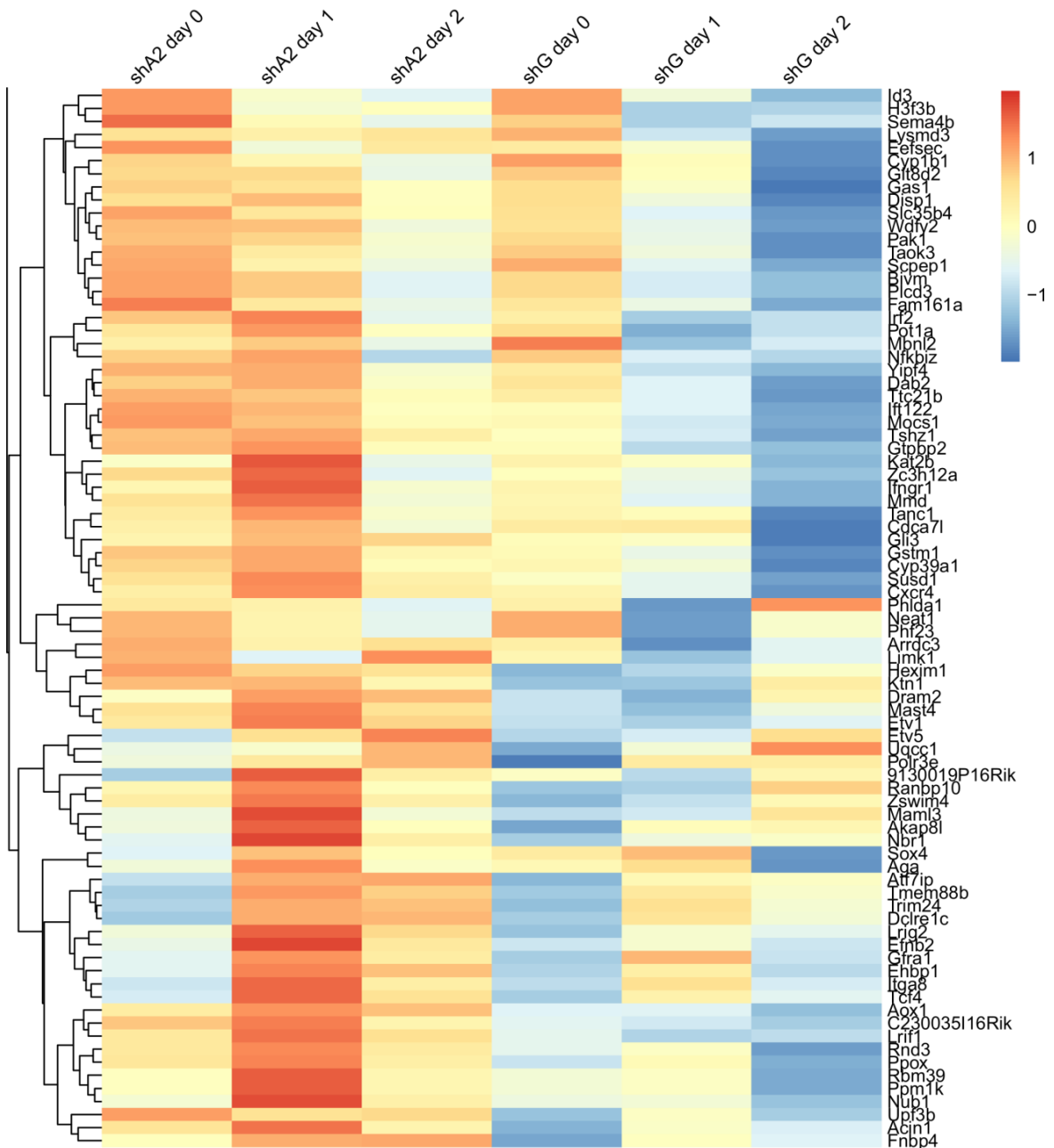


Figure 7.1. Heatmap from the list of differentially expressed APOBEC2 target genes. APOBEC2 target genes were defined from the ChIP-Seq DiffBind analysis (34 h vs 14 h time point). Differentially expressed genes were taken from DESeq2 analysis, shA2 vs shGFP, with $\text{padj} < 0.05$ at any time point, day 0, day 1, or day 2. Gene expression values were taken from the RNA-Seq experiment from APOBEC2 knockdown (shA2) and control shGFP C2C12 myoblasts at corresponding time points of differentiation, day 0 to day 2. Heatmap colors correspond to mean normalized read counts scaled by row (z-score). Heatmap was prepared using the pheatmap package in R (Kolde, 2019).

