SynTEF1 Overrides the HP1 Mediated Epigenetic Transcriptional Silencing at the Microsatellite Repeat Expansion in Friedreich's Ataxia

by

Christopher J. Brandon

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

University of Wisconsin-Madison

2022

Date of final oral examination: 06/20/2022

The dissertation is approved by the following members of the Final Oral Committee:

Aseem Z. Ansari, Professor, Biochemistry

Judith Kimble, Professor, Biochemistry

John M. Denu, Professor, Biochemistry

David A. Wassarman, Professor, Genetics

Acknowledgements

I would like to thank Aseem Ansari for his mentorship throughout the years. You have been a source of encouragement and excitement and have helped me to engage this difficult and often counter-intuitive project. It has been a pleasure working with you and I count myself lucky to have been a part of your lab. I also cherished the many discussions I have had with the current lab members (Sandra Kietlinska, Ashraf Mohammed, Mangesh Kaulage, Adithi Danda, Ryan Kempen, Preeti Dabas, Steven Phillips, Sarah Robinson-Thiewes, Devesh Bhimsaria, Suresh Kandikonda, and Nilanjana Chowdhury) and former lab members (Asfa Ali, Mackenzie Spurgat, and Evan Heiderscheit) which have been critical for my development as a scientist. I would like to thank the undergrads I have worked with (Sam Rider and Moua Lee) as well as the Khorana scholars (Mitali Padhi and Aarathi Menon Sankar). I would like to thank Natalie Racine for her mentorship inside and outside of lab. Finally, I would like to thank Professors Judith Kimble, John Denu, and David Wassarman for their critical analysis and insight into my research.

Dedication

The work contained in this thesis could not have been possible without the unwavering support of my wife and closest companion Ruth Brandon. Her confidence, patience, and eagerness to help has been the foundation from which I begin every day and has allowed me to perform the research presented here. Thank you for everything that you have done throughout the years.

Table of Contents

Abstract

Chromatin is generally separated into open, active euchromatin and compressed, silent heterochromatin. These regions contain disparate post-translational modifications on the DNA bound histones which signal for the specific recruitment of regulatory proteins that are needed to activate or repress local transcription. Methylation of the lysine 9 residue of histone H3 (H3K9me3) is an epigenetic signal for chromatin compaction and silencing. The HP1 family of proteins recognize and bind this modification and mediate compaction through homo- and hetero-dimerization. The replacement of methylation at H3K9 with active acetyl modifications engages positive transcriptional machinery such as BRD4, which signals for the activation of RNA polymerase.

Friedreich's Ataxia (FRDA) is a rare, progressive, and lethal neurodegenerative disorder that results from an aberrant trinucleotide repeat expansion in the first intron of the *FXN* gene which results in its heterochromatinization. We have found that the synthetic transcription elongation factor SynTEF1 licenses transcription across *FXN* without the expunction of the H3K9me3 repressive modification or the loss of HP1 proteins. SynTEF1 generates this dualistic state by preferentially recognizing the repeat expansion within the heterochromatic region which creates an ingress for BRD4 accumulation. Both of these transcriptionally opposed proteins form phase separated condensates in regions of active or repressive transcription, respectively. SynTEF1 mediates the co-condensate formation of BRD4 and HP1, likely mirroring other cellular mechanisms that enable RNA synthesis within heterochromatin. The co-treatment of SynTEF1

with the histone deacetylase inhibitor 109 results in an unexpected rise in H3K9me3 across the repeat region concomitant with the highly synergistic activation of *FXN* expression. The active reinforcement of heterochromatic modifications which are progressively overcome by SynTEF1 demonstrates the context-dependence of epigenetic modifications in their role of transcriptional regulation. The observations discussed in this thesis contribute to a deeper understanding of the role of epigenetic gene regulation at the interface of transcriptional activation and repression.

Chapter 1 | The epigenetic mechanisms of transcriptional repression at *Frataxin* in Friedreich's ataxia

1.1 Introduction

When the 6.4 billion nucleotides of the human genome were fully sequenced it was determined that there exist only about 20,000 protein coding genes (although this number is still debated), which make up approximately 1.5% of the genome [1]. In contrast, gene-poor and transcriptionally repressed repetitive DNA regions account for at least 45-54% of the human genome [2, 3]. These repetitive sequences are necessary to maintain the structure and function of the human chromosomal landscape and have helped direct human evolution [4, 5]. This repetitive DNA is largely made up of transposable, or mobile elements that can rearrange, insert, and duplicate themselves within the genome. Transposable elements may be considered selfish, or in some cases are even referred to as genomic parasites, in that their mobility allows them to spread pervasively, separately from the genetic elements of the organism [6]. This selfishness explains why they are found ubiquitously in all eukaryotic genomes and why they constitute such a large portion of the human genome. Transposable elements are not just benign expansions but can be found in gene bodies and important regulatory regions. Their insertion nearby or within genes affect local expression, can result in mutations that change the protein function, and change long range DNA interactions. Over many generations these changes caused by the insertion of transposable elements is stabilized in the genome which drives evolution and contributes to normal gene function, but aberrant insertions can result in the misregulation of gene transcription. The change in the modulation of essential genes through the insertion of transposable elements leads to several types of human diseases [5, 7, 8].

There are two major classes of transposable elements that differ by length, mechanism of insertion, and method of transcription. Long Interspersed Nuclear Elements (LINEs) are 4-7 kb

sequences transcribed by RNA polymerase II and comprise over 20% of the human genome [2]. They contain two open reading frames and a regulatory CpG island in the promoter [8]. The proteins translated from the two open reading frames are used to reverse-transcribe and insert the element back into the genome at a different locus [9]. The most common and only active family of LINE elements is LINE-1, while the less common LINE-2 and LINE-3 elements have become inert due to genomic mutations [9].

Small Interspersed Nuclear Elements (SINEs) are much smaller, about 300 nucleotides, and comprise 13% of the repetitive DNA sequences in the human genome [2]. The most abundant family of SINE elements is the Alu repeat, a descendent of the 7SL RNA gene [10]. Alu repeats contain a CpG regulatory element, an internal polymerase III binding sequence, and consist of two structured domains separated by a poly-A sequence [10]. The CpG regions found in transposable elements comprise over 1/3 of the total CpG islands in the human genome, and their methylation can be used to estimate global DNA methylation [9]. Transposable elements are highly prevalent sequences that are variable in size, contain internal regulatory elements, and are located irregularly in the genome, all of which leads to their involvement in disease etiology. They contribute to Autism Spectrum Disorder, Schizophrenia, bipolar disorder, Alzheimer's disease, and Friedreich's Ataxia amongst others [8, 11-13].

Another common repetitive element found in eukaryotes are Short Tandem Repeats (STRs), which are short DNA sequences of up to 100 bp that contain 1-13 repeating nucleotides. Despite their small nature, STRs make up about 3% of the human genome [2]. The repeating sequence of STRs primarily consist of short poly-A stretches and are frequently found near or inside SINE transposable elements [14-16]. The correlation between STRs and SINE elements

suggests that STRs result primarily from the mutation of poly-A stretches of mobile SINE elements [15]. STRs are prone to DNA polymerase slippage during replication which leads to both their expansion as well as the variability in their size between individuals [17]. Multiple repeat length differences can be used to identify people by forensic analysis based on collected DNA samples, genome mapping, parental verification, population genetics, and other various applications [18, 19]. Despite the variations in repeat length, their localization in the genome is not random. Trimer and hexamer repeats enrich in exonic regions which result in amino-acid insertions, while other more translationally disruptive repeat lengths are primarily found in intronic and intergenic regions [14]. STR expansions can change the transcriptional regulation or coding sequence of necessary genes, generating either loss of function or gain of function mutations. Over the last 15 years disorders associated with STR repeat expansion have doubled, and now account for more than 40 repeat expansion diseases (Fig 1.1) [20, 21].

One disease, Friedreich's Ataxia (FRDA), results from an STR expansion at an internal SINE element [22]. It is a rare autosomal recessive neurodegenerative disorder that affects 2-3 people per 100,000 [23]. FRDA was first described by Dr. Friedreich in 1863 as a disease resulting in the loss of feeling in the lower limbs and scoliosis of the spine [24]. The first symptoms begin during childhood, generally between 10 and 15 years of age [23]. As the disease progresses, patients exhibit muscle weakness which frequently results in the loss of motor control in the peripheral limbs and scoliosis, leaving the individual wheelchair bound. As the loss of muscle control worsens, patients can have difficulty controlling facial movements including swallowing and exhibit increased tiredness. About 30% of patients also develop diabetes mellitus as the disease progresses [23]. In the later stages FRDA patients frequently develop different cardiovascular

diseases such as hypertrophic cardiomyopathy, myocardial fibrosis, and tachycardia which eventually results in a shortened lifespan [25].

For over one-hundred years doctors have documented the disease progression of FRDA patients, with only palliative care available to ease their suffering. It was not until 1995 that the Pandolfo lab identified a 150 kb region of chromosome 9 that was associated with the disease [26]. The next year they identified the gene responsible for FRDA and coined it Frataxin (*FXN*), after Dr. Friedreich and the ataxia he first described [27]. Pandolfo identified a large $(GAA)_n$ STR repeat expansion in the first intron of the Frataxin gene in patient tissue, which was found to be present on both *FXN* alleles. Healthy individuals carry <10 (GAA) STR repeats found near an inactive, ~55M year old SINE transposable element in the Alu family [7]. This (GAA) STR is much larger in FRDA affected individuals and consists of 50 to over 1000 consecutive trinucleotide repeats. Patients homozygous for the (GAA) expansion suffer from FRDA, but heterozygous individuals are healthy and show no adverse phenotypes. In rare circumstances, affected FRDA individuals can be heterozygous for the GAA repeat expansion and contain a deleterious mutation in the second allele that leads in its inactivation, which results in the depletion of functional *FXN* protein and the progression of the disease [28].

In FRDA, the intronic (GAA) expansion results in decreased levels of *FXN* transcription, which cumulates in a global cellular depletion of both transcript and protein [29]. FXN is a necessary protein for growth and development, and the complete loss of FXN is embryonically lethal [30]. The GAA expansion attenuates expression but does not completely silence the gene; transcriptional levels are instead reduced to <5-25% of healthy individuals [31]. The repression

of *FXN* transcription occurs in a GAA repeat length-dependent manner, with individuals who contain longer repeats showing a deeper level of suppression [31].

The loss of *FXN* transcript results in a depletion of FXN protein, which is required for iron trafficking and mitochondrial iron-sulfur cluster biogenesis [32, 33]. The misregulation of iron trafficking results in a depletion of iron-sulfur clusters, placing a strain on the electron transport chain as well as disrupting general mitochondrial iron homeostasis. FRDA patient cells show signs of increased oxidative stress, increased iron uptake, mitochondrial structural and replication defects, and a reduction of the electron transport chain ATP generation [32-35].

Since the GAA repeat expansion mediates the loss of functional *FXN* transcript resulting in FRDA disease progression, understanding and alleviating that transcriptional block has been the primary focus of research in treating FRDA patients. The mature *FXN* transcript is exceptionally stable, with a $T_{1/2}$ of 11 hours [36]. Therefore, even partially relieving the transcriptional block could slow or even reverse disease progression. Unfortunately, the (GAA) repeat expansion results in multiple types of transcriptional inhibition which makes it difficult to treat. RNA Polymerase II is observed at the transcription start site (TSS) but elongation through the repeats appears to be stalled, and to this day it is still unclear whether one or multiple mechanisms of repression are inhibiting polymerase progression [36, 37]. There is strong evidence for the loss of active epigenetic modifications (Fig 1.2A) and the placement of inhibitory marks (Fig. 1.2B) which shift the DNA from an euchromatic to heterochromatic state (further discussed in 1.2). RNA:DNA duplex formation and antisense transcription mediating silencing have also been observed, presenting additional physical barriers to transcription (Fig 1.2B and further discussed in 1.3) [38-40].

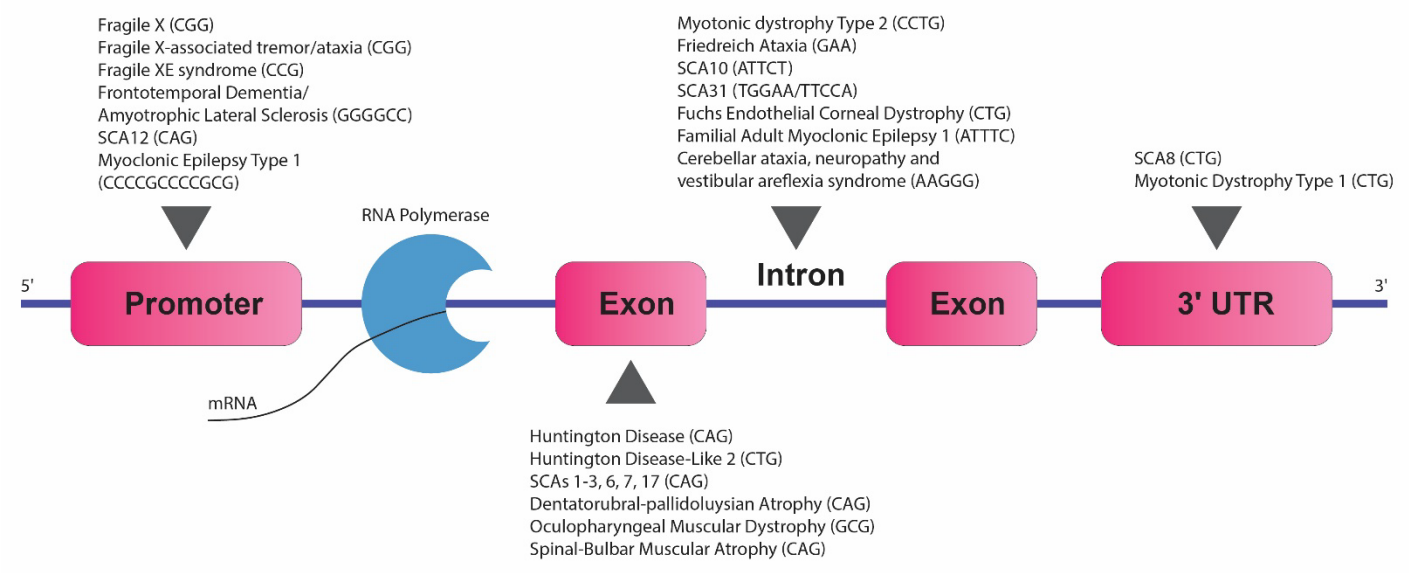

Fragile X (CGG)
Fragile X-associated tremor/ataxia (CGG)
Fragile XE syndrome (CCG)
Frontotemporal Dementia/
Amyotrophic Lateral Sclerosis (GGGGCC)
SCA12 (CAG)
Myoclonic Epilepsy Type 1
(CCCCGCCCCGCG)
RNA Polymerase
Myotonic dystrophy Type 2 (CCTG)
Friedreich Ataxia (GAA)
SCA10 (ATTCT)
SCA31 (TGGAA/TTCCA)
Fuchs Endothelial Corneal Dystrophy (CTG)
Familial Adult Myoclonic Epilepsy 1 (ATTTC)
Cerebellar ataxia, neuropathy and
vestibular areflexia syndrome (AAGGG)
SCA8 (CTG)
Myotonic Dystrophy Type 1 (CTG)
5'
Promoter
Exon
Intron
Exon
3' UTR
3'
mRNA
Huntington Disease (CAG)
Huntington Disease-Like 2 (CTG)
SCAs 1-3, 6, 7, 17 (CAG)
Dentatorubral-pallidoluysian Atrophy (CAG)
Oculopharyngeal Muscular Dystrophy (GCG)
Spinal-Bulbar Muscular Atrophy (CAG)

Figure 1.1 | Diseases resulting from repeat expansions

Diseases resulting from short-tandem repeat expansions. Repeat-expansion diseases result from the improper insertion or expansion of a repetitive DNA sequence into the gene. Shown are common repeat expansion–based diseases with the expanded DNA sequence and the location of the expansion. Repeats that occur in the promoter region either positively or negatively regulate gene transcription. Those that occur in the exon result in translated amino acid insertions, and those found in introns affect transcription and exon splicing. Although less common, 3’-UTR expansions can cause translational defects. Abbreviations: SCA, spinocerebellar ataxia; UTR, untranslated region.

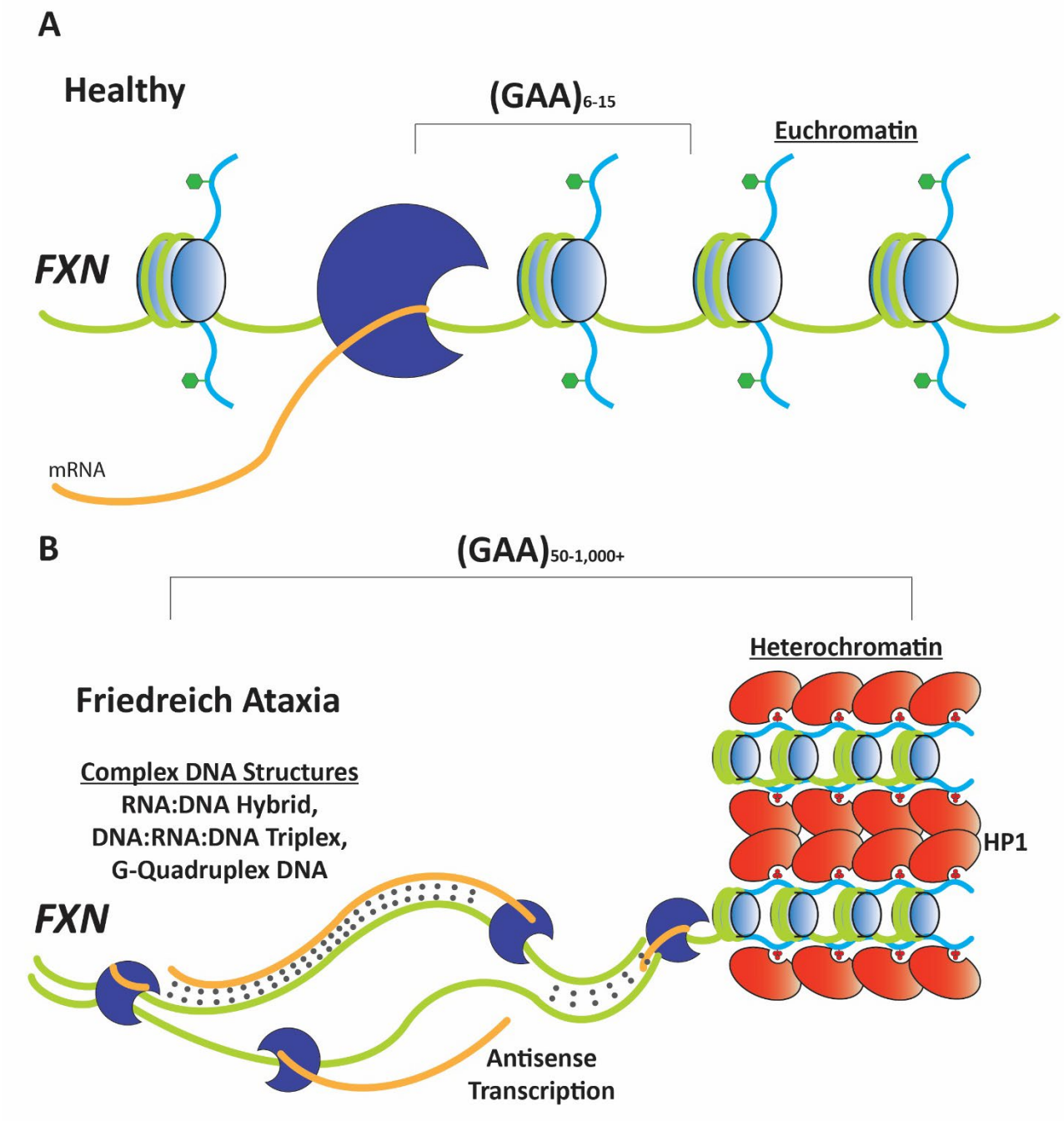

A
Healthy
(GAA)6-15
Euchromatin
FXN
mRNA
B
(GAA)50-1,000+
Heterochromatin
Friedreich Ataxia
Complex DNA Structures
RNA:DNA Hybrid,
DNA:RNA:DNA Triplex,
G-Quadruplex DNA
HP1
FXN
Antisense
Transcription

Figure 1.2 | Many mechanisms of repression at FXN caused by the (GAA) repeat expansion

(a) The transcription of *FXN* proceeds normally in healthy individuals. Histones are acetylated and open, and polymerase (blue crescent) transcribes through the repeat region. **(b)** The expanded repeat region in Friedreich ataxia results in the formation of complex DNA structures, antisense transcription, and heterochromatin, all of which negatively regulate progressive transcription of the gene. Acetyl groups are green hexagons; methyl groups are red circles; nucleosomes are light blue ovals; mRNA is yellow; and DNA is green. Polymerase (dark blue crescents) is transcribing the *FXN* gene. Heterochromatin is bound by HP1 (red ovals).