7. Appendix

7.2 Appendix II

Re: APOBEC2

Papavasiliou, Fotini

Wed 9/2/2020 8:21 PM

To: ALKURAYA, FOWZAN SAMI <FAIKuraya@kfshrc.edu.sa>;
Cc: SATEESH, MADDIREVULA <msateesh@kfshrc.edu.sa>; Lorenzo, Jose <jose.lorenzo@dkfz-heidelberg.de>; Ruf, Sandra <sandra.ruf@dkfz-heidelberg.de>;

Dear Prof. Alkuraya,

We'd be more than happy to collaborate and help you figure out the mechanism behind the mutation you've identified.

Would it be possible to get some more information (perhaps on a zoom call) so that we can best decide how to proceed?


With best regards,

Nina P.

=====
Professor F. Nina Papavasiliou, Ph.D.
Head, Division of Immune Diversity (D150)

German Cancer Research Center (DKFZ)
Foundation under Public Law
Im Neuenheimer Feld 280
69120 Heidelberg
Germany
phone: +49 6221 42-1390

n.papavasiliou@dkfz.de
www.dkfz.de



Management Board: Prof. Dr. med. Michael Baumann, Ursula Weyrich
VAT-ID No.: DE143293537

On Sep 1, 2020, at 4:03 PM, ALKURAYA, FOWZAN SAMI <FAIKuraya@kfshrc.edu.sa> wrote:

Dear Dr. Papavasiliou,

We read with interest your paper in BioRxiv on the role of APOBEC2 in myoblast differentiation. We thought you'd be interested to know that we have a family with a lethal form of myopathy in whom we identified a homozygous APOBEC2:NM_006789.3:c.190G>A;(p.Gly64Arg) variant, which we flagged as a likely candidate but didn't know what to make of it until we read your paper. Perhaps we can interest you in collaborating with us to figure out if this variant does impair APOBEC2 capacity to function in myoblast differentiation?

Very best,
Fowzan

Fowzan S Alkuraya, MD (Hons) ABP ABMGG (Clinical Genetics and Clinical Molecular Genetics)
Professor of Human Genetics, Principal Clinical Scientist and Senior Consultant
Head, Developmental Genetics Unit
King Faisal Specialist Hospital and Research Center
MBC-03 PO BOX 3354
Riyadh 11211, Saudi Arabia
Tel: +966 11 442 7875
Fax: +966 11 442 4585
Twitter: @alkuraya
falkuraya@kfshrc.edu.sa
alkuraya@outlook.com
[Google Scholar](#)

Figure 7.2. Mail from Dr. Fowzan Alkuraya regarding APOBEC2 G64R mutation.