1.2 Epigenetic modifications and the histone code

Nucleosomes are the most basic subunit of chromatin, consisting of 147 bp of DNA wrapped around a core histone octamer. The octamer is composed of two copies each of the evolutionarily conserved histone H2A, H2B, H3, and H4 [41]. The DNA is wrapped in a left-handed manner about 1.8 turns around the nucleosome core, with Lysine and Arginine positively charged histone amino acids contacting and directing the DNA winding [41, 42]. Only ~75% of each histone protein is folded and contained within the DNA bound nucleosome core. The other 25% consists of an unstructured N-terminal tail that extends out past the core particle. The solvent exposed histone tail is available for Post Translational Modifications (PTMs) and other protein-protein interactions [41]. This includes inter-nucleosomal interactions which stabilize the chromatin and mediate secondary and tertiary nucleosomal structures [41, 42].

Histone tails play a vital role in the regulation of transcription. They function as signposts, signaling to other factors about the nature and function of the wrapped DNA through their post-translational modification at specific residues. With 8 histone proteins per nucleosome, there exist many possibilities for both variation and redundancy at a locus, which allows for different signals and signal strengths at various locations along each gene. Histone tail PTMs are involved in regulating polymerase progression during transcription, the recruitment of transcription factors, regulating DNA availability, winding, structure, and compaction [43]. Some of the more common PTMs include the following: The phosphorylation of serine, threonine, and tyrosine results in the placement of a large, negatively charged modification. Positively charged lysine residues may have their charge stabilized through methylation, neutralized by acetylation or crotonylation, or provide a ligation site for the addition of the ubiquitin protein. Arginines can be

methylated which increases their bulkiness and stabilizes their positive charge, or citrullinated which eliminates it. Prolines can undergo isomerization which changes the structure of the tail, changing the availability of binding sites. And there are even more, less common modifications [43, 44]. The staggering complexity of potential epigenetic PTMs provides the cell a way to barcode regions of DNA for active transcription or to signal for its compaction and transcriptional silencing. The idea that multiple modifications co-occur and function combinatorially and/or sequentially in order to regulate downstream processes is called the histone code hypothesis [45].

Histone modifications work to regulate transcription and, while their function is frequently conserved between species and directly heritable, the signal is not directly encoded into our genome. Therefore, these modifications along with DNA methylation and regulatory RNA molecules are termed epigenetic modifications [46, 47]. Epigenetic modifications modulate most aspects of mRNA transcription: they recruit the proteins necessary for DNA compaction and silencing, signal remodeling complexes to reposition and rearrange the nucleosomes, and provide binding sites for necessary transcription factors [44]. The correct epigenetic signaling is necessary for homeostatic survival, and when it is disrupted in the cell it can lead to apoptosis, phenotypic changes, or abnormal growth culminating in various disease states including many types of cancer [48]. The (GAA) repeat expansion in FRDA results in many epigenetic changes at *FXN*, which have been characterized extensively (Fig. 1.3A).

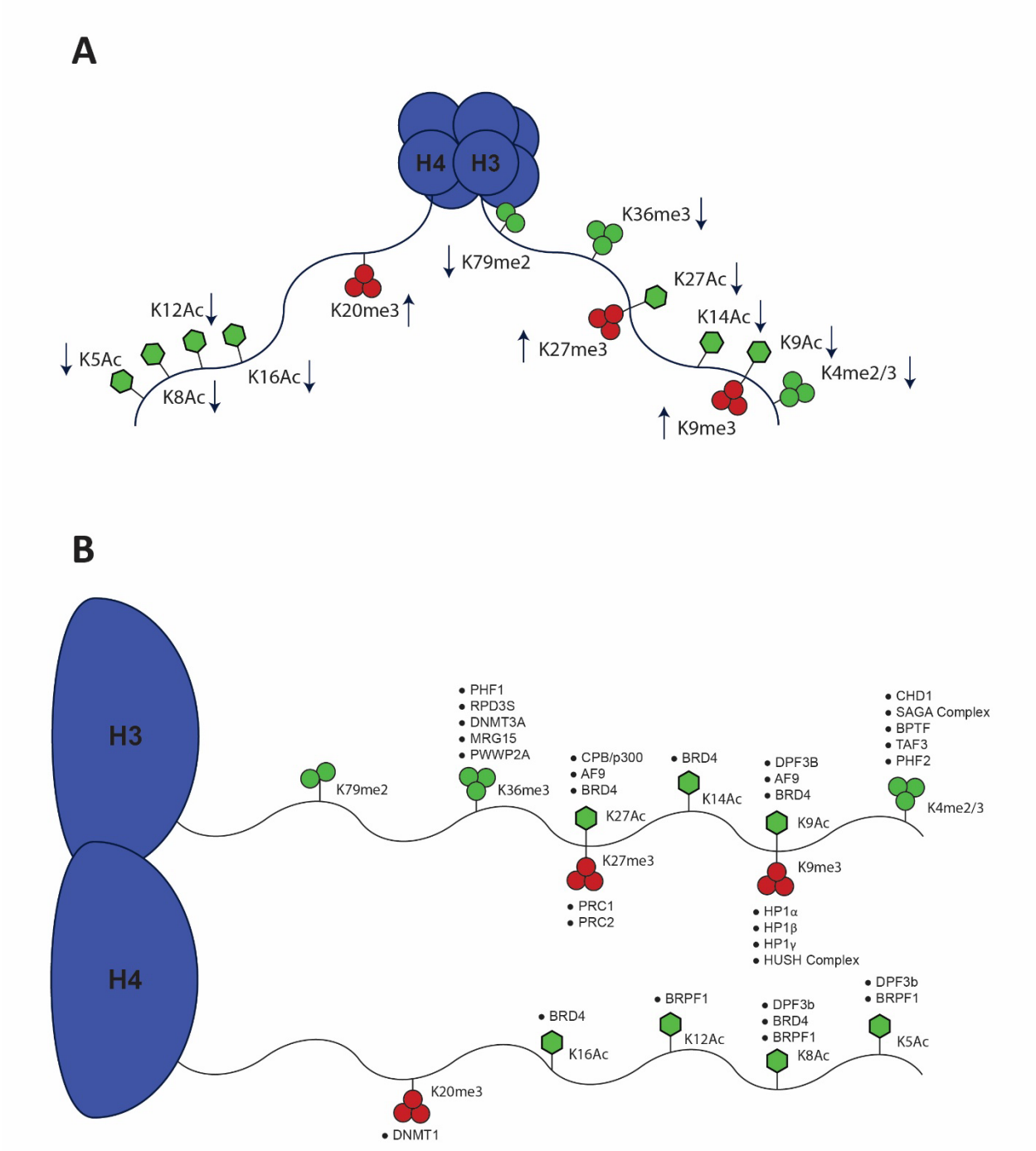

A
H4
H3
K36me3
K79me2
K27Ac
K12Ac
K20me3
K14Ac
K5Ac
K27me3
K9Ac
K16Ac
K4me2/3
K8Ac
K9me3
B
H3
H4
• PHF1
• RPD3S
• DNMT3A
• MRG15
• PWWP2A
• CPB/p300
• AF9
• BRD4
• BRD4
• DPF3B
• AF9
• BRD4
• CHD1
• SAGA Complex
• BPTF
• TAF3
• PHF2
K79me2
K36me3
K27Ac
K14Ac
K9Ac
K4me2/3
K27me3
K9me3
• PRC1
• PRC2
• HP1α
• HP1β
• HP1γ
• HUSH Complex
• BRD4
• BRPF1
• DPF3b
• BRD4
• BRPF1
• DPF3b
• BRPF1
K16Ac
K12Ac
K8Ac
K5Ac
K20me3
• DNMT1

Figure 1.3 | The changes in common epigenetic modifications that change during FRDA and their associated reader proteins

(a) Common epigenetic modifications that occur on histones H3 and H4. Red indicates modifications commonly associated with transcriptional repression, and green indicates groups commonly associated with transcriptional activation; methyl groups (me; circles) and acetyl groups (Ac; hexagons) on lysine (K) residues are shown. Arrows indicate the changes observed in FRDA cells, relative to healthy cells. (b) Reader proteins bind epigenetic modifications on specific lysine residues to regulate transcription.

1.3 The silencing of *FXN* through epigenetic transcriptional regulation

The silencing of *FXN* through histone epigenetic changes occurs in two parts: the removal of transcriptionally "active" modifications and their subsequent replacement with transcriptionally "repressive" ones. This transition from an active to repressive state is not entire, which therefore results in the attenuation but not the complete silencing of *FXN* transcription. The histone modifications that change with the expansion of the (GAA) repeat sequence have been characterized extensively (Fig 1.3B). While an exhaustive description of every epigenetic modification and associated protein is outside the scope of this work, I will attempt to summarize the current state of the field on the epigenetic regulation at *FXN*, provide a brief summary of important "reader" proteins, and describe how these changes affect chromatin compaction and transcription elongation in FRDA (Table 1). I will then describe research centered on the development of novel therapeutics that are being designed to overcome this repression.

Epigenetic Modification	Reader Proteins	Writer Proteins	Change During Friedreich's Ataxia
H3K4me2	• CHD1 [180] • SGF29 (SAGA Complex) [97]	• SET1 A/B [181] • MLL 1-4 [181]	Promoter - Unchanged [36, 37] Repeat Region - Decreased [36, 37, 108]
H3K4me3	• BPTF (NURF Complex) [98,99] • TAF3 [94] • PHF2 [182] • CHD1 [180] • SGF29 (SAGA Complex) (97)	• SET1 A/B [183] • MLL 1-4 [183]	Decreased [37, 108]
H3K9Ac	• BRD4 [142,184] • DPF3B [147] • AF9 [136]	• CBP/p300 [80, 185] • PCAF [80] • TAF1 [185] • BRD4 [185] • GCN5 [106]	Decreased [72, 74, 108]
H3K14Ac	• BRD4 [184]	• CBP/p300 [185] • TAF1 [185] • GCN5 [105]	Decreased [38, 72, 74, 108]
H3K27Ac	• BRD4 [142] • AF9 [136]	• CBP/p300 [80] • TAF1 [185] • BRD4 [185]	Present in Healthy Cells Not Measured in FRDA
H3K36me3	• PHF1 (PRC2) [125] • DNMT3A [186, 187] • RPD3s [123] • MRG15 [188, 189] • PWWP2A [124]	• SETD2 [190, 191] • SETD5 [192]	Decreased [36, 37, 108]
H3K79Ac		• DOT1L [129]	Decreased [37]
H4K5Ac	• DPF3b [147] • BRPF1 [193]	• HAT1 [194] • HBO1 [195]	Decreased [38, 72-74, 108]
H4K8Ac	• BRD4 [147] • DPF3b [147] • BRPF1 [193]		Decreased [38, 72-74]
H4K12Ac	• BRPF1 [193]	• HBO1 [195]	Decreased [38, 74]
H4K16Ac	• BRD4 [196]	• MSL Complex [197]	Decreased [38, 72, 74]
H3K9me2	• HP1α/HP1β/HP1γ [56, 57]	• G9a/GLP [198-200]	Increased [72, 74, 89]
H3K9me3	• HP1α/HP1β/HP1γ [56, 57] • MPP8 (HUSH Complex) [201]	• SETDB1 [68] • SETDB2 [69, 70] • SUV39H1 [198, 199] • SUV39H2 [198, 199]	Increased [36-38, 40, 72-76]
H3K27me3	• PRC1 [202] • PRC2 [202]	• PRC2 [79]	Increased [40, 72, 76]
H4K20me3	• DNMT1 [87]	• SMYD5 [203] • SUV420H1/2 [86]	Increased [37]

Table 1 | List of epigenetic modifications that change in response to (GAA) repeat expansion and their associated proteins

Common epigenetic modifications that have been measured in FRDA along with their associated reader and writer proteins. Changes in most reader and writer protein association with *FXN* have not been directly measured. A more complete list allows for an analysis of reader proteins that may be important in regulating epigenetic changes and potential writer proteins that can be targeted to help restore homeostasis.

1.3.1 Heterochromatin formation and histone 3 methylation at K9

Functionally, cellular chromatin can be broadly partitioned into two separate categories. Euchromatin is gene rich, actively transcribed chromatin that is mobile, diffuse, and localized generally in the interior of the nucleus. Alternatively, heterochromatin is composed of tightly bundled regions of DNA that tend to be gene poor, lowly transcribed, and isolated to the nuclear periphery and puncta within the nucleus. Heterochromatin is further subdivided into two important classes depending on its function: facultative and constitutive. Facultative heterochromatin has regions that are transcribed infrequently, usually during development or differentiation, but are otherwise silent. HOX gene cluster silencing and X chromosome inactivation are examples of facultative heterochromatin which are differentially expressed during development [44]. Constitutive heterochromatin is instead tightly wound, compacted DNA that is generally not transcribed.

The establishment and maintenance of constitutive heterochromatin is a necessary function, and we are constantly learning more about the locations and types of DNA that undergo this type of compaction. Restricting access to this DNA and separating it from the rest of the nuclear milieu allows the cell to inhibit the transcription of gene poor regions such as centromeric, telomeric, satellite DNA, and LINE, SINE, and other transposable elements [3, 49-51].

A characteristic of constitutive heterochromatin is di- and tri-methylation of histone H3 at lysine 9 (H3K9me2/3) [52, 53]. H3K9me2/3 colocalizes with constitutive heterochromatin in repetitive regions of DNA including alpha satellite centromeric repeats, retrotransposons, LINE and SINE elements, all gene poor lowly transcribed regions of DNA [51]. H3K9 methylation is a

crucial signal for maintaining the inactive state in these regions through the recruitment of the heterochromatin protein 1 (HP1) family of reader proteins, which both restrict access to DNA and function as physical barriers to transcription [54, 55].

The three human HP1 paralogues, HP1α, HP1β, and HP1γ have a highly conserved sequence with an N-terminal chromodomain, a flexible linker domain, and a C-terminal chromoshadow domain (Figure 1.4A) [56, 57]. Chromodomains are common among proteins that recognize and bind methylated lysine residues. The HP1 chromodomain specifically recognizes and binds to H3K9me2/3 modifications which recruits HP1 to regions of marked heterochromatin inhibiting transcription and facilitating condensation. Once bound to H3K9me2/3, HP1 proteins recognize and interact with each other through their highly conserved chromoshadow domain (Figure 1.4B). This creates complex networks of homo- and heterodimerized HP1 proteins spanning large regions of H3K9me2/3 modified nucleosomes [58, 59]. HP1α oligomerization is regulated through a small, serine rich N-terminal extension. The phosphorylation of these specific serine residues promotes oligomerization, which may prove to be an important yet not well understood biological process [60]. The HP1 family also contains an unstructured linker domain which functions to recognize and bind nucleotides. The additional interactions with RNA and DNA aids in transcription regulation by stabilizing the bound protein at the DNA (Figure 1.4B) [61]. All three of the HP1 paralogues are found in constitutive heterochromatic DNA, facilitating its compaction, and creating a physical barrier to transcription throughout the region. One paralogue, HP1γ, is more promiscuous and can also be found in regions of euchromatin and active transcription, which suggests this protein may play multiple roles depending on the state of the chromatin [62].

An interesting compartmentalization mechanism that has recently been observed with the HP1 family is their ability to phase-separate [60]. The N-terminal, hinge, and C-terminal regions are all highly disordered, a common trait of proteins that have a propensity to form phase-separated condensates [63, 64]. Despite their sequence homology and common ability to bind nucleic acids, the Narlikar lab has shown that the HP1 paralogues have different propensities to phase separate. HP1α readily forms phase separated condensates when complexed with short (147 bp) and long (2.7 kb) DNA sequences, where HP1γ will only coacervate with the longer sequences, and HP1β doesn't phase separate with free DNA at all [65]. They were able to describe another important characteristic for HP1β (and to a lesser extent HP1γ), that their addition to DNA-bound HP1α condensates rapidly dissolves the condensate. This type of relationship provides insight as to why we observe paralog specific binding patterns and allows for subtle, biophysical tuning of epigenetic repression at regions containing H3K9 methylation.

The placement of H3K9 methylation can occur either in a complete or in a stepwise manner. The heteromeric methyltransferases G9a/GLP catalyze the initial placement of mono- and di-methyl modifications [66]. These sites can be further tri-methylated through the recruitment of methyltransferases Suv39H1 and Suv39H2 [52, 67]. Bypassing the stepwise route, SETDB1 and SETDB2 can directly di- and tri-methylate H3K9 without the need of initial priming [68-70]. SETDB1 and SETDB2 can form a multimeric complex with the HP1 family which facilitates the spread of H3K9 methylation as well as the recruitment of HP1, resulting in the rapid compaction of DNA [71].

In healthy individuals the levels of HP1 and H3K9me2/3 would be expected to be low at the *FXN* gene as it is a necessary mitochondrial protein [36]. FRDA patient derived cells that are

challenged with the (GAA) repeat expansion see an aberrant placement of H3K9me2/3 in the first intron and through the repetitive region [36-38, 40, 72-76]. The placement of H3K9me2/3 is not sufficient for transcriptional repression though, HP1 proteins must be recruited to condense the DNA and act as a barrier to transcription. As anticipated, HP1β condenses (GAA) repeat-containing chromatin and FRDA patients display an increase in HP1α and HP1γ recruitment to and around the repeat region of *FXN* [8, 40]. These observations show that H3K9me2/3 is functional and working to suppress transcription at the GAA repeat region. Importantly though, HP1 recruitment is not sufficient to completely silence transcription. There is enough remaining active transcriptional machinery to partially overcome the presence of HP1 and H3K9me2/3, which gives rise to low levels of transcription.

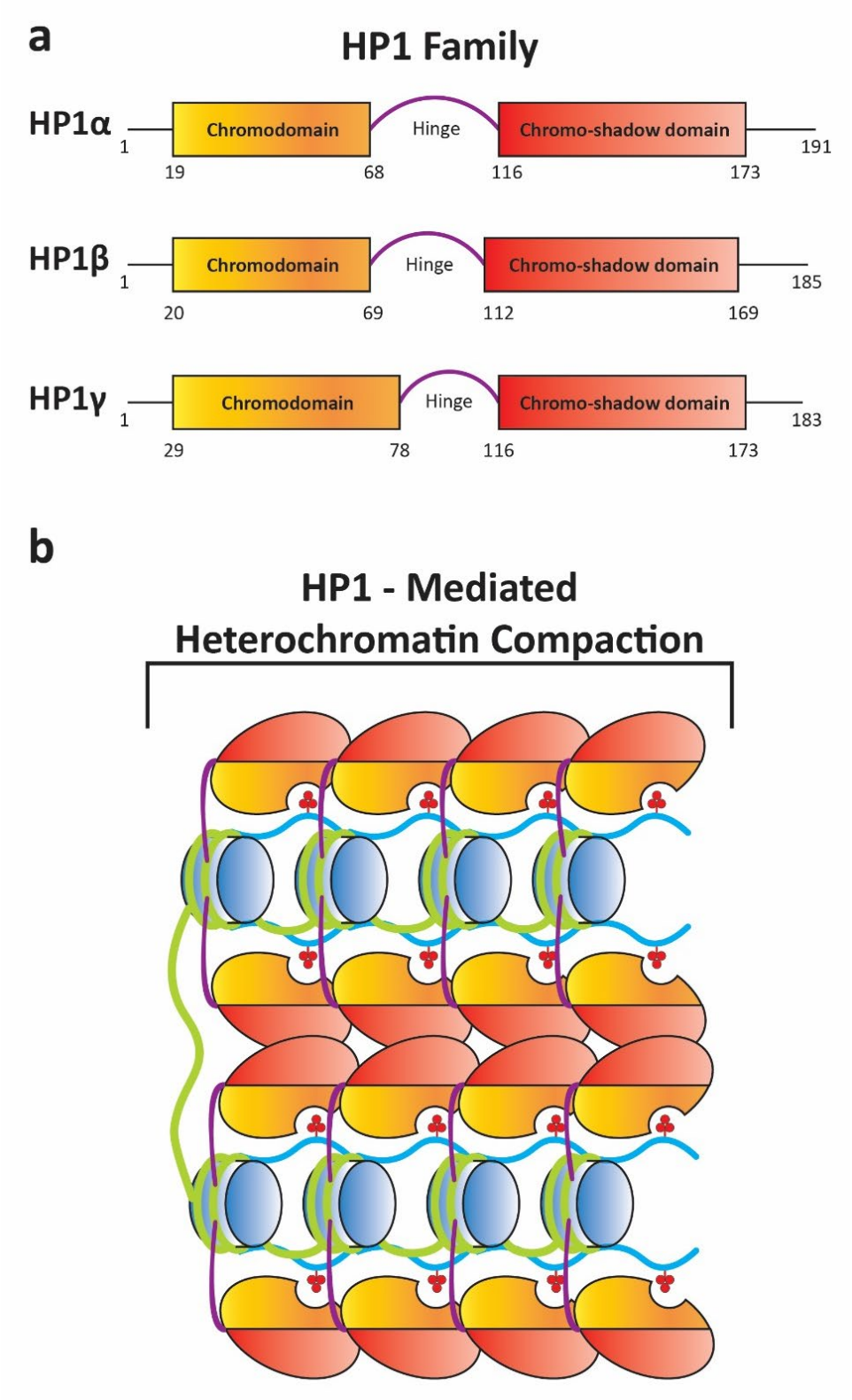

a
HP1 Family
HP1α
1
Chromodomain
Hinge
Chromo-shadow domain
191
19
68
116
173
HP1β
1
Chromodomain
Hinge
Chromo-shadow domain
185
20
69
112
169
HP1γ
1
Chromodomain
Hinge
Chromo-shadow domain
183
29
78
116
173
b
HP1 - Mediated
Heterochromatin Compaction

Figure 1.4 | The structure and function of the heterochromatic HP1 family of epigenetic reader proteins

(a) The domains of the 3 highly complementary human HP1 paralogues (HP1α, HP1β, and HP1γ) are shown with their domain boundaries indicated below each structure. The chromodomain recognizes and binds histones modified with H3K9me2/3, which stabilizes the histones at these loci. The chromo-shadow domain promotes homo- and heterodimerization, which facilitates chromatin compaction. The hinge region binds nucleic acids to help stabilize the proteins bound to chromatin. (b) HP1 mediates the compaction of heterochromatin by binding H3K9me2/3 (red circles) with the chromodomain (orange) and homo- and heterodimerizes through interactions with the chromo-shadow domain (red). The hinge region (purple) interacts with DNA to stabilize its binding. Histones are blue and DNA is green.