7.3 Appendix III

qPCR: A2 expression in C2C12

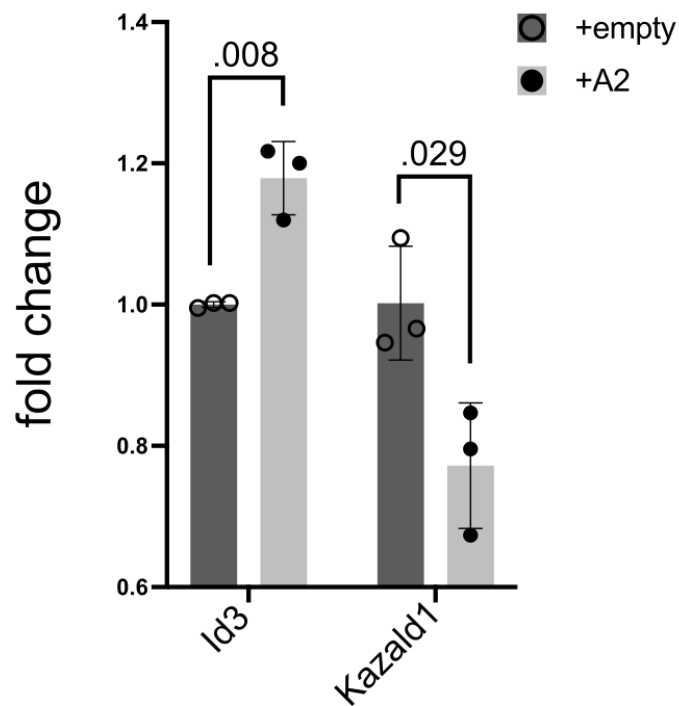



Figure 7.3 Effect of exogenous expression of APOBEC2 in C2C12 fibroblasts on APOBEC2 target genes. A2 expression in RT-qPCR (qPCR) was done on selected target genes from total RNA extracted from C2C12 cells expressing APOBEC2. C2C12 cells were transduced with pMX-A2 (+A2) or pMX-empty (+empty) retroviral vectors. Bars represent mean fold change (normalized expression, $\Delta\Delta Cq$) between +empty and +A2 cells with Tbp as a reference housekeeping gene, calculated in BioRad CFX Maestro software. Error bars represent standard deviation. Bar plots and statistics were prepared using Prism 9 software; statistical analysis: multiple t-tests. Values above each pairwise comparison represent adjusted P values (Holm-Sidak method).

7. Appendix

7.4 Appendix IV





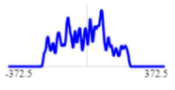
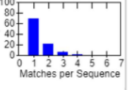



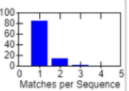



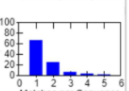






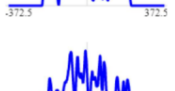
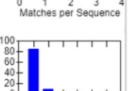


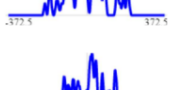
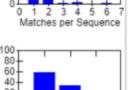


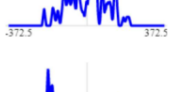
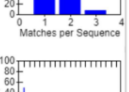


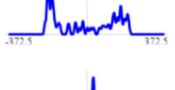
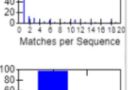


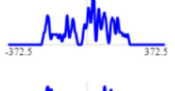
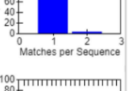



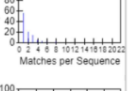



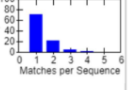



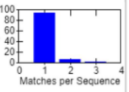


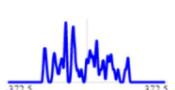



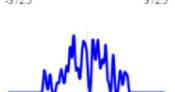
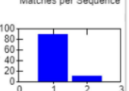
STREME
Sensitive, Thorough, Rapid, Enriched Motif Elicitation

For further information on how to interpret these results please access <https://meme-suite.org/meme/doc/streme.html>.
To get a copy of the MEME software please access <https://meme-suite.org>.

If you use STREME in your research, please cite the following paper:
Timothy L. Bailey, "STREME: accurate and versatile sequence motif discovery", *Bioinformatics*, Mar. 24, 2021. [\[full text\]](#)

[DISCOVERED MOTIFS](#)
[INPUTS & SETTINGS](#)
[PROGRAM INFORMATION](#)
[MOTIFS IN MEME TEXT FORMAT](#)
[MATCHING SEQUENCES](#)
[RESULTS IN XML FORMAT](#)

DISCOVERED MOTIFS [Next Top](#)