1.3.2 H3K27me3 / H3K27Ac and PRC2 transcriptional repression

Histone H3 tri-methylation at lysine 27 (H3K27me3) is another prevalent modification for epigenetic transcriptional repression. It is placed and read by the Polycomb Repressive Complex 2 (PRC2), which contains proteins that can both read H3K27me3 as well as write new modifications onto nearby lysine residues. PRC2 is a large multimeric complex that consists of 4 core proteins: the methyltransferases EZH1/2, the RNA and DNA binding protein SUZ12, the H3K27me3 recognition protein EED, and RBBP4/7 which binds another epigenetic modification H3K36me3 to be described later [77]. PRC2 along with the long non-coding RNA XIST, are necessary for X-chromosome inactivation in females, an evolutionarily conserved process by which one of the X chromosomes is condensed and inactivated during development. This reduces the transcription of genes located on the second female X-chromosome and ensures there is not overproduction of unnecessary mRNA [78]. The method by which the PRC2 complex results in gene repression begins with the EED protein recognizing and binding regions rich in H3K27me3 marks. EZH1/2 then recognizes unmodified H3K27 and catalyzes the transfer of methyl groups from the cofactor S-adenosyl-L-methionine (SAM) to the lysine on the nearby histones through its catalytic SET domain. This process of recognition and subsequent placement of H3K27me3 spreads the epigenetic mark to surrounding histones and creates a feedback loop with the additional recruitment of PRC2 complexes, which both stabilizes the inhibition and triggers its spread [79]. The promoters and gene bodies of transcriptionally silent genes commonly enrich for H3K27me3.

Conversely, acetylation of H3K27 (H3K27Ac) is catalyzed by the ubiquitous histone acetyltransferases CBP/P300 and TAF1 [80, 81]. When acetyl marks are placed they must be

removed before the gene can be methylated, which inhibits both and the recruitment of the PRC2 complex and the spread of additional methyl modifications. H3K27Ac is broadly found in actively transcribed genes and colocalizes with other similar active epigenetic modifications, as well as Mediator, a large multicomponent complex that interfaces directly with polymerase. Mediator recruits a large host of proteins which signal for the polymerase to begin transcription. H3K27Ac, along with H3K4me1 has been used as a standard epigenetic modification to identify enhancers and super-enhancers, regions of DNA that are open, hyperacetylated, and act as hubs for long range control of active transcription [82, 83].

In healthy individuals H3K27Ac is present in the first intron of *FXN* and through the repeat region [84]. This is not surprising due to its role in Mediator recruitment and signaling for active transcription, but unfortunately it is not known whether the (GAA) repeat expansion in FRDA patients affects this. Since H3K27Ac is found to be enriched in the promoters and enhancers of highly expressed genes, I hypothesize FRDA patients would display a significant decrease of H3K27Ac at the *FXN* promoter which would result in the failure to recruit the necessary transcription machinery to the locus, limiting transcription.

While changes in acetylation have not been examined, it has been shown that FRDA leads to a large increase of H3K27me3 in patient derived cells [40, 72, 76]. This change in epigenetic regulation generally leads to the recruitment of PRC2 and other binding partners, thus inhibiting transcription. But while examples of HP1 recruitment to H3K9me3 at *FXN* have been demonstrated, at present the recruitment of PRC2 to the H3K27me3 has yet to be shown. The (GAA) expansion at the *FXN* gene provides an intriguing target for both mechanisms of repression since transcription across the repeat region produces a large repetitive intronic RNA which must

be spliced out. This long RNA can stabilize the binding of both HP1 and PRC2, providing an additional target for them to associate with. The increase in both histone binding partners and the stabilizing mRNA will restrict polymerase elongation into the gene.

1.3.3 H4K20me3 and DNA methylation in FRDA

One of the earliest histone modifications observed is methylation of H4K20 [85]. Mono-, Di- and Trimethylation are both present at nucleosomes and have different functions. Di-methylation of H4K20 is the most abundant modification, found on 80% of all H4 proteins [85]. Mono- and Dimethylation of H4K20 is generally found to aid in DNA replication by recruiting the origin replication complex (ORC), or signal for DNA damage repair through the association of 53BP1 [85]. Trimethylation on the other hand, is catalyzed by the methyltransferases SUV420H1/2 and is a hallmark of heterochromatic regions including pericentric heterochromatin, telomeres, and repetitive elements [86]. While H4K20me3 is a well-established heterochromatic mark, a specific reader protein was only recently found. The DNA methyltransferase DNMT1 reads H4K20me3 and catalyzes the methylation of DNA, preferentially at LINE1 repetitive elements [87].

Since approximately one-third of all the CpG dinucleotide regulatory elements originate from the insertion and replication of transposable elements, the cell needs a mechanism to ensure they are not inadvertently transcribed. The targeted recruitment of DNMT1 through the binding of H4K20me3 maintains the methylation of LINE CpG islands and inhibits their transcription. In FRDA, the (GAA) repeat expansion is located near the Alu-SINE element. Just 5’

of the Alu element in the first intron is a CpG island that makes for a prime target for H4K20me3 regulation. To date only one measurement of H4K20me3 has been published, which showed an increase just upstream of the GAA repeat expansion [37]. Since H4K20me3 recruits DNMT1, the increase at *FXN* should lead to increased DNA methylation across the local CpG island. Bisulfite sequencing has shown exactly that; concomitantly with the increase in H4K20me3, FRDA cells and tissues do have increased levels of DNA methylation within the *FXN* CpG island [88-90]. The specific, targeted methylation of this CpG island with a fusion dCas9-DNMT3A results in the downregulation of *FXN* [88]. Therefore, in FRDA it is likely that DNMT1 is recruited at regions enriched with H4K20me3, resulting in the methylation of the regulatory CpG island.

1.3.4 The active promoter, bivalency, and H3K4me2/3 function at *FXN*

The di- and trimethylation of histone H3 at lysine 4 is a hallmark of promoters, transcription start sites (TSS), and early regions of actively transcribed genes [44, 91, 92]. The enrichment of H3K4me2/3 occurs in similar patterns as H3K9Ac and H3K14Ac, other epigenetic modifications that signal for transcriptional activation [93]. TAF3, a component of the general transcription factor TFIID, is recruited to H3K4me3 modifications through its plant homeodomain (PHD) [94]. TFIID bound at gene promoters helps direct the assembly of the remaining general transcription factors as well as RNA polymerase II into the preinitiation complex, a large multiprotein complex that primes polymerase for promoter escape and active transcription [95, 96]. In addition to helping to recruit general transcription machinery, other large protein which facilitate active transcription such as the Nucleosome Remodeling Complex (NURF) and the Sot-Ada-Gcn5 (SAGA) complexes, recognize and associate with H3K4me2/3 modifications [97-99].

Nucleosomal DNA is wrapped around histones, which must be moved out of the way during transcription so that polymerase can transcribe unimpeded through the gene. The Nucleosome Remodeling Factor (NURF) complex functions by sliding nucleosomes along the DNA, decreasing the local nucleosome occupancy and making the DNA accessible for transcription factors and RNA polymerase [100, 101]. The largest NURF subunit is Bromodomain PHD Finger Transcription Factor (BPTF), a bifunctional recognition protein. BPTF contains two histone recognition domains, a PHD domain which binds H3K4me3 and a Bromodomain which binds acetylated lysines on Histone H4, another mark associated with active transcription [98, 100, 102]. The dual-recognition capabilities of BPTF links these modifications at the promoter of actively transcribed genes. Once NURF is assembled, it is responsible for altering nucleosome position by sliding them along the DNA.

The SAGA complex is a much larger multi-protein complex consisting of multiple component responsible for different activating functions. Chd1 and Sgf29 are important for the recognition of H3K4me2/3, the histone acetyltransferase (HAT) Gcn5 catalyzes the placement and spread of histone acetylation, and other SAGA components signal for the de-ubiquitination of H2B, the recruitment of TATA-binding protein associated factors, and the recruitment of other proteins that directly bind to general transcription machinery [103, 104]. For an in-depth review see Weake and Workman 2012. The acetyltransferase Gcn5 that is recruited with the SAGA complex catalyzes the acetylation of H3K9 and H3K14, both marks of active transcription that logically colocalize with H3K4me3 at active promoters [105, 106].

The SAGA and NURF complexes play critical roles in regulating transcription, and mutations that disrupt their function can lead to developmental defects and cancer [107]. Since

H3K4me2/3 plays such an important role in the recruitment of these transcriptional coregulators, it is perhaps very telling that we do not see a complete loss and instead only an attenuation of H3K4 methylation in FRDA at the promoter [36, 37, 108]. Instead of at the promoter, we see a major loss of H3K4me2/3 ChIP signal after the first exon, just upstream of the (GAA) repeats. The H3K4me2/3 loss in this region suggests that the modifications are failing to spread 3' onto neighboring histones during transcription. Broad H3K4me3 peaks are associated with higher levels of transcription elongation, and the sharpening of the H3K4me3 peak in FRDA may play a role in the loss of progressive transcription after polymerase pause-release [109]. It is not currently known whether the change to the H3K4me2/3 peaks directly affects the recruitment of the SAGA or NURF complexes at *FXN*, although their loss would have a detrimental effect on *FXN* expression.

The present, albeit depleted H3K4me3 modification at *FXN* along with the increase in H3K9me3 and H3K27me3 produces a bivalent system, one that signals for both active gene transcription and gene silencing concomitantly [36, 37, 108] . These systems occur infrequently at other genomic loci to allow for the loading of polymerase at the promoter, stalling its progression and preparing it for activation [110]. While uncommon, H3K4me3/H3K27me3 bivalent modifications are observed more frequently at regions of lowly transcribed and developmental genes, which primes and readies them for activation. This epigenetic bivalency used for transcriptional priming is similar to the epigenetic state at *FXN*, where the gene is partially but not completely activated.

H3K4me3/H3K9me3 bivalent domains are less common but have been found at regions of poised transcription in lineage-committed mesenchymal stem cells, attenuating the level of

positive transcription at those loci [111]. In Neurospora, H3K4me3/H3K9me3 bivalent domains have also been observed, where the placement of H3K4me3 restricts the spreading of H3K9me3 into active genes [112]. In this system the loss of methylation at K4 causes H3K9me3 encroachment into actively transcribed regions. Very few proteins have been described that recognize both of the H3K4me3/H3K9me3 modifications in bivalent domains. Recently the protein Spindlin1 was found to associate with C11orf84 and bind the H3K4me3/H3K9me3 bivalent modification at this rRNA loci [113]. It displaces HP1γ and activates transcription in these regions, which shows mechanisms exist, even if they are not well characterized, that can recognize and activate transcription at this type of bivalent chromatin. The persistence of the H3K4me3 modification in FRDA likely inhibits the spread of H3K9me3 into the promoter which maintains its openness. Therefore, the *FXN* gene appears to be poised, with H3K4me3 present at the promoter and H3K9me3 present in the gene, which results in polymerase loading at the promoter but fails to progress into the gene. H3K9me3 encroaches onto the promoter, reducing H3K4me3 signal but failing to completely silence the gene. HP1 proteins are bound to H3K9me3 modifications within the gene body, further evidence that the repression by H3K9me3 is active, but unable to completely disable the active promoter.

1.3.5 The loss of H3K36me3 and H3K79me2 and transcription elongation defects in FRDA

After transcription is initiated, polymerase will associate with the Negative Elongation Factors (NELF) and DRB Sensitivity Inducing Factor (DSIF), which cause polymerase to pause and inhibits its progression into the gene body [114]. This occurs about 50 nucleotides proximal to

the promoter, and these factors must be released before polymerase can continue [115]. Cyclin dependent kinase 9 (CDK9) phosphorylates NELF and DSIF which releases them from polymerase, allowing transcription elongation [116]. In addition, CDK9 phosphorylates serine-2 (S2P) on the C-Terminal Domain (CTD) of polymerase, another mark of active transcription [116]. CDK9 is recruited to the paused polymerase by Bromodomain Containing Protein 4 (BRD4), a protein that contains two bromodomains which recognize multiple acetylated histone lysine residues [116, 117]. Specific histone acetylation modifications recruits BRD4 to regions marked for active transcription, where it subsequently recruits the Positive Transcription Elongation Factor b complex (pTEFb), of which a major part is CDK9 [116, 118, 119]. The release of NELF/DSIF and the polymerase S2P are canonical markers of active transcription.

The CTD phosphorylation on serine-2 is recognized by the Set Domain containing protein (SETD2) as polymerase moves through the gene. As transcription occurs, SETD2 continuously methylates Histone 3 at Lysine 36 (H3K36me3), coating the gene body and making H3K36me3 a hallmark of transcription elongation [120, 121]. The role of H3K36me3 in transcription is intriguing as it is recognized by proteins whose primary role is to inhibit transcription, despite being placed during active transcription. In *Saccharomyces cerevisiae* the histone deacetylase (HDAC) RPD3s is recruited to H3K36me3 to inhibit spurious transcription, the improper recruitment of polymerase to regions 3’ of the TSS [122, 123]. The recruitment of polymerase to the wrong location results in erroneous truncated transcripts. Recently a similar pathway has been identified in humans, where H3K36me3 is recognized by the protein PWWP2A, which associates with specific components of the Nucleosome Remodeling and Development (NuRD) complex including HDACs 1 and 2 [124]. RPD3 and the HDACs 1 and 2 remove the activating

acetyl-lysine modifications in the gene body, tempering the active signal and inhibiting spurious transcription.

H3K36me3 is also recognized by PHF1, a component of the repressive PRC2 complex as discussed previously [125]. Interestingly PRC2 recruitment at regions of active elongation does not facilitate the spread of H3K27me3. Instead H3K36me3 antagonizes the methyltransferase activity of PRC2. Although the role of PRC2 in elongating transcription is still not fully understood, it likely plays a role in further inhibiting spurious transcription [126]. There is evidence that the PRC2 complex can recruit the H3K36me3 demethylase NO66, which leads to the loss of H3K36me3 and the placement of H3K27me3, inhibiting further transcription [127].

In FRDA cells, even though the promoter is partially active, H3K4me2/3 is decreased with a left shift of the peak, accompanied by increases in inhibitory methylation through the gene body [37, 108]. *FXN* is necessary for mitochondrial function, and some minimal amount of transcript must be made even in affected individuals, which means polymerase successfully elongates through the repeat expansion at some level. Along with the decrease in transcription, there is about a 50% decrease in H3K36me3 placement through the gene body [36, 37, 108]. It has not been explored whether H3K36me3 is actively working to repress spurious transcription through the recruitment of PRC2 or if it is involved with HDAC recruitment to the gene, although HDAC inhibitors are being developed and have been partially successful in relieving epigenetic inhibition [128].

Separate from the role of H3K36me3 in tempering spurious transcription, the di-methylation of H3k79 (H3K79me2) is also found at actively transcribed genes. The role

H3K79me2 plays in transcription is not fully understood. The placement of H3K79me2 is catalyzed by the methyltransferase DOT1L. DOT1L interacts with AF9, a protein shared with the Super Elongation Complex (SEC), which allows it to place H3K79me2 modifications specifically on histones that have not been methylated at H3K27 [129, 130]. The interaction of DOT1L with a component of the SEC suggests the placement of H3K79me2 may facilitate the interaction of the SEC with polymerase during elongation. Recently DOT1L has been shown to play a role in transcription initiation through direct interactions with TFIID, TBP, and TAF80 which suggests the placement of H3K79me2 at regions of progressive elongation can also function to maintain expression at key genes [131]. In FRDA cells, the H3K79me2 epigenetic modification is almost completely lost [108]. Without H3K79me2 to act as a positive feedback mechanism, proteins that stabilize transcription initiation will be depleted and polymerase will have difficulty elongating through the gene body.

1.3.6 Histone acetylation at H3K9 and H3K14 is lost during FRDA which is restored with small molecule HDAC inhibitors

The covalent acetylation of lysine residues is a common modification which changes the physiochemical properties of the residue, neutralizing the positive charge of the ε-amino group. This is accomplished through the recruitment of Histone Acetyltransferases (HATs), which catalyze the covalent attachment of the acetyl group from a donor acetyl-CoA to the amino-acid substrate. The histone acetyl-lysine regulates transcription through one of two general mechanisms. The first and most direct is by decreasing the general positive charge of the histone and enhancing its hydrophobicity, which decreases its affinity to the negatively charged DNA [44,

48, 132]. This decreases histone-DNA interactions and frees the DNA so that it can be more readily accessed. Additionally, since histone acetylation occurs on the ε-amine, it must be removed before an inhibitory methyl mark can be placed. The second mechanism of transcriptional regulation is through its interaction and recruitment of binding proteins that contain either a bromodomain, YEATS domain, or a double PHD finger (DPF) domain [132]. These domains are conserved, acetyl-lysing binding protein subunits found in a highly abundant and diverse set of proteins that regulate transcription. Bromodomains specifically are found in 46 human proteins and consist of 4 alpha helices connected through two interhelical loops, forming a hydrophobic cavity [133]. This cavity binds the uncharged acetylated lysine and can differentiate between surrounding amino acids, allowing it to modulate the recognition of specific histone tail segments [134]. The YEATS and DPF domains have deeper binding pockets which allow them to recognize acetylated residues as well as accommodate larger PTMs like lysine crotonylation [132]. Through the destabilization of the histone-DNA interaction as well as the specific engagement of acetyl-lysine recognition proteins, histone acetylation is a crucial epigenetic modification that signals for the targeted increase in gene transcription.

The TSS of actively transcribed genes is highly enriched for the acetylation of Histone H3 at both Lysine 9 and Lysine 14 (H3K9Ac/H3K14Ac) with much lower levels as you progress to the 3’ end of the gene body [135]. H3K9Ac/H3K14Ac colocalize with H3K4me3 at active promoters, and can also be found in lowly transcribed bivalent genes that contain moderate levels of H3K27me3 [135]. Acetyl marks are recognized by proteins that stabilize and facilitate transcription initiation and elongation such as BRD4, AF9, and MLL1. BRD4 contains two bromodomains, BD1 and BD2. BD1 recognizes histone acetylation, recruiting the protein to

regions of active transcription while the BD2 domain can recruit other acetylated proteins to regions of active transcription [133]. As discussed previously, BRD4 recruits pTEFβ and CDK9 to the paused polymerase to signal for transcription elongation at these regions of acetylation.

AF9 contains a YEATS domain which primarily recognizes lysine crotonylation but will also bind acetylated lysines including H3K9Ac/H3K14Ac [136]. AF9 is part of the SEC as discussed previously, a large multiprotein complex that facilitates progressive transcription. H3K9Ac is also recognized by the methyltransferase MLL1 which catalyzes and stabilizes the trimethylation of H3K4 at the promoter of genes, priming them for active transcription. Histone H3 tails that are acetylated at lysine 9 are better substrates for the placement of trimethyl marks by MLL1, further facilitating the open and active gene promoter [137].

The loss of *FXN* transcript in FRDA coincides with a loss of H3K9Ac/H3K14Ac at the promoter and through the repeat region [38, 72, 108, 138]. This suggests that both BRD4 and AF9 are not properly recruited, resulting in a depletion of both pTEFb and the SEC and the subsequent failure of polymerase to progress into the gene. This theory has driven the development of small molecule histone deacetylase inhibitors (HDACi) to restore histone acetylation, and later synthetic transcription elongation factors (SynTEFs) which directly bind to BRD4 and recruit it to the repeat region [139, 140]. Both designs have shown some success in restoring *FXN* expression which will be discussed in greater detail later, but the effectiveness of SynTEF1 and its specificity for BRD4 suggests at a minimum pTEFb and CDK9 recruitment to the stalled polymerase are necessary to facilitate pause-release and signal for elongation through the (GAA) repeats.

1.3.7 The acetylation of histone H4

Histone H4 has 4 lysine amino-acids clustered in a dense group near the N-terminal end of its unstructured tail. Each lysine residue is separated by 2-3 small, primarily glycine, uncharged amino acids. These lysines are located at K5, K8, K12, and K16 and are generally acetylated in actively transcribed genes [141]. This acetylation patch is separated from the next modified lysine, and general mark for transcriptional repression, H4K20me3 by two positively charged arginine residues on either side of a large, charged histidine residue. This difference in polarity between active and repressive epigenetic modifications changes the landscape of the histone H4 tail and its ability to interact with DNA.

H4 acetylation is recognized by a myriad of bromodomain containing proteins and other large complexes. As with other acetyl-lysine modifications, BRD4 recognizes H4K8Ac and H4K16Ac and is more actively recruited by these modifications when H3K9Ac and H3S10P is present [142, 143]. H3K9Ac facilitates active transcription through the recruitment of transcription elongation machinery while the H3S10P modification inhibits HP1 protein binding and stamps the locus as active, allowing BRD4 to bind more readily [144]. Additionally, H4K5Ac, H4K8Ac, and H4K12Ac are all recognized by the Bromodomain Containing and PHD Finger Containing protein BRPF1. The bromodomain in BRPF1 associates with acetyl modifications, while a PWWP domain recognizes H3K36me3, which couples progressive elongation with histone acetylation [145]. BRPF1 will also associate with other acetyltransferases, facilitating the spread of histone acetylation throughout the region [146]. In muscle tissue, H4 acetylation has been found to recruit DPF3, part of the ATP dependent nucleosome remodeling complex BAF [147, 148]. The BAF complex is massive, consisting of up to 15 subunits and is up to 2 MDa in size. Its

primary role is to reorganize nucleosomes in order to increase DNA accessibility, which clears the way for polymerase to transcribe through the gene [149].

Improper regulation of H4 histone acetylation will affect the recruitment of these important activating proteins and complexes. FRDA patients see a significant loss in H4 acetylation, although this may not occur equally at all loci. The Gottesfeld lab first examined these modifications by ChIP-seq and showed that there was a 50% loss of acetylation at H4K5/K8/K12/K16 in lymphoblast cells when compared to healthy individuals. Using an HDAC inhibitor (compound **4b**) they were able to show a minor restoration of H4K5Ac at *FXN* [38]. A recombinant GFP minigene with an artificial intron inserted that contained a $(GAA)_{560}$ repeat sequence displayed a significant loss of H4K5Ac and H4K8Ac, as well as a decrease in GFP expression [73]. This construct shows that the (GAA) expansion directly leads to a loss of acetylation. Subsequent experiments investigating the efficacy of HDAC inhibitors agreed with previous data and showed that FRDA fibroblasts have depleted H4K5/K8/K12/K16Ac, which is partially restored after HDACi treatment [72]. The loss of H4 acetylation resembles the loss of other acetyl-marks, following the general depletion of active epigenetic modifications. Whether there is a specific epigenetic reader that is crucial for *FXN* transcription, or a general loss of multiple factors that drive transcription failure needs to be examined further.

1.4 The role of non B-form DNA, RNA:DNA hybrids, and G-quadruplex DNA in FRDA

RNA:DNA hybrids are naturally produced and separated during transcription [150]. Both mechanisms have internal controls which facilitate the displacement of the RNA from the RNA:DNA duplex and subsequently stabilize the DNA:DNA duplex. Occasionally during

transcription, longer RNA:DNA duplexes can be formed which displace the non-template strand of DNA with the nascent RNA. Purine rich RNA forms more stable RNA:DNA hybrids than pyrimidine rich RNA and is more thermodynamically stable than duplex DNA [151]. Therefore, purine rich nascent RNA can stably anneal with the exposed single stranded DNA (ssDNA) template strand in the transcription bubble, which inhibits the reformation of duplex DNA. Because the RNA:DNA hybrid is more stable, the RNA can further unwind and displace the non-template strand from the duplex, 5' into the gene and "zip up", disrupting the DNA duplex. This results in two structures: a stable RNA:DNA duplex and an unbound and exposed ssDNA non-coding strand, both of which must be resolved before the next round of transcription can occur [152]. Additionally, the free, unprotected coding strand is likely to form further, more thermodynamically stable secondary structures. It can integrate with the RNA:DNA duplex forming a DNA:RNA:DNA triplex, form a G-quadruplex secondary structure, or form other stem-loop secondary structures. These further prevent RNA:DNA duplex resolution, and will all act as barriers to gene transcription (Fig 1.2B).

During the transcriptional repression that occurs in FRDA, polymerase must, at some low-level successfully progress through the gene in order to maintain a basal level of *FXN* transcript. As polymerase transcribes, it reaches the first intron and begins to move through the repeat region, producing a (GAA) RNA repeat. This (GAA) sequence is a long and unstructured purine sequence, which will favor the thermodynamically stable RNA:DNA duplex [151]. Deleterious RNA:DNA duplex formation is resolved through multiple redundant mechanisms in the cell. There are both RNA and DNA helicases that can recognize and resolve RNA:DNA hybrids, which release the RNA and facilitate the reformation of the duplex DNA. Perhaps the most well characterized

mechanism is through the enzymatic cleavage of the bound RNA strand by RNase H [152]. RNase H is sequence independent, and recognizes and subsequently cleaves the RNA strand of the RNA:DNA duplex. The free DNA strands are then able to re-anneal, reforming the DNA duplex. The importance of RNase H in clearing RNA:DNA hybrids has led to the speculation that it could play a beneficial role in FRDA. When polymerase slows down to navigate the (GAA) repeats, RNase H targets and cleaves RNA:DNA hybrids that are blocking polymerase progression [39].

While RNA:DNA hybrid formation is frequently deleterious, it can also play a protective role in the cell. CpG islands are enriched in the promoter regions of approximately 60% of ubiquitously and highly transcribed genes [153]. It is necessary for these CpG islands to remain unmethylated in order to maintain the openness of the promoter. If instead, these CpG islands are incorrectly methylated it can result in considerable changes in transcription at necessary genes, and frequently result in cancer [154]. The open and actively transcribed promoters of essential genes are thought to be at least partially protected from DNA methylation through R-Loop formation, which acts as an inhibitor to DNA methyltransferase function [155].

R-Loop structures were first hypothesized to be a major cause of transcriptional repression in FRDA when it was observed that transcription initiation is likely not affected by the (GAA) repeat expansion [156]. Instead, it was suggested that the formation of non-B form DNA structures may play a role in FRDA. The direct observation of RNA:DNA hybrids at *FXN* was first documented by the Gromak lab nearly 15 years later [39]. They used a RNA:DNA specific antibody, S9.6, to immunoprecipitate the RNA bound chromatin through the DNA:RNA Immunoprecipitation (DRIP) technique. They were able to show that RNA:DNA hybrids formed at *FXN* in FRDA patient derived cells centered around the repeat region, and that they were sensitive

to RNase H digestion. These observations suggest that the stabilization of duplex DNA may help to relieve the transcriptional block, an idea which has become the target of antisense oligonucleotides for drug design discussed below.

1.5 Small molecule drug development for Friedreich's Ataxia

Currently there are no approved therapeutics to treat FRDA. Patients receive palliative care focused on relieving symptoms associated with peripheral muscle decline, but this only provides comfort and does not increase patient mobility or survival. There are five different types of therapeutics targeting specific aspects of FRDA [157]. First, there are gene therapies under development which would deliver a healthy *FXN* gene to the cell using Adeno-associated virus treatments (AAV). FRDA is an autosomal recessive disease, so delivering one functional *FXN* gene would be enough to alleviate the disease. Second, research is being done to stabilize the existing low levels of mRNA that is produced. Increasing its half-life will result in an increase of the total cellular level of functional transcript and resulting FXN protein. Third, small molecules that reduce the oxidative stress in the cell are being developed and tested. These target the reactive oxygen species, decreasing cellular damage and extending the lifespan of FRDA cells. Fourth, therapeutics that directly target the mitochondria are being developed to increase electron transport chain activity. These target the main cellular dysfunction that results from the loss of *FXN* and restoring mitochondrial function will help overcome the loss of FXN protein. The fifth set of compounds are designed to increase *FXN* expression at a transcriptional level through modulating endogenous pathways and relieving the blockages that exist at the gene. I will further

discuss this fifth group, focusing on a subset of small molecules that have been developed specifically to treat either the epigenetic or RNA:DNA blockages found in FRDA.

1.5.1 HDAC inhibitors increase histone acetylation to restore active epigenetic modifications

The Gottesfeld group was the first to probe for changes in histone acetylation in FRDA patients at *FXN*. They observed a general decrease in H3K9Ac and H3K14Ac which has been replicated in multiple cell and tissue types [38, 72, 108, 138]. To increase histone acetylation, specific therapeutics to inhibit HDACs have become a source of great interest with the expectation that inhibiting the removal of histone acetylation will lead to a local increase at *FXN* and relieve the transcription blockage. Trichostatin A (TSA) was an early HDAC inhibitor originally isolated as an antifungal and antibiotic drug. It competes for the active site of the HDAC and inhibits it from binding histone acetyl groups. TSA is a broad, non-specific, and reversable inhibitor of all 10 of the class 1 and class 2 HDACs, with a low nanomolar affinity [158]. Suberoylanilide hydroxamic acid, (SAHA, vorinostat) is another pan HDAC class 1 and 2 inhibitor which was designed from TSA. SAHA is currently undergoing testing in multiple phase I trials as an anti-cancer agent [159, 160]. Unfortunately, neither TSA nor SAHA were able to restore *FXN* expression, but they were able to partially increase histone acetylation in FRDA lymphoblast cell lines [38]. Surprisingly, treatment with the SAHA derivative BML-210 showed a 2-fold increase in *FXN* transcription which suggested selectively targeting HDACs may be effective to treat FRDA.

Using BML-210 as a scaffold, the Gottesfeld group developed the compound **4b**, which was able to partially restore both *FXN* expression as well as histone acetylation in FRDA cells [38].

They further refined this compound through the addition of a para-methyl group on the unmodified phenyl ring, yielding a new molecule called **106**. Like SAHA and TSA, **106** binds tightly to the active site of the class 1 HDACs, but more specifically to HDAC1 and HDAC3, and disassociates much more slowly [161]. **106** was successfully able to restore histone acetylation at *FXN* as well as significantly increase FXN protein levels more robustly than **4b** [161]. The success of **106** in cell lines prompted further research into mouse models to determine if the results could be replicated in a living system.

The Pandolfo group had developed two new transgenic mouse lines, the knock-in/knock-in (KIKI) mouse and the knock-in/knock-out (KIKO) mouse [162]. The KIKI animal carries two copies of a $(GAA)_{230}$ repeat expansion in the first intron of the Frataxin gene, and the KIKO carries one copy, with the second *FXN* gene removed. The Gottesfeld and Pandolfo groups were able to show that **106** restores histone acetylation in KIKI mouse brains as well as increases FXN protein levels [163]. Interestingly, along with the observed increases in acetylation and protein, extremely high levels of H3K9me3 persisted at *FXN*. This provided one of the first observations that the H3K9me3 modification at *FXN* is not imperforate, and transcription can occur despite its presence.

Two new HDAC inhibitors were developed shortly afterwards based on the **106** compound, **109** and **136** [140]. Both compounds inhibited HDACs 1 and 3 just like **106**, but **109** bound at low nanomolar affinity and 20-40 times more tightly. In patient-derived peripheral blood mononuclear cells (PBMCs), **109** was able to increase *FXN* expression in a dose-dependent manner, equally or slightly better than **136**. In KIKI mouse splenocytes, **106** and **136** both restored expression while **109** was ineffective which suggested mice and humans have varying responses

to the different HDACi [140]. The effectiveness of **109** above what was achieved with **106** in human tissues sparked great interest as a likely target for drug development. Gottesfeld, Pandolfo, and Rusche further characterized **109**, showing it was effective in both increasing histone acetylation as well as activating *FXN* expression in iPSCs and differentiated neuronal cells [72].

HDAC inhibition is still being developed as a treatment for FRDA, but these compounds are not currently undergoing clinical trials. **109** targets the Class-1 HDACs, HDAC1 and HDAC3, but targeting both of these ubiquitous HDACs will result in other general side effects such as global transcriptome perturbation resulting from large changes in the epigenome (see Chapter 2). While this may make HDAC inhibitors difficult to develop as a potential long-term therapy, they have taught us some very important lessons about the transcriptional repression that occurs at *FXN.* The loss of histone acetylation at *FXN* is a functional loss, detrimental for the transcription of the gene; it is not a passive loss resulting from the failure of polymerase to transcribe. Many factors (discussed previously) that are recruited by histone acetylation help to restore the activity of the stalled polymerase. Due to the nature of HDAC inhibition, it should restore binding sites for all these available histone acetyl-binding factors that can facilitate transcription. But the restoration of histone acetylation does not completely restore *FXN* transcription, which suggests that it is a necessary but not a sufficient component in FRDA patient treatments. This may be due to other mechanisms of inhibition that remain at the gene such as H3K9me3 and H3K27me3. Increasing acetylation without removing these methyl marks results in the generation of a bivalent system. This may present enough of a barrier that the necessary transcription factors cannot be recruited to a sufficient level, partially but not completely restoring transcription.

1.5.2 The action of polyamides and synthetic transcription elongation factors in FRDA

Netropsin and distamycin are antibiotics secreted from *Streptomyces netropsis* that bind in the minor groove of B-form DNA and inhibit transcription. Netropsin and distamycin both utilize a short, repeating 2 or 3 N-methylpyrrole structure connected by amide bonds to selectively bind AT rich DNA [164]. Replacing the N-methylpyrrole with an N-methylimidazole changes the affinity of the compound to preferentially recognize GC rich sequences which became the basis for molecular recognition of DNA sequences by polyamide compounds [164, 165]. This led to the development of a molecular recognition code of polyamide binding patterns by Peter Dervan [166]. This code describes different methods of developing polyamides that target specific DNA sequences by utilizing N-methylpyrrole, N-methylimidazole, and beta alanine interactions along both DNA strands. Soon after this code was developed, Aseem Ansari in the Ptashne group, in collaboration with the Dervan group developed a targeted polyamide conjugated with a VP2 activation domain that was able to induce the transcription of a reporter construct containing multiple polyamide binding sites [167]. This was the first time a polyamide was used to both target a specific DNA sequence and recruit transcriptional machinery to that site, activating transcription.

Once the rules for DNA targeting had been established, polyamides were developed to recognize and stabilize duplex DNA and interfere with proteins that may interact with the minor groove. Emmanual Kas's group used a polyamide **P9** which targeted satellite repeats. They demonstrated that treatment of **P9** on drosophila diploid cells would displace both the satellite binding D1 protein and its binding partner HP1 from compressed repetitive DNA sequences while

being retained on active rDNA sequences [168]. This suggested that HP1 may be displaced from tightly compressed heterochromatic DNA by polyamides but persist in transcribed regions.

Polyamides can detect short DNA sequences with sub-nanomolar affinity and are tune-able to target specific sequences. This makes them an ideal compound to target disease causing STRs which are generally 2-6 nucleotides in length. The (GAA) STR that causes FRDA is repeated 50-1000+ times, which provides hundreds of binding sites for polyamides to recognize and at which to enrich. Since polyamides stabilize B-form DNA and bind in the minor groove, the Gottesfeld and Dervan groups developed the first polyamides to target the repeat expansion at *FXN*, FA1 and FA3, hypothesizing that stabilizing the duplex B-form DNA would restore expression [169]. They were able to show nuclear enrichment of fluorescently labeled polyamide and sequence specific DNA recognition, but this only resulted in a minor increase in *FXN* expression. In a plasmid construct they found evidence that FA1 and FA3 stabilized linear DNA and depleted RNA:DNA formation. These results show polyamides can enter cells, translocate to the nucleus, and recognize and bind to tandem repeat DNA sequences making them a powerful and specific DNA recognition molecule.

The development of Synthetic Transcription Elongation Factors (SynTEFs) by the Ansari group relied on the polyamide recognition structure as well as a strong understanding of the epigenetic state of the cell [139]. FRDA results in increased methylation and decreased acetylation at the *FXN* gene focused around the TSS through the repeat region, which results in the loss of multiple activating factors. The common acetyl-lysine recognition protein BRD4 associates with 5 of the 7-histone acetyl-lysine residues outlined above. BRD4 recruits pTEFb and CDK9 to phosphorylate the CTD of polymerase as well as NELF and DSIF, all signals for polymerase

to continue elongation. HDAC treatment restored acetylation but relied on BRD4 recognizing and binding to the newly acetylated histones, despite the presence of other repressive epigenetic factors. Increasing acetylation then, only resulted in a partial but not a complete restoration of transcription. The Ansari group recognized that BRD4 deficiency may be responsible for the transcription defect, and this provided a unique approach to target the problem.

BRD4 contains two bromodomains which recognize and bind acetyl-lysines. The Ansari lab coupled the polyamide FA1 to a small molecule bromodomain inhibitor JQ1, which binds with nanomolar affinity to the BET family of proteins, of which BRD4 is a member [170]. The high affinity binding of JQ1 to BRD4 inhibits it from recognizing and binding acetylated lysine residues. This makes JQ1 a potential chemotherapeutic that is being developed to target cancer cells which present with epigenetic dysregulation [171]. When complexed with a GAA-targeting polyamide, instead of acting as an inhibitor JQ1 binds tightly to the BRD4 bromodomain and recruits the protein to the (GAA) STR region of *FXN*. BRD4 is then able to recruit pTEFb and CDK9, which signals for the progressive elongation of the stalled polymerase. This signal occurs despite the epigenetic repression of H3K9me3 and H3K27me3, which makes SynTEF1 treatment sufficient to overcome the transcriptional block in FRDA patient cells. *In vitro* experiments completely restore both *FXN* mRNA and protein to the same level as healthy individuals [139].

1.5.3 Disruption of RNA:DNA hybrids can restore *FXN* expression

RNA interference (RNAi) is a relatively well understood antiviral pathway that was first discovered in plants, subsequently described by Fire and Mello in *C. elegans*, and is now

understood to be conserved among the different kingdoms [172, 173]. Briefly, RNA viruses replicate a long double-stranded RNA which is subsequently packaged into viral particles. The endonuclease DICER recognizes these duplex RNA strands and cleaves them into 21-23nt long small interfering RNA fragments (siRNA), which are unwound and loaded into the largest subunit of the RNA-Induced Silencing Complex (RISC), Argonaut 2 (Ago2). The RISC complex then uses the individual siRNA strand as a template to recognize and degrade other complementary viral RNA [173]. The position of the RISC complex as a singularly antiviral response has become a debated subject, as other functions for the complex have been discovered. The Janowski group reported that the micro-RNA 589 (miR-589) recruits Ago2 and activates COX-2 transcription. miR-589 has central mismatches that, when bound to Ago2, allow it to recognize but not cleave the transcript, which instead binds and disrupts any RNA:DNA hybrid that occurs at the COX-2 promoter [174].

During the transcription of FRDA, RNA:DNA hybrids are formed which produce a physical barrier to polymerase progression. The Corey group reasoned that if this impediment is resolved it might restore *FXN* expression. They used dsRNA to activate the RNAi pathway, targeting and degrading the (GAA) repeat RNA. In this manner they were able to show a dose-dependent increase in both *FXN* mRNA and protein, and they surmised that this was due to the elimination of the RNA:DNA hybrid [175]. To show that this was the case and that B-form DNA was restored, they used a dsRNA with specific central mismatches similar to miR-589, which allowed Ago2 to associate with but not cleave the (GAA) repeat. This method was just as effective at restoring *FXN* transcript levels as the dsRNA, which showed that inhibiting the formation of RNA:DNA hybrid is a viable method to restore transcription. Unfortunately, dsRNA is difficult to advance as a therapeutic, so the Corey lab instead designed an Antisense Oligonucleotide (ASO) from stable

locked nucleic acids (LNA) to target the repeat [175]. They showed these ASOs were also capable of binding to (GAA) nascent RNA and restoring transcription. Surprisingly that increase in *FXN* mRNA does not coincide with the restoration of the epigenetic state of the cell, and instead, similar to the other treatments, occurs despite the negative regulation.

The Corey lab has screened ASOs with different linkers and chemical modifications and identified multiple structural iterations that can activate expression by resolving the RNA:DNA hybrid [176, 177]. They also developed “gapmer” complexes which contain a central DNA portion flanked by modified RNA nucleotides, which increases (GAA) specificity [178]. The Corey lab achieved cellular uptake of their ASO through electroporation or complexed with cationic lipids that facilitate cellular entry. ASOs directly injected intracerebroventricularly (i.e., without these methods to facilitate uptake) failed to activate *FXN* expression in the brain stem, cerebrum, or cerebellum of adult or neonatal mice [179]. While unfortunate, the controls they used were effective which suggests ASOs may still enter the tissue unaided but fail to activate transcription. The authors speculated that the failure to increase *FXN* expression might be due to dosing issues, or differences between human and transgenic mouse tissue, and more work will need to be done to determine how to increase efficacy.

1.6 Conclusions

The repression of *FXN* in FRDA originates from the (GAA) repeat expansion, but it is mediated through a multitude of other factors. There is ample evidence that epigenetic regulation has been altered, which results in a general loss of histone acetylation, an increase in

histone methylation, and the restriction of polymerase movement into the gene. The observed recruitment of HP1 proteins at the *FXN* locus presents another barrier for active transcription and confirms there is silencing machinery in place to restrict access to the gene. Additionally, RNA:DNA hybrids are formed through interactions between the DNA and the nascent transcript, which must be unwound before polymerase can progress. To FRDA patients, any increase to cellular *FXN* levels can mean adding years to their life, but a cure will need to overcome all these obstacles or bypass the diseased locus altogether.