Motif	Logo	RC Logo	P-value	E-value	Sites	More	Submit/Download	Positional Distribution	Matches per Sequence
1-GCCCCCCCCCN			1.5e-008	2.2e-007	320 (33.0%)	↓	→		
2-RTGASTCAY			5.4e-005	8.1e-004	221 (22.8%)	↓	→		
3-AGCCAATCAGA			3.9e-004	5.9e-003	286 (29.5%)	↓	→		
4-CCCCCCCCBCCC			1.4e-003	2.1e-002	137 (14.1%)	↓	→		
5-GCGCAGGCCGC			1.6e-003	2.4e-002	72 (7.4%)	↓	→		
6-ACTACAWTCCCA			7.0e-003	1.0e-001	97 (10.0%)	↓	→		
7-AAAAAAAAAAAA			9.2e-003	1.4e-001	100 (10.3%)	↓	→		
8-ACCGAAGTG			1.4e-002	2.2e-001	63 (6.5%)	↓	→		
9-CCTCCTCCTCCTC			1.9e-002	2.9e-001	344 (35.5%)	↓	→		
10-AGCAGAGCTGB			2.1e-002	3.2e-001	231 (23.8%)	↓	→		
11-AGATGGCGGC			3.2e-002	4.8e-001	99 (10.2%)	↓	→		
12-CCCGCCTC			3.2e-002	4.8e-001	65 (6.7%)	↓	→		
13-RTGAGTCAY			5.9e-002	8.9e-001	89 (9.2%)	↓	→		
14-AAACCACAA			8.2e-002	1.2e+000	167 (17.2%)	↓	→		

Stopped because 3 consecutive motifs exceeded the p-value threshold (0.05).
STREME ran for 85.06 seconds.

(Continued to the following page)