Attempts to restore histone acetylation through targeted small molecule HDAC inhibitors has been partially successful in restoring both *FXN* transcription and protein levels. Concerns remain about inhibiting ubiquitous HDAC proteins and how that will change the global transcriptome levels, and what off-target effects may occur. ASOs designed to deliberately disrupt the RNA:DNA hybrid have had some success in cells but need more research to become an effective treatment. Cell permeable and gene specific SynTEFs have shown to be effective in animal studies and are moving into phase I clinical trials. At this moment, they look to be the most promising drug in development for FRDA patients. Additionally, they provide a new technology that, due to their modular structure, can be redesigned to target other repeat-based diseases.

Rare diseases provide an extraordinary opportunity to teach us new things about biology; how sometimes small changes can result in major phenotypic outcomes. Along with the scientific advances that can be made, these are people who live their lives with the specter of their disease looming over their head. Research that progresses our understanding of FRDA along with other orphan diseases brings us closer every day to discovering a cure.

1.7 Bibliography

1. Willyard, C., *Expanded human gene tally reignites debate.* Nature, 2018. **558**: p. 354-355.
2. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
3. Hoyt, S.J., et al., *From telomere to telomere: The transcriptional and epigenetic state of human repeat elements.* Science, 2022. **376**(6588): p. eabk3112.
4. Choi, J.Y. and Y.C.G. Lee, *Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements.* PLOS Genetics, 2020. **16**(7): p. e1008872.
5. Jakubosky, D., et al., *Properties of structural variants and short tandem repeats associated with gene expression and complex traits.* Nature Communications, 2020. **11**(1): p. 2927.
6. Kramerov, D.A. and N.S. Vassetzky, *Origin and evolution of SINEs in eukaryotic genomes.* Heredity, 2011. **107**(6): p. 487-495.
7. Batzer, M.A. and P.L. Deininger, *Alu repeats and human genomic diversity.* Nature Reviews Genetics, 2002. **3**(5): p. 370-379.
8. Misiak, B., L. Ricceri, and M.M. Sąsiadek, *Transposable Elements and Their Epigenetic Regulation in Mental Disorders: Current Evidence in the Field.* Frontiers in Genetics, 2019. **10**.
9. Cordaux, R. and M.A. Batzer, *The impact of retrotransposons on human genome evolution.* Nature Reviews Genetics, 2009. **10**(10): p. 691-703.
10. Dewannieux, M., C. Esnault, and T. Heidmann, *LINE-mediated retrotransposition of marked Alu sequences.* Nature Genetics, 2003. **35**(1): p. 41-48.
11. Paulson, H., *Chapter 9 - Repeat expansion diseases*, in *Handbook of Clinical Neurology*, D.H. Geschwind, H.L. Paulson, and C. Klein, Editors. 2018, Elsevier. p. 105-123.
12. Kim, S., et al., *Structural Variation of Alu Element and Human Disease.* Genomics & informatics, 2016. **14**(3): p. 70-77.
13. Deininger, P., *Alu elements: know the SINEs.* Genome Biology, 2011. **12**(12): p. 236.

14. Subramanian, S., R.K. Mishra, and L. Singh, *Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions.* Genome Biology, 2003. **4**(2): p. R13.
15. Kejnovský, E., et al., *Expansion of microsatellites on evolutionary young Y chromosome.* PloS one, 2013. **8**(1): p. e45519-e45519.
16. Nadir, E., et al., *Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(13): p. 6470-6475.
17. Gymrek, M., *A genomic view of short tandem repeats.* Current Opinion in Genetics & Development, 2017. **44**: p. 9-16.
18. Cavanaugh, S.E. and A.S. Bathrick, *Direct PCR amplification of forensic touch and other challenging DNA samples: A review.* Forensic Science International: Genetics, 2018. **32**: p. 40-49.
19. Hodel, R.G.J., et al., *The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century.* Applications in Plant Sciences, 2016. **4**(6): p. 1600025.
20. Paulson, H., *Repeat expansion diseases.* Handb Clin Neurol, 2018. **147**: p. 105-123.
21. Depienne, C. and J.-L. Mandel, *30 years of repeat expansion disorders: What have we learned and what are the remaining challenges?* The American Journal of Human Genetics, 2021. **108**(5): p. 764-785.
22. Montermini, L., et al., *The Friedreich Ataxia GAA Triplet Repeat: Premutation and Normal Alleles.* Human Molecular Genetics, 1997. **6**(8): p. 1261-1266.
23. Robert Power, S.I.B. *Friedreich's Ataxia*. 2018 [cited 2022; Available from: https://rarediseases.org/rare-diseases/friedreichs-ataxia/.
24. Friedreich, N., *Ueber degenerative Atrophie der spinalen Hinterstränge.* Archiv für pathologische Anatomie und Physiologie und für klinische Medicin, 1863. **27**(1): p. 1-26.
25. *Friedreich Ataxia Fact Sheet | National Institute of Neurological Disorders and Stroke*. 2019 [cited 2019 04/09/2019]; NIH Friedreich's Ataxia Fact Sheet]. Available from:

https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Fact-Sheets/Friedreichs-Ataxia-Fact-Sheet.

26. Montermini, L., et al., *The Friedreich ataxia critical region spans a 150-kb interval on chromosome 9q13.* Am J Hum Genet, 1995. **57**(5): p. 1061-7.
27. Campuzano, V., et al., *Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion.* Science, 1996. **271**(5254): p. 1423-7.
28. Bidichandani, S.I., T. Ashizawa, and P.I. Patel, *Atypical Friedreich ataxia caused by compound heterozygosity for a novel missense mutation and the GAA triplet-repeat expansion.* American journal of human genetics, 1997. **60**(5): p. 1251-1256.
29. Plasterer, H.L., et al., *Development of frataxin gene expression measures for the evaluation of experimental treatments in Friedreich's ataxia.* PLoS One, 2013. **8**(5): p. e63958.
30. Cossée, M., et al., *Inactivation of the Friedreich ataxia mouse gene leads to early embryonic lethality without iron accumulation.* Human Molecular Genetics, 2000. **9**(8): p. 1219-1226.
31. Campuzano, V., et al., *Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes.* Hum Mol Genet, 1997. **6**(11): p. 1771-80.
32. Puccio, H., et al., *Mouse models for Friedreich ataxia exhibit cardiomyopathy, sensory nerve defect and Fe-S enzyme deficiency followed by intramitochondrial iron deposits.* Nat Genet, 2001. **27**(2): p. 181-6.
33. Colin, F., et al., *Mammalian Frataxin Controls Sulfur Production and Iron Entry during de Novo Fe4S4 Cluster Assembly.* Journal of the American Chemical Society, 2013. **135**(2): p. 733-740.
34. Belbellaa, B., et al., *High Levels of Frataxin Overexpression Lead to Mitochondrial and Cardiac Toxicity in Mouse Models.* Mol Ther Methods Clin Dev, 2020. **19**: p. 120-138.
35. Bulteau, A.-L., et al., *Frataxin Acts as an Iron Chaperone Protein to Modulate Mitochondrial Aconitase Activity.* Science, 2004. **305**(5681): p. 242-245.
36. Punga, T. and M. Buhler, *Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation.* EMBO Mol Med, 2010. **2**(4): p. 120-9.

37. Kim, E., M. Napierala, and S.Y. Dent, *Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia.* Nucleic Acids Res, 2011. **39**(19): p. 8366-77.

38. Herman, D., et al., *Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia.* Nat Chem Biol, 2006. **2**(10): p. 551-8.

39. Groh, M., et al., *R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome.* PLoS Genet, 2014. **10**(5): p. e1004318.

40. De Biase, I., et al., *Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription.* PLoS One, 2009. **4**(11): p. e7914.

41. Cutter, A.R. and J.J. Hayes, *A brief review of nucleosome structure.* FEBS Lett, 2015. **589**(20 Pt A): p. 2914-22.

42. Korolev, N., A.P. Lyubartsev, and L. Nordenskiöld, *A systematic analysis of nucleosome core particle and nucleosome-nucleosome stacking structure.* Scientific Reports, 2018. **8**(1): p. 1543.

43. Morgan, M.A.J. and A. Shilatifard, *Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation.* Nature Genetics, 2020. **52**(12): p. 1271-1281.

44. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Research, 2011. **21**(3): p. 381-395.

45. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-5.

46. Esteller, M., *Epigenetics in evolution and disease.* The Lancet, 2008. **372**: p. S90-S96.

47. Waddington, C.H., *An Introduction to Modern Genetics.* Proceedings of the Royal Entomological Society of London. Series A, General Entomology, 1939. **14**(4-6): p. 82-82.

48. Dawson, Mark A. and T. Kouzarides, *Cancer Epigenetics: From Mechanism to Therapy.* Cell, 2012. **150**(1): p. 12-27.

49. Yunis, J.J. and W.G. Yasmineh, *Heterochromatin, Satellite DNA, and Cell Function.* Science, 1971. **174**(4015): p. 1200-1209.
50. Gershman, A., et al., *Epigenetic patterns in a complete human genome.* Science, 2022. **376**(6588): p. eabj5089.
51. Martens, J.H., et al., *The profile of repeat-associated histone lysine methylation states in the mouse epigenome.* Embo j, 2005. **24**(4): p. 800-12.
52. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases.* Nature, 2000. **406**(6796): p. 593-9.
53. Nakayama, J., et al., *Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly.* Science, 2001. **292**(5514): p. 110-3.
54. Machida, S., et al., *Structural Basis of Heterochromatin Formation by Human HP1.* Molecular Cell, 2018. **69**(3): p. 385-397.e8.
55. Erdel, F., et al., *Mouse Heterochromatin Adopts Digital Compaction States without Showing Hallmarks of HP1-Driven Liquid-Liquid Phase Separation.* Mol Cell, 2020. **78**(2): p. 236-249 e7.
56. Bannister, A.J., et al., *Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain.* Nature, 2001. **410**(6824): p. 120-4.
57. Lachner, M., et al., *Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.* Nature, 2001. **410**(6824): p. 116-20.
58. Smothers, J.F. and S. Henikoff, *The HP1 chromo shadow domain binds a consensus peptide pentamer.* Current Biology, 2000. **10**(1): p. 27-30.
59. Cowieson, N.P., et al., *Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis.* Current Biology, 2000. **10**(9): p. 517-525.
60. Larson, A.G., et al., *Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin.* Nature, 2017. **547**(7662): p. 236-240.
61. Muchardt, C., et al., *Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha.* EMBO Rep, 2002. **3**(10): p. 975-81.

62. Vakoc, C.R., et al., *Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin.* Mol Cell, 2005. **19**(3): p. 381-91.

63. Velez, G., et al., *Evidence supporting a critical contribution of intrinsically disordered regions to the biochemical behavior of full-length human HP1γ.* J Mol Model, 2016. **22**(1): p. 12.

64. Alberti, S., A. Gladfelter, and T. Mittag, *Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates.* Cell, 2019. **176**(3): p. 419-434.

65. Keenen, M.M., et al., *HP1 proteins compact DNA into mechanically and positionally stable phase separated domains.* Elife, 2021. **10**.

66. Shinkai, Y. and M. Tachibana, *H3K9 methyltransferase G9a and the related molecule GLP.* Genes Dev, 2011. **25**(8): p. 781-8.

67. O'Carroll, D., et al., *Isolation and Characterization ofSuv39h2, a Second Histone H3 Methyltransferase Gene That Displays Testis-Specific Expression.* Molecular and Cellular Biology, 2000. **20**(24): p. 9423-9433.

68. Schultz, D.C., et al., *SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins.* Genes Dev, 2002. **16**(8): p. 919-32.

69. Falandry, C., et al., *CLLD8/KMT1F Is a Lysine Methyltransferase That Is Important for Chromosome Segregation*.* Journal of Biological Chemistry, 2010. **285**(26): p. 20234-20241.

70. Nishikawaji, T., et al., *Oncogenic roles of the SETDB2 histone methyltransferase in gastric cancer.* Oncotarget, 2016. **7**(41): p. 67251-67265.

71. Fritsch, L., et al., *A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex.* Mol Cell, 2010. **37**(1): p. 46-56.

72. Soragni, E., et al., *Epigenetic therapy for Friedreich ataxia.* Ann Neurol, 2014. **76**(4): p. 489-508.

73. Soragni, E., et al., *Long intronic GAA*TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia.* Nucleic Acids Res, 2008. **36**(19): p. 6056-65.

74. Al-Mahdawi, S., et al., *The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues.* Human Molecular Genetics, 2008. **17**(5): p. 735-746.

75. Polak, U., et al., *Alleviating GAA Repeat Induced Transcriptional Silencing of the Friedreich's Ataxia Gene During Somatic Cell Reprogramming.* Stem Cells Dev, 2016. **25**(23): p. 1788-1800.

76. Sherzai, M., et al., *HMTase Inhibitors as a Potential Epigenetic-Based Therapeutic Approach for Friedreich's Ataxia.* Front Genet, 2020. **11**: p. 584.

77. Schuettengruber, B., et al., *Genome Regulation by Polycomb and Trithorax: 70 Years and Counting.* Cell, 2017. **171**(1): p. 34-57.

78. Maclary, E., et al., *PRC2 represses transcribed genes on the imprinted inactive X chromosome in mice.* Genome Biology, 2017. **18**(1): p. 82.

79. Yu, J.R., et al., *PRC2 is high maintenance.* Genes Dev, 2019. **33**(15-16): p. 903-935.

80. Jin, Q., et al., *Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation.* Embo j, 2011. **30**(2): p. 249-62.

81. Devaiah, B.N., et al., *BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin.* Nat Struct Mol Biol, 2016. **23**(6): p. 540-8.

82. Beacon, T.H., et al., *The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes.* Clinical Epigenetics, 2021. **13**(1): p. 138.

83. Fox, S., et al., *Hyperacetylated chromatin domains mark cell type-specific genes and suggest distinct modes of enhancer function.* Nature Communications, 2020. **11**(1): p. 4544.

84. Li, J., et al., *Defining Transcription Regulatory Elements in the Human Frataxin Gene: Implications for Gene Therapy.* Hum Gene Ther, 2020. **31**(15-16): p. 839-851.

85. Jørgensen, S., G. Schotta, and C.S. Sørensen, *Histone H4 Lysine 20 methylation: key player in epigenetic regulation of genomic integrity.* Nucleic Acids Research, 2013. **41**(5): p. 2797-2806.

86. Schotta, G., et al., *A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin.* Genes Dev, 2004. **18**(11): p. 1251-62.

87. Ren, W., et al., *DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation.* Nature Communications, 2021. **12**(1): p. 2490.

88. Rodden, L.N., et al., *DNA methylation in Friedreich ataxia silences expression of frataxin isoform E.* Scientific Reports, 2022. **12**(1): p. 5031.

89. Greene, E., et al., *Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia.* Nucleic Acids Res, 2007. **35**(10): p. 3383-90.

90. Evans-Galea, M.V., et al., *FXN methylation predicts expression and clinical outcome in Friedreich ataxia.* Annals of Neurology, 2012. **71**(4): p. 487-497.

91. Santos-Rosa, H., et al., *Active genes are tri-methylated at K4 of histone H3.* Nature, 2002. **419**(6905): p. 407-411.

92. Barski, A., et al., *High-Resolution Profiling of Histone Methylations in the Human Genome.* Cell, 2007. **129**(4): p. 823-837.

93. Bernstein, B.E., et al., *Genomic Maps and Comparative Analysis of Histone Modifications in Human and Mouse.* Cell, 2005. **120**(2): p. 169-181.

94. Lauberth, S.M., et al., *H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation.* Cell, 2013. **152**(5): p. 1021-36.

95. Burley, S.K. and R.G. Roeder, *BIOCHEMISTRY AND STRUCTURAL BIOLOGY OF TRANSCRIPTION FACTOR IID (TFIID).* Annual Review of Biochemistry, 1996. **65**(1): p. 769-799.

96. Liu, X., et al., *Initiation Complex Structure and Promoter Proofreading.* Science, 2011. **333**(6042): p. 633-637.

97. Chen, Y.-J.C. and S.Y.R. Dent, *Conservation and diversity of the eukaryotic SAGA coactivator complex across kingdoms.* Epigenetics & Chromatin, 2021. **14**(1): p. 26.

98. Wysocka, J., et al., *A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.* Nature, 2006. **442**(7098): p. 86-90.

99. Sun, G., et al., *An H3K4me3 reader, BAP18 as an adaptor of COMPASS-like core subunits co-activates ERα action and associates with the sensitivity of antiestrogen in breast cancer.* Nucleic Acids Research, 2020. **48**(19): p. 10768-10784.

100. Alkhatib, S.G. and J.W. Landry, *The nucleosome remodeling factor.* FEBS Lett, 2011. **585**(20): p. 3197-207.

101. Zahid, H., N.M. Olson, and W.C.K. Pomerantz, *Opportunity knocks for uncovering the new function of an understudied nucleosome remodeling complex member, the bromodomain PHD finger transcription factor, BPTF.* Current Opinion in Chemical Biology, 2021. **63**: p. 57-67.

102. Ruthenburg, Alexander J., et al., *Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions.* Cell, 2011. **145**(5): p. 692-706.

103. Weake, V.M. and J.L. Workman, *SAGA function in tissue-specific gene expression.* Trends in Cell Biology, 2012. **22**(4): p. 177-184.

104. Pray-Grant, M.G., et al., *Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation.* Nature, 2005. **433**(7024): p. 434-438.

105. Grant, P.A., et al., *Expanded Lysine Acetylation Specificity of Gcn5 in Native Complexes *.* Journal of Biological Chemistry, 1999. **274**(9): p. 5895-5900.

106. Bonnet, J., et al., *The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription.* Genes Dev, 2014. **28**(18): p. 1999-2012.

107. Koutelou, E., C.L. Hirsch, and S.Y.R. Dent, *Multiple faces of the SAGA complex.* Current Opinion in Cell Biology, 2010. **22**(3): p. 374-382.

108. Kumari, D., R.E. Biacsi, and K. Usdin, *Repeat expansion affects both transcription initiation and elongation in friedreich ataxia cells.* J Biol Chem, 2011. **286**(6): p. 4209-15.

109. Chen, K., et al., *Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes.* Nature Genetics, 2015. **47**(10): p. 1149-1157.

110. Bernstein, B.E., et al., *A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells.* Cell, 2006. **125**(2): p. 315-326.

111. Matsumura, Y., et al., *H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation.* Molecular Cell, 2015. **60**(4): p. 584-596.

112. Zhu, Q., et al., *Histone H3 lysine 4 methyltransferase is required for facultative heterochromatin at specific loci.* BMC Genomics, 2019. **20**(1): p. 350.

113. Du, Y., et al., *Structural mechanism of bivalent histone H3K4me3K9me3 recognition by the Spindlin1/C11orf84 complex in rRNA transcription activation.* Nature Communications, 2021. **12**(1): p. 949.

114. Sims, R.J., 3rd, R. Belotserkovskaya, and D. Reinberg, *Elongation by RNA polymerase II: the short and long of it.* Genes Dev, 2004. **18**(20): p. 2437-68.

115. Yu, M., et al., *RNA polymerase II - associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II.* Science, 2015. **350**(6266): p. 1383-1386.

116. Zhou, Q., T. Li, and D.H. Price, *RNA Polymerase II Elongation Control.* Annual Review of Biochemistry, 2012. **81**(1): p. 119-143.

117. Dey, A., et al., *The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(15): p. 8758-8763.

118. Yang, Z., et al., *Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4.* Mol Cell, 2005. **19**(4): p. 535-45.

119. Adelman, K. and J.T. Lis, *Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans.* Nat Rev Genet, 2012. **13**(10): p. 720-31.

120. Venkatesh, S. and J.L. Workman, *Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development.* Wiley Interdiscip Rev Dev Biol, 2013. **2**(5): p. 685-700.

121. Li, B., et al., *The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II *.* Journal of Biological Chemistry, 2003. **278**(11): p. 8897-8903.

122. Carrozza, M.J., et al., *Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription.* Cell, 2005. **123**(4): p. 581-592.

123. Li, B., et al., *Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription.* J Biol Chem, 2009. **284**(12): p. 7970-6.

124. Zhang, T., et al., *A variant NuRD complex containing PWWP2A/B excludes MBD2/3 to regulate transcription at active genes.* Nature Communications, 2018. **9**(1): p. 3798.

125. Cai, L., et al., *An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2 Complex Targeting.* Molecular Cell, 2013. **49**(3): p. 571-582.

126. Yuan, W., et al., *H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation *.* Journal of Biological Chemistry, 2011. **286**(10): p. 7983-7989.

127. Brien, G.L., et al., *Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation.* Nature Structural & Molecular Biology, 2012. **19**(12): p. 1273-1281.

128. Soragni, E., et al., *Rationale for the development of 2-aminobenzamide histone deacetylase inhibitors as therapeutics for Friedreich ataxia.* J Child Neurol, 2012. **27**(9): p. 1164-73.

129. Feng, Q., et al., *Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain.* Current Biology, 2002. **12**(12): p. 1052-1058.

130. Chen, S., et al., *The PZP Domain of AF10 Senses Unmodified H3K27 to Regulate DOT1L-Mediated Methylation of H3K79.* Molecular Cell, 2015. **60**(2): p. 319-327.

131. Wu, A., et al., *DOT1L complex regulates transcriptional initiation in human erythroleukemic cells.* Proceedings of the National Academy of Sciences, 2021. **118**(27): p. e2106148118.

132. Sabari, B.R., et al., *Metabolic regulation of gene expression through histone acylations.* Nature Reviews Molecular Cell Biology, 2017. **18**(2): p. 90-101.

133. Engelberg, I.A., et al., *Improved methods for targeting epigenetic reader domains of acetylated and methylated lysine.* Current Opinion in Chemical Biology, 2021. **63**: p. 132-144.

134. Filippakopoulos, P., et al., *Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family.* Cell, 2012. **149**(1): p. 214-231.

135. Karmodiya, K., et al., *H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells.* BMC Genomics, 2012. **13**(1): p. 424.

136. Li, Y., et al., *AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation.* Cell, 2014. **159**(3): p. 558-71.

137. Southall, S.M., et al., *Structural Basis for the Requirement of Additional Factors for MLL1 SET Domain Activity and Recognition of Epigenetic Marks.* Molecular Cell, 2009. **33**(2): p. 181-191.

138. Al-Mahdawi, S., et al., *The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues.* Hum Mol Genet, 2008. **17**(5): p. 735-46.

139. Erwin, G.S., et al., *Synthetic transcription elongation factors license transcription across repressive chromatin.* Science, 2017. **358**(6370): p. 1617-1622.

140. Rai, M., et al., *Two new pimelic diphenylamide HDAC inhibitors induce sustained frataxin upregulation in cells from Friedreich's ataxia patients and in a mouse model.* PLoS One, 2010. **5**(1): p. e8825.

141. Lloyd, J.T. and K.C. Glass, *Biological function and histone recognition of family IV bromodomain-containing proteins.* J Cell Physiol, 2018. **233**(3): p. 1877-1886.

142. Olley, G., et al., *BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange-like syndrome.* Nat Genet, 2018. **50**(3): p. 329-332.

143. Zippo, A., et al., *Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation.* Cell, 2009. **138**(6): p. 1122-36.

144. Fischle, W., et al., *Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation.* Nature, 2005. **438**(7071): p. 1116-1122.

145. Vezzoli, A., et al., *Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1.* Nat Struct Mol Biol, 2010. **17**(5): p. 617-9.

146. Laue, K., et al., *The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity.* Development, 2008. **135**(11): p. 1935-46.

147. Lange, M., et al., *Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex.* Genes Dev, 2008. **22**(17): p. 2370-84.

148. Zeng, L., et al., *Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b.* Nature, 2010. **466**(7303): p. 258-262.

149. Alfert, A., N. Moreno, and K. Kerl, *The BAF complex in development and disease.* Epigenetics & Chromatin, 2019. **12**(1): p. 19.

150. Westover, K.D., D.A. Bushnell, and R.D. Kornberg, *Structural Basis of Transcription: Separation of RNA from DNA by RNA Polymerase II.* Science, 2004. **303**(5660): p. 1014-1016.

151. Roberts, R.W. and D.M. Crothers, *Stability and Properties of Double and Triple Helices: Dramatic Effects of RNA or DNA Backbone Composition.* Science, 1992. **258**(5087): p. 1463-1466.

152. Aguilera, A. and T. García-Muse, *R Loops: From Transcription Byproducts to Threats to Genome Stability.* Molecular Cell, 2012. **46**(2): p. 115-124.

153. Illingworth, R.S. and A.P. Bird, *CpG islands--'a rough guide'.* FEBS Lett, 2009. **583**(11): p. 1713-20.

154. Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer.* Nature Reviews Genetics, 2002. **3**(6): p. 415-428.

155. Ginno, Paul A., et al., *R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters.* Molecular Cell, 2012. **45**(6): p. 814-825.

156. Ohshima, K., et al., *Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo.* J Biol Chem, 1998. **273**(23): p. 14588-95.

157. *Friedreich's Ataxia Research Alliance Research Pipeline*. [Website] 2022 [cited 2022 5/23/2022]; Available from: https://curefa.org/pipeline.

158. Vanhaecke, T., et al., *Trichostatin A-like hydroxamate histone deacetylase inhibitors as therapeutic agents: toxicological point of view.* Curr Med Chem, 2004. **11**(12): p. 1629-43.

159. NIH. *Clinical Trials Using Vorinostat*. 2022 [cited 2022 5/3/2022]; Available from: https://www.cancer.gov/about-cancer/treatment/clinical-trials/intervention/vorinostat.

160. Marks, P.A., *Discovery and development of SAHA as an anticancer agent.* Oncogene, 2007. **26**(9): p. 1351-1356.

161. Chou, C.J., D. Herman, and J.M. Gottesfeld, *Pimelic diphenylamide 106 is a slow, tight-binding inhibitor of class I histone deacetylases.* J Biol Chem, 2008. **283**(51): p. 35402-9.

162. Miranda, C.J., et al., *Frataxin knockin mouse.* FEBS Lett, 2002. **512**(1-3): p. 291-7.

163. Rai, M., et al., *HDAC inhibitors correct frataxin deficiency in a Friedreich ataxia mouse model.* PLoS One, 2008. **3**(4): p. e1958.

164. Kopka, M.L., et al., *The molecular origin of DNA-drug specificity in netropsin and distamycin.* Proc Natl Acad Sci U S A, 1985. **82**(5): p. 1376-80.

165. Lown, J.W., et al., *Molecular recognition between oligopeptides and nucleic acids: novel imidazole-containing oligopeptides related to netropsin that exhibit altered DNA sequence specificity.* Biochemistry, 1986. **25**(23): p. 7408-7416.

166. Dervan, P.B. and R.W. Bürli, *Sequence-specific DNA recognition by polyamides.* Current Opinion in Chemical Biology, 1999. **3**(6): p. 688-693.

167. Ansari, A.Z., et al., *Towards a minimal motif for artificial transcriptional activators.* Chemistry & Biology, 2001. **8**(6): p. 583-592.

168. Blattes, R., et al., *Displacement of D1, HP1 and topoisomerase II from satellite heterochromatin by a specific polyamide.* EMBO J, 2006. **25**(11): p. 2397-408.

169. Burnett, R., et al., *DNA sequence-specific polyamides alleviate transcription inhibition associated with long GAA.TTC repeats in Friedreich's ataxia.* Proc Natl Acad Sci U S A, 2006. **103**(31): p. 11497-502.

170. Filippakopoulos, P., et al., *Selective inhibition of BET bromodomains.* Nature, 2010. **468**(7327): p. 1067-73.

171. Damiani, E., et al., *Targeting Epigenetic 'Readers' with Natural Compounds for Cancer Interception.* J Cancer Prev, 2020. **25**(4): p. 189-203.

172. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-811.

173. Berkhout, B., *RNAi-mediated antiviral immunity in mammals.* Current Opinion in Virology, 2018. **32**: p. 9-14.

174. Matsui, M., et al., *Promoter RNA links transcriptional regulation of inflammatory pathway genes.* Nucleic Acids Res, 2013. **41**(22): p. 10086-109.

175. Li, L., M. Matsui, and D.R. Corey, *Activating frataxin expression by repeat-targeted nucleic acids.* Nat Commun, 2016. **7**: p. 10606.

176. Li, L., et al., *Activation of Frataxin Protein Expression by Antisense Oligonucleotides Targeting the Mutant Expanded Repeat.* Nucleic Acid Ther, 2018. **28**(1): p. 23-33.

177. Shen, X., et al., *Activating frataxin expression by single-stranded siRNAs targeting the GAA repeat expansion.* Bioorganic & Medicinal Chemistry Letters, 2018. **28**(17): p. 2850-2855.

178. Shen, X., et al., *Progress towards drug discovery for Friedreich's Ataxia: Identifying synthetic oligonucleotides that more potently activate expression of human frataxin protein.* Bioorganic & Medicinal Chemistry, 2020. **28**(11): p. 115472.

179. Kilikevicius, A., et al., *Difficulties translating antisense-mediated activation of Frataxin expression from cell culture to mice.* RNA Biology, 2022. **19**(1): p. 364-372.

180. Sims, R.J., 3rd, et al., *Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains.* J Biol Chem, 2005. **280**(51): p. 41789-92.

181. Shinsky, S.A., et al., *Biochemical reconstitution and phylogenetic comparison of human SET1 family core complexes involved in histone methylation.* The Journal of biological chemistry, 2015. **290**(10): p. 6361-6375.

182. Wen, H., et al., *Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation.* J Biol Chem, 2010. **285**(13): p. 9322-9326.

183. Ernst, P. and C.R. Vakoc, *WRAD: enabler of the SET1-family of H3K4 methyltransferases.* Brief Funct Genomics, 2012. **11**(3): p. 217-26.

184. Zhang, W., et al., *Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells.* J Biol Chem, 2012. **287**(51): p. 43137-55.

185. Devaiah, B.N., et al., *BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin.* Nature Structural &Amp; Molecular Biology, 2016. **23**: p. 540.

186. Dhayalan, A., et al., *The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation*.* Journal of Biological Chemistry, 2010. **285**(34): p. 26114-26120.

187. Weinberg, D.N., et al., *The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape.* Nature, 2019. **573**(7773): p. 281-286.

188. Luco, R.F., et al., *Regulation of Alternative Splicing by Histone Modifications.* Science, 2010. **327**(5968): p. 996-1000.

189. Hayakawa, T., et al., *RBP2 is an MRG15 complex component and down-regulates intragenic histone H3 lysine 4 methylation.* Genes to Cells, 2007. **12**(6): p. 811-826.

190. Edmunds, J.W., L.C. Mahadevan, and A.L. Clayton, *Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation.* The EMBO Journal, 2008. **27**(2): p. 406-420.

191. Yuan, W., et al., *Heterogeneous nuclear ribonucleoprotein L Is a subunit of human KMT3a/Set2 complex required for H3 Lys-36 trimethylation activity in vivo.* J Biol Chem, 2009. **284**(23): p. 15701-7.

192. Sessa, A., et al., *SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring.* Neuron, 2019. **104**(2): p. 271-289.e13.

193. Poplawski, A., et al., *Molecular Insights into the Recognition of N-Terminal Histone Modifications by the BRPF1 Bromodomain.* Journal of Molecular Biology, 2014. **426**(8): p. 1661-1676.

194. Shang, W.-H., et al., *Acetylation of histone H4 lysine 5 and 12 is required for CENP-A deposition into centromeres.* Nature Communications, 2016. **7**(1): p. 13465.

195. Miotto, B. and K. Struhl, *HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin.* Mol Cell, 2010. **37**(1): p. 57-66.

196. Zippo, A., et al., *Histone Crosstalk between H3S10ph and H4K16ac Generates a Histone Code that Mediates Transcription Elongation.* Cell, 2009. **138**(6): p. 1122-1136.

197. Smith Edwin, R., et al., *A Human Protein Complex Homologous to the Drosophila MSL Complex Is Responsible for the Majority of Histone H4 Acetylation at Lysine 16.* Molecular and Cellular Biology, 2005. **25**(21): p. 9175-9188.

198. Rice, J.C., et al., *Histone Methyltransferases Direct Different Degrees of Methylation to Define Distinct Chromatin Domains.* Molecular Cell, 2003. **12**(6): p. 1591-1598.

199. Tachibana, M., et al., *SET Domain-containing Protein, G9a, Is a Novel Lysine-preferring Mammalian Histone Methyltransferase with Hyperactivity and Specific Selectivity to Lysines 9 and 27 of Histone H3.* Journal of Biological Chemistry, 2001. **276**(27): p. 25309-25317.

200. Tachibana, M., et al., *Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9.* Genes Dev, 2005. **19**(7): p. 815-26.

201. Kokura, K., et al., *Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion.* The EMBO Journal, 2010. **29**(21): p. 3673-3687.

202. Margueron, R. and D. Reinberg, *The Polycomb complex PRC2 and its mark in life.* Nature, 2011. **469**(7330): p. 343-349.

203. Stender, Joshua D., et al., *Control of Proinflammatory Gene Programs by Regulated Trimethylation and Demethylation of Histone H4K20.* Molecular Cell, 2012. **48**(1): p. 28-38.

Chapter 2 | Overriding Transcriptional Silencing with Synthetic Transcription Elongation Factors

This chapter has been adapted from "Partitioning of BRD4 into HP1 condensates overrides persistent repressive chromatin at microsatellite repeats in Friedreich's Ataxia". Christopher J. Brandon, Wojciech Rosikiewicz, Matthew J. Cuneo, Sam Rider, Ashraf Mohammed, Marcus Valentine, Mangesh Kaulage, Sandra J. Kietlinska, Burkhard Hoeckendorf, Wenwei Lin, Khaled Khairy, Taosheng Chen, Beisi Xu, Tanja Mittag, Aseem Z. Ansari. *Submitted Nature 2022*

Author contributions: CJB, TM, WR, MC, TC, WL, and AZA designed research; CJB, WR, MC, SR, MK, MV performed research; AM contributed new reagents; CJB, WR, WR, SR, BH, MC, KK, SJK, BX, TM and AZA analyzed the data; and CJB with AZA wrote the paper.

Acknowledgements

I would like to recognize the committed and dedicated hard work performed by my co-authors on this manuscript, which has been adapted for this thesis chapter. Your insight has been invaluable, and without your superb effort and attention to detail we would not have been able to make these important observations. I truly appreciate everything each one of you has contributed.

2.1 Introduction

More than half of the human genome is composed of repetitive sequence elements that are concealed in repressive heterochromatin [1, 2]. A signature mark of repressive heterochromatin at repeat elements is the trimethylated lysine 9 of histone H3 (H3K9me3) [3, 4]. Not only are these modifications found at regions of constitutive heterochromatin and at pericentromeric repeats, H3K9me3 is also found at genes that are repressed in specific cell lineages [5, 6]. This latter form of facultative heterochromatin occurs at genes that define fate and function of distinct cell types. Consistent with their critical role in regulating chromatin availability, aberrant placement or erasure of heterochromatin marks is linked to multiple human diseases [3].

In Friedreich's ataxia (FRDA/FA), a terminal neurodegenerative disease, expression of frataxin (*FXN*) is downregulated by expansion of GAA trinucleotide repeats within the first intron of the *FXN* gene as discussed in Chapter 1 [7]. This expansion is accompanied by increased deposition of H3K9me3 through the repeat region [8-13]. The size of the of GAA repeat expansion positively correlates with H3K9me3 enrichment, suggesting a functional role of this signature repressive mark in silencing *FXN* expression [12]. Placing expanded GAA repeats at different genomic loci [8] or downstream of heterologous gene promoters [10] drives a concomitant increase in H3K9me3 marks at the new site. Furthermore, augmenting the levels of active acetyl marks with histone deacetylase (HDAC) inhibitors alleviates the repressive impact of H3K9me3 on *FXN* in patient-derived cells and in animal models of the disease [9, 13-15].

We previously restored frataxin expression in a variety of patient-derived cells with SynTEF1, a hetero-bifunctional molecule composed of a GAA-repeat binding polyamide (PA1)

[16-18] tethered to JQ1 [19], a small molecule that binds BET proteins (Extended Data Fig. 2.1a). BET proteins, including BRD4, interact with acetylated histone tails and recruit multiple components of the transcriptional machinery [19, 20]. Within hours of treatment, SynTEF1 traffics to the nucleus, enriches at the disease-causing GAA repeat expansion, localizes BRD4 to the repressive heterochromatin, and licenses RNA polymerase II transcriptional elongation across the repressive GAA repeats [16]. Here, we investigated whether SynTEF1-aided expression of *FXN* would trigger erasure of H3K9me3 and removal of HP1 proteins from its disease-causing repeats.

2.2 Persistence of H3K9me3 despite *FXN* expression

To monitor the rewiring of epigenetic marks upon SynTEF1 treatment, we performed chromatin-immunoprecipitation (ChIP-seq) of H3K9me3 and H3K4me3, chromatin marks of transcription repression and activation respectively (Fig. 2.1a). In patient-derived GM15850 lymphoblastoid cells, the H3K4me3 profile shifted subtly downstream, in agreement with SynTEF1-licensed transcription elongation at *FXN*. A computed pseudo-velocity plot highlights the 3' vectorial displacement of H3K4me3 enrichment (Fig. 2.1a right panel and Extended Data Figs. 2.2a-g). Unexpectedly, rather than being erased, the levels of repressive H3K9me3 increased upon SynTEF1-induced transcription of *FXN* (Fig. 2.1b). While uncommon, pre-existing H3K9me3 marks can persist at some actively transcribed genes [21]. Rather than the mark itself, heterochromatin proteins (HP1) that read H3K9me3 marks and form phase-separated repressive condensates are the key effectors that present a physical barrier to transcribing polymerases [22-26]. Consistent with previous reports [8, 27], we observed enrichment of all three paralogs of HP1 (a, b, g) at *FXN*. While all three paralogs bind H3K9me3 efficiently (Fig. 2.1h), they display

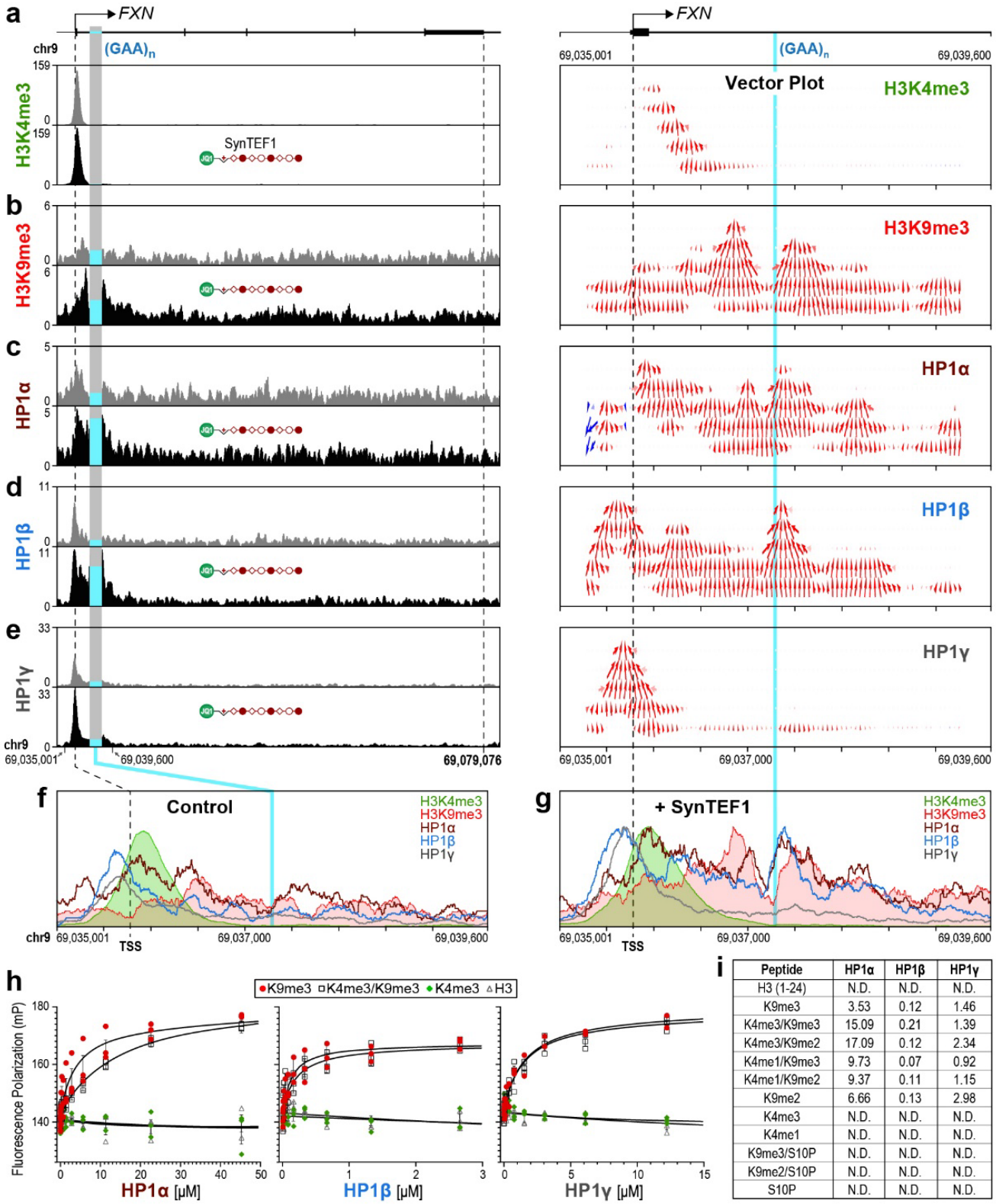
non-identical enrichment profiles across *FXN*. HP1γ enriched at the transcription start site (TSS), HP1α enriched downstream of the TSS and flanking the GAA repeat expansions, while HP1β appeared to bridge the profiles of the other two paralogs (Figs. 2.1c-f). Consistent with their repressive function, siRNA-mediated knockdown of HP1α and HP1γ isoforms resulted in significantly increased expression of *FXN* in diseased cells (Extended Data Fig. 2.1h).