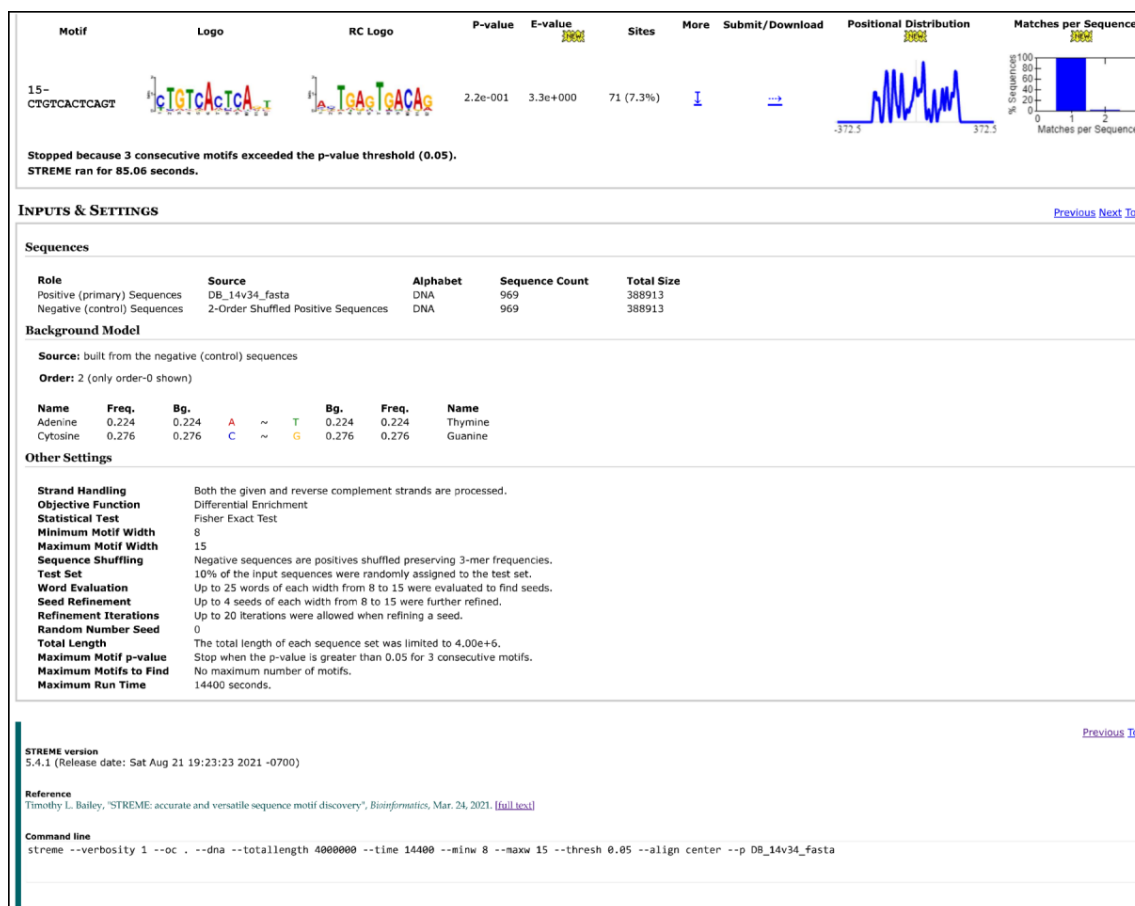


Figure 7.4. STREME analysis for motif discovery in APOBEC2 target chromatin regions. APOBEC2 binding sites, 969 sites with length 401 bp centered at the peak, were defined from the ChIP-Seq analysis in section 3.2.1. The enriched motifs were ranked by the p-values (adjusted p-values). The sites column, column 6, represents the frequency of the motifs within the APOBEC2 binding sites. Sequence logos depict the frequency of each nucleotide base at each position scaled to bits, 0 to 2. The positional distribution represents the frequency (out of 969) and location of the motifs along the 401 bp site centered at the peak. Matches per Sequence represents the number of times a certain motif is present in the same sequence. STREME analysis was done with default settings and used through the MEME Suite webpage (<https://meme-suite.org/meme/doc/streme.html>).

7. Appendix

8. Acknowledgments

First, I would like to acknowledge **Prof. Dr. Nina Papavasiliou** for allowing me to seek out the function of APOBEC2! Her mentorship was invaluable.

I would also like to thank my Thesis Advisory Committee, **Prof. Dr. Frank Lyko** and **Prof. Dr. Christoph Dieterich**, for their guidance.

I would like to thank **Damir Kronic** and **Manuela Brom** of the DKFZ Light Microscopy Facility for the training and guidance with microscopy and image analysis.

I would like to thank **Frank Schwarz** and GPCF@DKFZ for the training and assistance with the MST experiments.

I would like to thank **Julia Flock** and **Dr. Kim Remans** from EMBL PepCore for producing the recombinant APOBEC2 and its mutants in insect cells.

I would like to thank **Ignacio Ibarra** and **Dr. Moritz Mall** for the advice and the discussions regarding transcriptional regulators.

I would like to acknowledge **Linda Molla** who paved the way for this project. Thank you for always pushing and challenging me.

I would like to thank **Sandra Ruf**. We did this together.

I would like to thank **Esteban Erben** for the tips regarding the co-IP experiments.

I would like to thank **Ricca Pecori** for listening to my ideas and problems. I could always count on your wise advice.

I would like to thank **Ana Maia** and **Shravan Venkateswaran** for always being there.

I would like to thank **Alex Hempelmann** for translating my abstract. (I forgot we started together in the lab.)

I would also like to thank **Anastasia Gkeka** and **Monica Chandra** for making life calm like Animal Crossing during the lockdown.

I also appreciate commiserating with **Taga Lerner** and **Ksenia Myacheva**. We started this Ph.D. journey together and we will end it together.

I also must thank **Xico Aresta Branco**. Thank you for pretending to remember me when I asked you regarding openings in Nina's lab.

I would also like to thank **Joey Verdi** for guiding me on Cy5-labeling APOBEC2 for the MST experiments. The sushi nights were also great!

I also must thank **Salvo di Giorgio!** Thank you for the timely help with the alignments!

I would also like to thank **Annette** and **Monique** for making sure materials and equipment are readily available for experiments.

To the past and current members of D150/D160, thank you for bringing life to the lab! Thank you **Gianna, Bea, Rafi, Kathrin, Hamid, Evi, and Dimitra** for the memories!

I would also like to thank **Elias Amro** who joined me on the APOBEC2 project as I was finishing this thesis.

Salamat to the Pinoys of Heidelberg: **Arlou, Fidel, Josh, and Aeiou!**

Lastly, I would like to thank my mom and dad, and the rest of my family, for their support. Thank you for believing in me!

And **Angelica**. Ito na yun.