2.3 HP1 enrichment accompanies *FXN* expression

DNA binding polyamides can displace HP1 from chromatin [28]. Therefore, we hypothesized that SynTEF1 would overcome the blockade to transcription elongation by displacing HP1 proteins without the need to erase underlying H3K9me3 marks. Contrary to paradigms of HP1 function, SynTEF1-treatment *enhanced* the enrichment of all three paralogs (Figs. 2.1c-g). Moreover, the most repressive paralog, HP1α, enriched directly along the path of the elongating Pol II (Fig. 2.1c and ref.[16]). HP1γ, which can permit transcription elongation when bound within coding regions [21, 29], displayed increased promoter binding consistent with its well-defined role in repressing transcription initiation (Figs. 2.1e and 2.1g). To enable comparisons, we overlaid individually scaled profiles and observed a coherent increase in repressive chromatin features accompanying active transcription elongation across the GAA repeats (Figs. 2.1f and 2.1g).

Given the overlap between active (H3K4me3) and repressive (H3K9me3) marks at the *FXN* transcription start site, we measured the ability of the three HP1 paralogs to bind orthogonally marked histone tails. Binding of 12 modified histone H3 peptides to each of the full-length HP1

paralogs was measured by fluorescence polarization (Fig. 2.1h and Extended Data Fig. 2.4). HP1 paralogs displayed high affinity for di- and tri-methylated Lysine-9 (H3K9me2/3) peptides but not for the unmodified (H3), mono/tri methylated Lysine-4 (H3K4me1/3), or the phospho-Serine10 (H3S10P) modified peptides that are known to block HP1 binding to nucleosomes (Fig. 2.1h and Extended Data Figs. 2.4a-c). All three HP1 paralogs bind peptides bearing the opposing bivalent K4me3/K9me3 marks, consistent with the ChIP-seq profiles observed at *FXN*.



i

Peptide	HP1α	HP1β	HP1γ
H3 (1-24)	N.D.	N.D.	N.D.
K9me3	3.53	0.12	1.46
K4me3/K9me3	15.09	0.21	1.39
K4me3/K9me2	17.09	0.12	2.34
K4me1/K9me3	9.73	0.07	0.92
K4me1/K9me2	9.37	0.11	1.15
K9me2	6.66	0.13	2.98
K4me3	N.D.	N.D.	N.D.
K4me1	N.D.	N.D.	N.D.
K9me3/S10P	N.D.	N.D.	N.D.
K9me2/S10P	N.D.	N.D.	N.D.
S10P	N.D.	N.D.	N.D.

Fig. 2.1 | Repressive chromatin marks and HP1 paralogs increase in response to *FXN* gene expression

a, ChIP-seq and pseudo-velocity vector analysis of H3K4me3 before and after treatment for 24 hours with DMSO vector (grey) or 1 μM SynTEF1 (black) (n=2; p=n.s.). Psuedo-velocity vector plot stratifies 50bp bins into 5 layers and compares peak changes after treatment to control peaks. The H3K4me3 vector plot visualizes a 3' movement into the gene body and towards the repeat region.

b-e, ChIP-seq and pseudo-velocity vector plot for (b) H3K9me3 (n=2; $p<1x10^{-5}$); (c) HP1α (n=2; $p<0.01$); (d) HP1β (n=2; $p<1x10^{-9}$)**;** (e) HP1γ (n=2; $p<0.05$). The all 4 peaks increase after SynTEF1 treatment at the *FXN* gene at and around the (GAA) repeat region. HP1α shows a large increase before the repeat region at a similar region as H3K9me3 peak. HP1α and HP1β show large increases after the repeats that correlate with the increase with H3K9me3. HP1γ increases at the promoter and TSS which overlaps with both peaks at H3K4me3 and H3K9me3 but shows much weaker enrichment around the (GAA) repeat.

f-g, Overlayed relative ChIP-seq peaks at *FXN* for (f) vector Control or (g) 1 μM SynTEF1 treated cells (b-e) which shows changes in peak enrichment after treatment.

h, Fluorescence polarization assays using synthetic 24 amino-acid H3 tail peptides conjugated with fluorescein at the C-terminus. Plots for H3K9me3, H3K4me3, bivalent H3K4me3+K9me3, and unmodified H3 (1-24aa) are displayed. Paired t-tests were performed and showed significance for HP1α ($p<0.05$) and HP1γ ($P<0.001$). **i,** Calculated Kd values using non-linear least squares curve fit.

2.4 *FXN* stimulation in diseased cells mirrors healthy cells

The co-localization of opposing regulatory marks detected in our ChIP studies could result from SynTEF1 activity in a limited number of "super-responder" cells that only bear active chromatin marks at the transcribed *FXN* gene. In this scenario, the seemingly co-localized repressive marks in ChIP studies would be contributed solely by unresponsive cells within the population. To investigate this possibility, we performed fluorescence in situ hybridization (FISH) using sequential probes to visualize the *FXN* genomic loci (DNA-FISH, in red) and the nascent, chromatin-associated *FXN* transcripts (via RNA-FISH, in green) (Figs. 2.2a-d). Visual inspection confirmed that in the absence of SynTEF1, *FXN* transcripts are rarely observed in untreated diseased FRDA cells (Fig. 2.2a). However, upon treatment with SynTEF1, an increase in nascent *FXN* transcript was readily evident (Fig. 2.2b). The pattern and extent of co-localization is indistinguishable from the merged DNA+RNA signals observed in healthy cells (Figs. 2b and 2.2c). Quantifying the co-localization of RNA at >13,000 *FXN* DNA loci in nine biological replicates demonstrates that SynTEF1 restores *FXN* expression across the population of diseased cells in a manner that mirrors the expression patterns in the healthy cells (Fig. 2.2e). These results from single cells and across the population of cells, unequivocally indicate that SynTEF1 overrides persistent repressive chromatin features to restore *FXN* expression in diseased cells.

Previous imaging studies of FRDA cells had suggested a causal relationship between the repression of the *FXN* gene and its position along the nuclear lamina of the nucleus [30].To determine how SynTEF1 impacts nuclear positioning of *FXN*, we measured the distance of each allele to the closest point on the nuclear periphery in treated or untreated FRDA cells (Fig. 2.2f). The distances of 1600 *FXN* alleles to the nuclear periphery are displayed as violin plots (Fig. 2.2g).

While the density profiles differed slightly, no significant repositioning of the *FXN* loci was observed upon SynTEF1-mediated expression of the gene. Consistent with this observation, healthy cells expressed both *FXN* alleles irrespective of their proximity to the nuclear periphery (Fig. 2.2d).

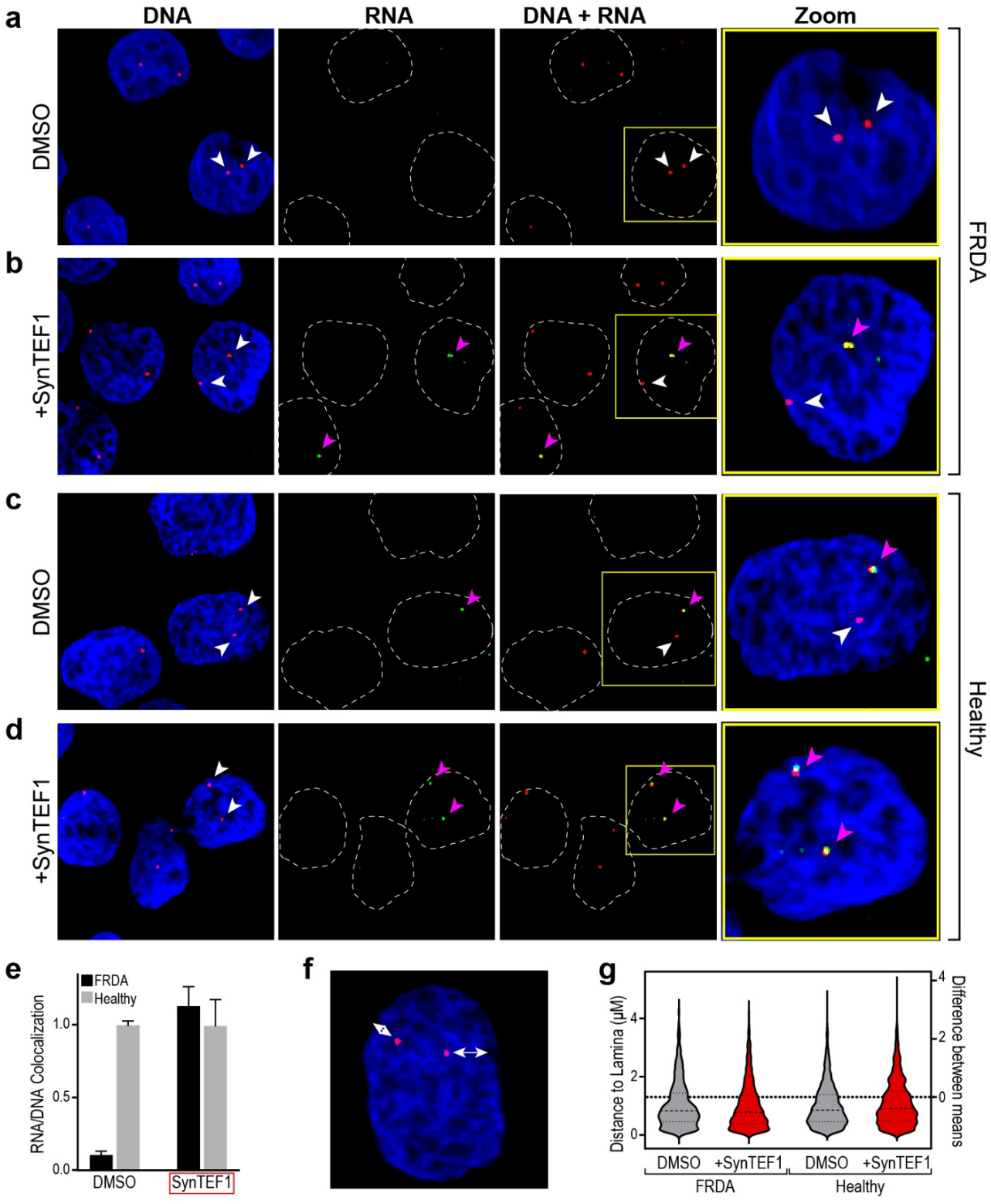

a
DNA
RNA
DNA + RNA
Zoom
DMSO
b
+SynTEF1
FRDA
c
DMSO
d
+SynTEF1
Healthy
e
FRDA
Healthy
RNA/DNA Colocalization
1.0
0.5
0.0
DMSO
SynTEF1
f
g
Distance to Lamina (μM)
4
2
0
Difference between means
4
2
0
DMSO
+SynTEF1
DMSO
+SynTEF1
FRDA
Healthy

Fig. 2.2 | FXN stimulation in diseased cells mirrors healthy cells

a-d, Sequential DNA (red) and RNA (green) Fluorescence *In situ* Hybridization (FISH) performed on (a-b) GM15850 (FRDA) and (c-d) GM15851 (Healthy) cells. The nuclei were visualized with DAPI. White arrows show inactive *FXN* alleles and pink arrows show genes with active transcription. Cells were treated with DMSO vector alone (a,c) or with 1μM SynTEF1(b,d).

e, Colocalization of RNA FISH with DNA FISH puncta normalized to healthy controls. Error bars are SEM. SynTEF1 restores nascent RNA expression to the same level as healthy cells.

f, Graphical representation of 3D DNA puncta localization measurements. Red dots are DNA FISH, whose relative location was measured to the periphery of the DAPI stain (white arrows).

g, 3D distance analysis of over 1400 DNA FISH puncta per sample using Imaris image analysis software. Nuclear lamina was determined by the periphery of DAPI stain.

2.5 Synergistic activation with an epigenetic inhibitor

To examine whether repressive chromatin attenuates SynTEF1 function, we used pharmacological agents to inhibit enzymes that place repressive marks. A sustained drug-development campaign has yielded a selective histone deacetylase inhibitor HDACi-109 (109), which stimulates the expression of *FXN* in FRDA cell lines and animal models of the disease [15]. 109 selectively inhibits members of class I histone deacetylases and prevents the erasure of transcriptionally conducive acetyl-lysine-9 (K9Ac) chromatin marks (Fig. 2.3a). Importantly, the retention of acetyl-K9 marks would prevent the subsequent placement of repressive K9me3 mark.

We observed synergistic activation of *FXN* upon co-treatment with SynTEF1 and 109, which greatly exceeded the maximal activation achieved by either small molecule alone (Fig. 2.3b, Extended Data Figure 2.8). This result is consistent with our observation that siRNA mediated knockdown of HP1α and HP1γ, which read the H3K9me3 mark, also significantly enhanced SynTEF1-dependent *FXN* expression (Extended Data Fig. 2.1i). Taken together, the results reveal that repressive chromatin attenuates *FXN* expression induced by SynTEF1. Using RNA-seq, we further examined whether the synergistic stimulation of *FXN* by 109-mediated inhibition of HDACs also triggers SynTEF1-responsive transcription initiation at other genomic loci. Consistent with previous studies, the transcriptome-wide profiles identified *FXN* as the primary target of SynTEF1 (Fig. 2.3c). On the other hand, 109 dramatically remodeled the transcriptome, with 2854 transcripts up-regulated and 1366 transcripts down-regulated (Fig. 2.3d and Extended Data Table 2.1). This level of perturbation by a freely diffusing HDAC inhibitor is not unexpected as blocking HDAC activity across the genome should have wide-ranging

disruptive consequences. Co-treatment with SynTEF1 and 109 synergistically raised *FXN* levels (Fig. 2.3e). When normalized against 109-treated cells, the residual transcriptome profile closely matched that of cells treated with SynTEF1 alone (Fig. 2.3f, Extended Data Fig. 2.7h). Reassuringly, while 109 relieved chromatin-based attenuation at *FXN*, it did not enable SynTEF1 to stimulate transcription at other genomic loci.

To investigate the basis of this remarkable synergy between SynTEF1 and 109, we measured the levels of active H3K9Ac marks across *FXN* by quantitative ChIP assays. In agreement with previous reports [9], 109 treatment leads to robust retention of H3K9Ac marks at the promoter (Fig. 2.3g, amplicon A). However, levels of H3K9Ac marks steadily declined as they approach the GAA repeats. This pattern of H3K9Ac enrichment provides a mechanistic explanation for the observed synergy in presence of 109 and SynTEF1. More importantly, it reveals the basis for the controlled restoration of *FXN* expression by SynTEF1 alone. In essence, in untreated FRDA cells, the low level of H3K9Ac at the *FXN* promoter gate the extent of transcription initiation. SynTEF1 does not alter this early regulatory step but only acts on Pol II that is stalled at the GAA repeats. In contrast, 109 treatment increases H3K9Ac levels at the promoter, disrupting promoter-gating and permitting higher levels of transcription initiation. A minor fraction of the promoter-loaded Pol II transcribes through the GAA repeats, resulting in modest increase in *FXN* levels. Thus, in the absence of 109, promoter gating and SynTEF1-mediated elongation of stalled Pol II act together to restore *FXN* expression to levels similar to those observed in healthy cells. However, in the presence of both small molecules, the 109-induced Pol II flux through the promoter is further licensed by SynTEF1 to transcribe past the repressive repeats, thus yielding dramatically increased levels of *FXN* transcripts.

2.6 Over-riding actively placed H3K9me3 marks

Given the high levels of *FXN* expression triggered by the combined action of SynTEF1 and 109, we examined if H3K9me3 repressive marks were erased under these conditions. Rather than being eroded or erased as predicted by current paradigms, the levels of H3K9me3 rose substantively over the region spanning the repeats (Fig. 2.3h, purple bars). The anti-correlated enrichment of H3K9Ac at the promoter and H3K9me3 at the expanded GAA repeats suggests regulated attenuation of transcription across *FXN*. Rather than a frozen or static epigenetic state, the data reveal that fresh repressive marks are placed in response to active transcription of disease-causing GAA repeats. The hyper-methylation of H3K9 upon co-treatment with 109 and SynTEF1 emerges as a consequence of inhibiting histone deacetylation and perturbation of the dynamic balance between erasure and re-writing of repressive marks.

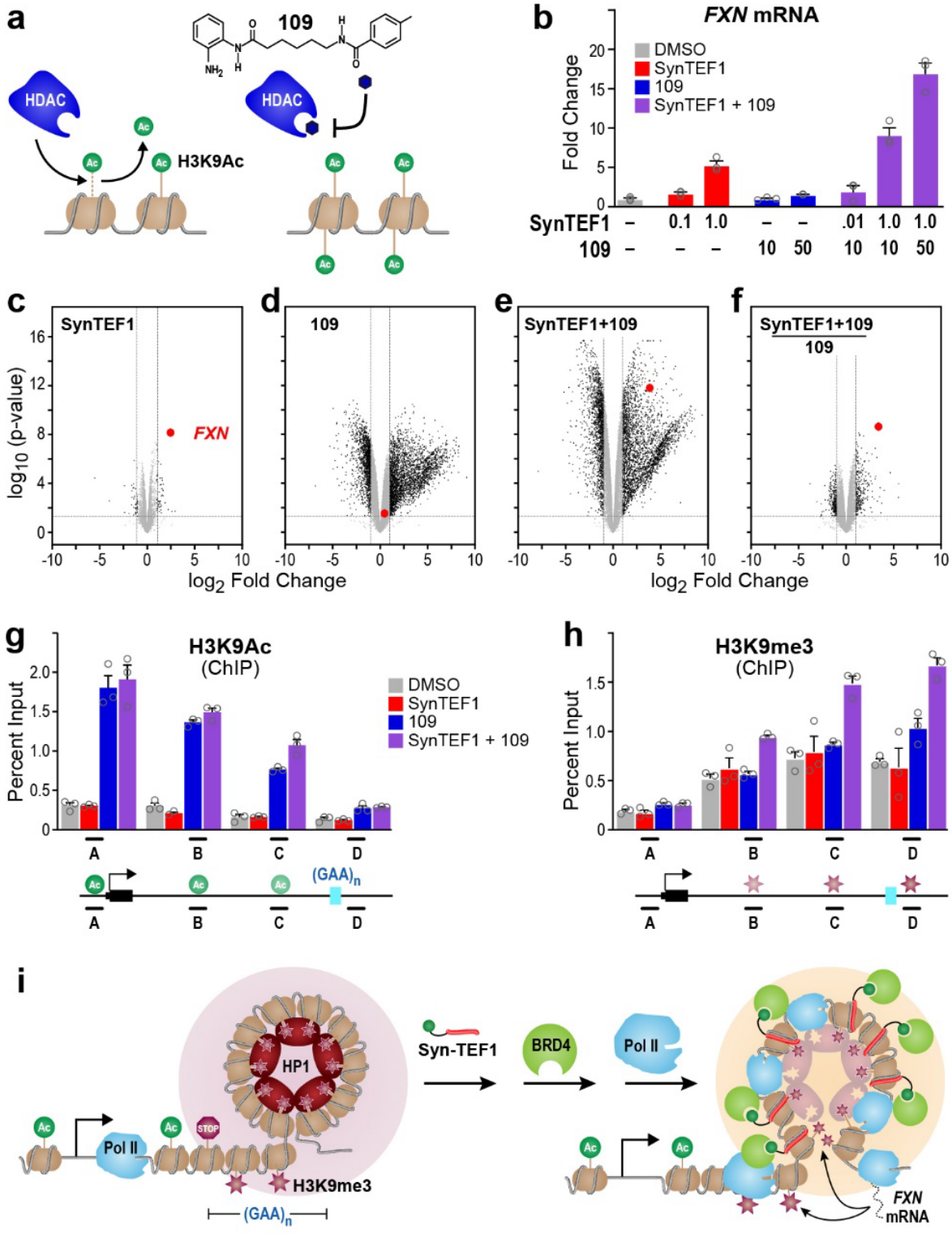

a
109
HDAC
HDAC
Ac
H3K9Ac
b
FXN mRNA
DMSO
SynTEF1
109
SynTEF1 + 109
Fold Change
SynTEF1 – 0.1 1.0 – – .01 1.0 1.0
109 – – – 10 50 10 10 50
c
SynTEF1
FXN
log10 (p-value)
log2 Fold Change
d
109
e
SynTEF1+109
log2 Fold Change
f
SynTEF1+109
109
g
H3K9Ac
(ChIP)
Percent Input
DMSO
SynTEF1
109
SynTEF1 + 109
A B C D
(GAA)n
h
H3K9me3
(ChIP)
Percent Input
A B C D
i
HP1
Pol II
STOP
H3K9me3
(GAA)n
Syn-TEF1
BRD4
Pol II
FXN
mRNA

Fig. 2.3 | Active and inhibitory epigenetic regulation calibrates polymerase licensing

a, Schematic showing how Histone Deacetylases (HDACs) catalyze the removal of acetylation from histone tails. 109 treatment binds the active site of HDACs which stabilizes histone acetylation.

b, qRT-PCR analysis of *FXN* expression normalized to DMSO control. GM15850 cells were treated for 24 hours with vehicle DMSO, SynTEF1, 109, or SynTEF1 + 109 at the indicated concentrations. Error bars are SEM of 3 replicates. 109 works synergistically with SynTEF1 to increase *FXN* expression but is does not show effects alone.

c-f, Volcano plots comparing RNAseq expression changes in response to either 1 μM SynTEF1 **(c)**, 50 μM 109 (**d)**, or 1 μM SynTEF1 + 50μM 109 **(e)** treatment. *FXN* is marked in red. Treatment with SynTEF1 and cotreatment with SynTEF1 + 109 positively regulate *FXN* expression. Both samples treated with 109 show large perturbations to the transcriptome. **f**, Transcriptional changes in SynTEF1 + 109 when corrected for 109 in co-treated samples. This shows the majority of changes that occur result from the addition of 109.

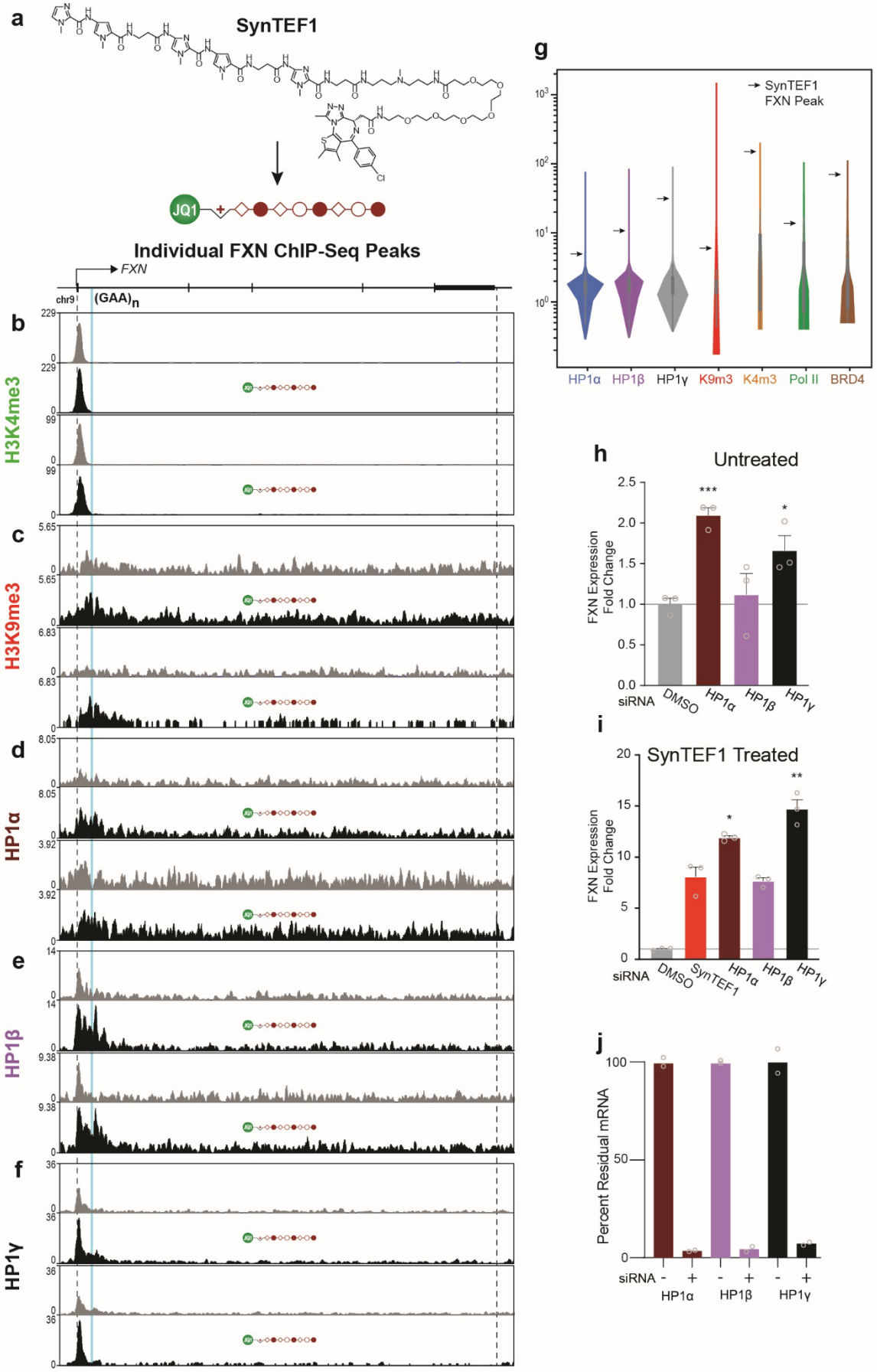
g-h, ChIP-qPCR for (g) H3K9Ac and (h) H3K9me3 normalized to percent input at specified amplicons across the *FXN* gene in GM15850 FRDA cells. Samples were treated for 24 hours with

DMSO vehicle, 1 µM SynTEF1, 50 µM 109, or 1 µM SynTEF1 + 50 µM 109. H3K9Ac is depleted in untreated and SynTEF1 treated cells and increases with 109 treatments at the promoter and TSS of the gene. H3K9me3 is present at the repeat region and increases when co-treated with SynTEF1 and 109.

i, Model of SynTEF1 mediated progressive transcription through repressive HP1 condensates during cellular feedback regulation at *FXN* in vivo. HP1 is red, Pol II is blue, and BRD4 is green. SynTEF1 treatment recruits BRD4 to the HP1 bound chromatin. This facilitates Pol II transcription through the heterochromatic region and increases the production of *FXN* transcript.

2.7 Mechanistic and Therapeutic implications

From the therapeutic perspective, our data indicate that “epigenetic drugs” that inhibit chromatin modifying enzymes will perturb promoter gating more readily than the GAA-dependent gating of elongation that we observe at *FXN* [13, 31-33]. Perturbing the flux between initiation and elongation results in loss of homeostatic control, overexpression of *FXN,* and undesirable therapeutic outcomes. Furthermore, freely diffusing epigenetic drugs globally dysregulate gene expression thereby limiting their therapeutic utility in non-acute diseases such as Friedreich’s ataxia. In contrast, SynTEF1 is designed to remove the elongation block placed by the GAA-repeat expansions in *FXN*. By enabling transcription elongation without eroding or erasing the repressive environment imposed at the repeat expansions, SynTEF1 provides a revealing insight into the fluidity and context-dependence of the histone code. While the signature epigenetic marks function as expected by convention, the ease with which they can be overridden to permit expression of critical genes was entirely unexpected.


a
SynTEF1
JQ1
Individual FXN ChIP-Seq Peaks
FXN
chr9
(GAA)n
b
H3K4me3
c
H3K9me3
d
HP1α
e
HP1β
f
HP1γ
g
SynTEF1
FXN Peak
HP1α HP1β HP1γ K9m3 K4m3 Pol II BRD4
h
Untreated
FXN Expression
Fold Change
siRNA DMSO HP1α HP1β HP1γ
i
SynTEF1 Treated
FXN Expression
Fold Change
siRNA DMSO SynTEF1 HP1α HP1β HP1γ
j
Percent Residual mRNA
siRNA - + - + - +
HP1α HP1β HP1γ

Extended Data Figure 2.1 | Epigenetic silencing caused by (GAA)n repeat expansion at FXN remains after SynTEF1 mediated transcriptional activation.

a, The chemical structure of SynTEF1 consists of three important subunits. The JQ1 "warhead" of the compound is attached to the N-terminus of the structure which binds to the bromodomains of the BET family of proteins. This is attached to a PEG6 linker which bridges the warhead and the polyamide. Lastly comes the polyamide is the largest part of the molecule and responsible for DNA-recognition and is located on the C-terminus consists of a linker dimethylaminopropylamine bound to β-Im-β-Py-Im-β-Py-Im where β is β-alanine, Py is N-methyl pyrrole, and Im is N-methyl imidazole.

b-f, Individual ChIP-seq tracks for H3K4me3 **(b)**, H3K9me3 **(c)**, HP1α **(d)**, HP1β **(e)**, and HP1γ **(f)**. Samples were either treated with the DMSO vehicle (grey) or with 1 μM SynTEF1 (black) for 24 hours. The GAA repeat region is marked in blue. The HP1 and H3K9me3 peaks increase after SynTEF1 treatment, while H3K4me3 shifts 3' into the gene.

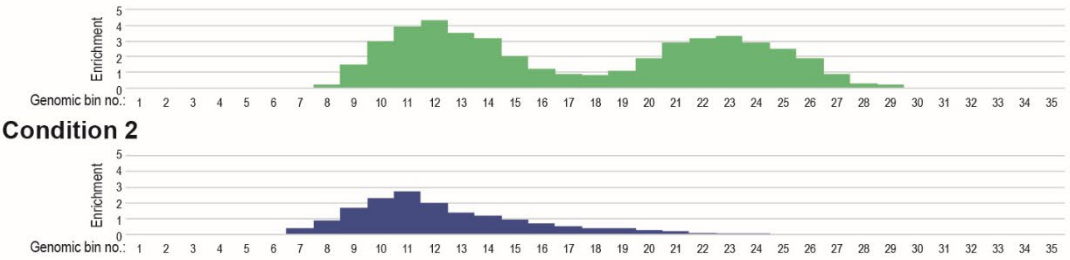
g, Violin plot of vehicle treated ChIP-seq peaks. Arrows show the signal enrichment at the *FXN* peak after SynTEF1 treatment.

h, *FXN* expression for HP1 proteins after siRNA treatment in GM15850 FRDA lymphoblasts. Expression is normalized to DMSO untreated samples. Error bars are SEM, p-values measured by unpaired t-test compared to DMSO. HP1α $p<0.01$, HP1γ $p<0.05$. The knockdown of HP1α and HP1γ increase *FXN* expression while HP1β shows no change.

i, *FXN* expression after siRNA knockdown and subsequent SynTEF1 treatment normalized to DMSO. Error bars are SEM, p-values measured by unpaired t-test compared to SynTEF1. HP1α $p<0.05$, HP1γ $p<0.01$. SynTEF1 works synergistically with the loss of HP1α and HP1γ to increase *FXN* expression.

j, Percent residual mRNA after siRNA treatment compared with untreated cells. All three knockdowns reduced mRNA levels to <10% of the control.

a. Classic visualization of bedGraph in genome browser



b. 1-D enrichment conversion to 2-D matrix

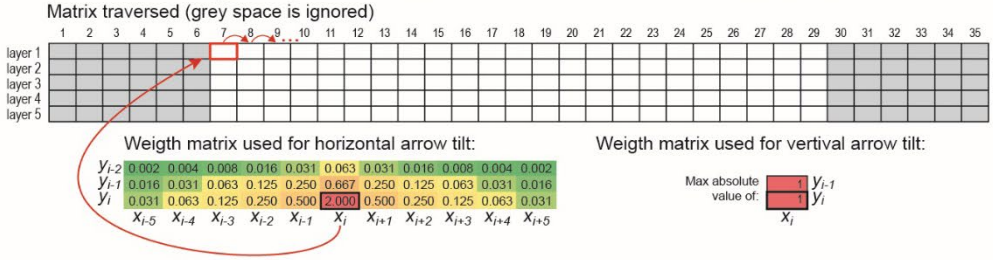
Genomic bin no.:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
layer 1	0	0	0	0	0	0	0	0	0	0	0	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
layer 2	0	0	0	0	0	0	0	0	0	0	0.9	1	0.5	0.2	0	0	0	0	0	0	0	0.2	0.3	0	0	0	0	0	0	0	0	0	0	0	0
layer 3	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0.9	1	1	0.9	0.5	0	0	0	0	0	0	0	0	0	0
layer 4	0	0	0	0	0	0	0	0.5	1	1	1	1	1	1	0.2	0	0	0.1	0.9	1	1	1	1	1	1	0.9	0	0	0	0	0	0	0	0	0
layer 5	0	0	0	0	0	0	0	0.2	1	1	1	1	1	1	1	1	0.9	0.8	1	1	1	1	1	1	1	1	0.9	0.3	0.2	0	0	0	0	0	0
Enrichment (value from bedGraph file)	0	0	0	0	0	0	0	0.2	1.5	3	3.9	4.3	3.5	3.2	2	1.2	0.9	0.8	1.1	1.9	2.9	3.2	3.3	2.9	2.5	1.9	0.9	0.3	0.2	0	0	0	0	0	0

Genomic bin no.:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
layer 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
layer 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
layer 3	0	0	0	0	0	0	0	0	0	0.3	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
layer 4	0	0	0	0	0	0	0	0	0.7	1	1	1	0.4	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
layer 5	0	0	0	0	0	0	0.4	0.9	1	1	1	1	1	1	1	0.7	0.5	0.4	0.4	0.3	0.2	0.1	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0	0
Enrichment (value from bedGraph file)	0	0	0	0	0	0	0.4	0.9	1.7	2.3	2.7	2	1.4	1.2	1	0.7	0.5	0.4	0.4	0.3	0.2	0.1	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0	0

c. Subtraction of the values of Condition 2 from Condition 1

Genomic bin no.:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
layer 1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
layer 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	1.0	0.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
layer 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.3	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	1.0	1.0	0.9	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
layer 4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.2	0.0	0.0	0.0	0.6	0.8	1.0	0.2	0.0	0.0	0.1	0.9	1.0	1.0	1.0	1.0	1.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
layer 5	0.0	0.0	0.0	0.0	0.0	0.0	-0.4	-0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.4	0.4	0.6	0.7	0.8	0.9	1.0	1.0	1.0	1.0	0.9	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0
Diff. Enrichment sum	0.0	0.0	0.0	0.0	0.0	0.0	-0.4	-0.7	-0.2	0.7	1.2	2.3	2.1	2.0	1.1	0.5	0.4	0.4	0.7	1.6	2.7	3.1	3.3	2.9	2.5	1.9	0.9	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0

d. Traverse path and weighting for horizontal / vertical arrow tilt calculations



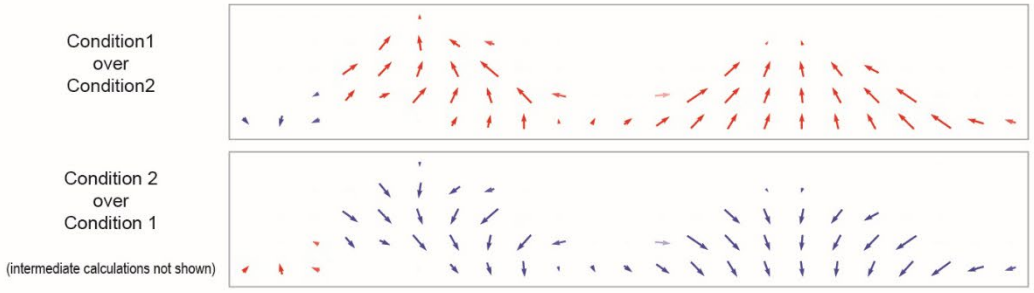
e. Calculated values for horizontal arrow tilt

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
						0.01	0.02	0.04	0.07	0.15	0.00	-0.15	-0.07	-0.04	-0.02	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00						
						0.09	0.20	0.41	0.81	0.73	-0.15	-0.70	-0.65	-0.43	-0.21	-0.09	-0.01	0.04	0.09	0.18	0.15	-0.10	-0.20	-0.10	-0.05	-0.02	-0.01	0.00						
						0.18	0.40	0.84	0.98	0.86	0.35	-0.48	-1.17	-1.12	-0.52	-0.17	0.11	0.43	0.94	0.98	0.41	-0.20	-0.71	-0.93	-0.72	-0.34	-0.15	-0.06						
						0.02	0.08	-0.39	0.60	0.67	0.87	0.47	-0.30	-0.95	-0.92	-0.33	0.31	1.01	1.33	0.94	0.37	-0.09	-0.55	-1.00	-1.32	-1.09	-0.52	-0.24						
						0.37	-0.18	-0.52	-0.07	0.15	0.34	0.33	0.16	-0.02	-0.05	0.22	0.51	0.84	0.86	0.61	0.36	0.02	-0.23	-0.54	-1.01	-1.29	-1.05	-0.70						

f. Calculated values for vertical arrow tilt

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
						0.00	0.00	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00						
						0.00	0.00	0.00	0.00	0.90	1.00	0.50	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.30	0.00	0.00	0.00	0.00	0.00	0.00						
						0.00	0.00	0.00	0.70	0.90	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.90	1.00	1.00	0.90	0.50	0.00	0.00	0.00	0.00						
						0.00	0.00	-0.20	0.70	0.30	1.00	1.00	1.00	1.00	0.20	0.00	0.00	0.10	0.90	1.00	1.00	1.00	1.00	1.00	0.90	0.00	0.00	0.00						
						-0.40	-0.70	-0.20	0.00	0.00	0.00	0.60	0.80	1.00	0.30	0.40	0.40	0.60	0.90	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.30	0.20						

g. Final visualization with "Quiver plot"



Extended Data Figure 2.2 | Velocity plots are a method to visualize the change in ChIP-seq peak distribution

a, Signal enrichment from two toy example conditions, both of which were pre-normalized to the same sequencing depth, and which both use the same bin size (in our case genome was separated in 50bp long bins), are visualized.

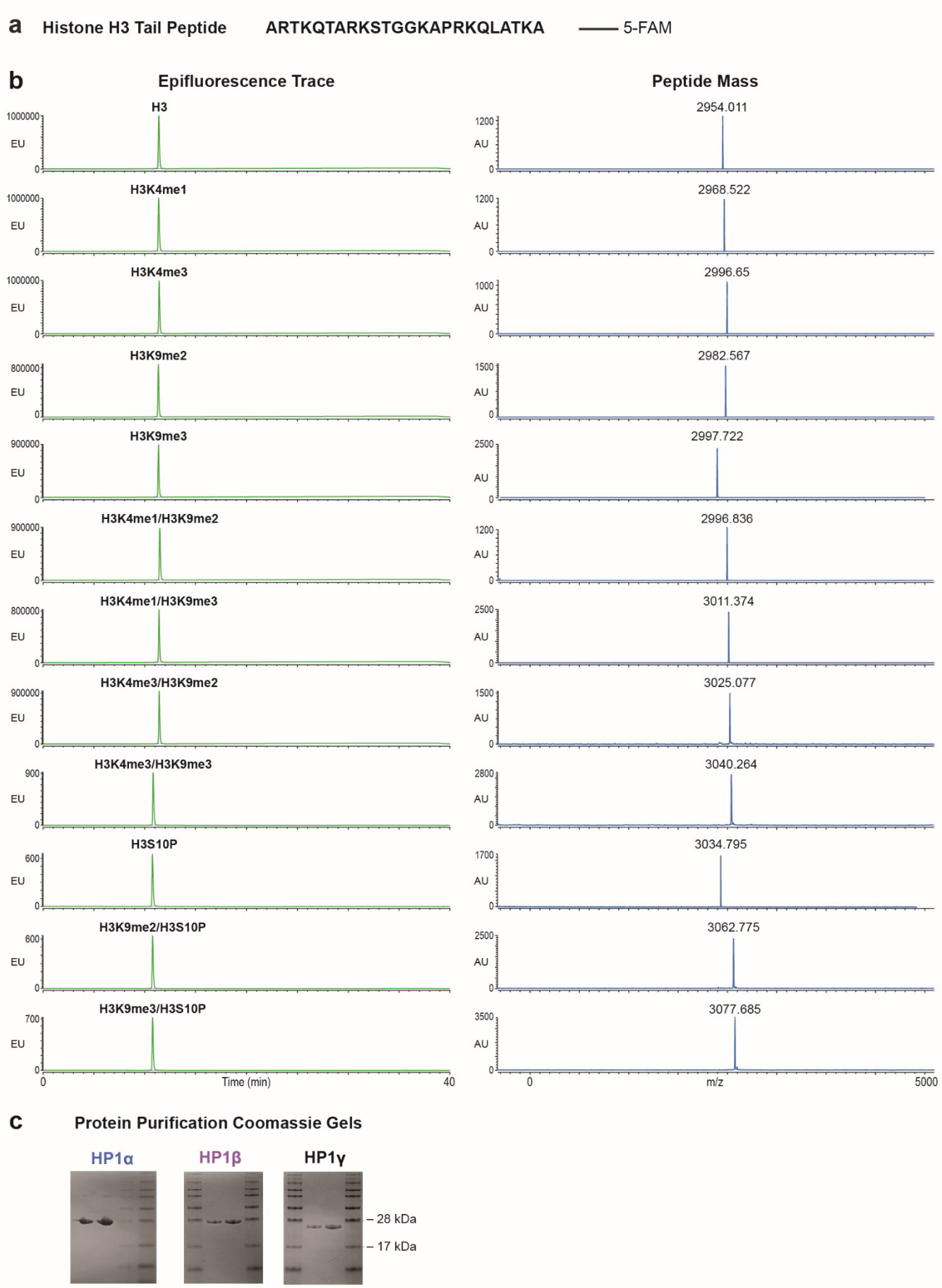
b, Layering of the max integer peak value. Each layer value is derived from the normalized value in each bin, with each layer value set from 0-1. When each layer reaches 1, the next layer is filled until the upper layer is reached.

c, Graphical representation of the difference between the 2D matrices (b). Positive values are in orange and negative values in purple.

d, Horizontal and vertical weight matrix to calculate arrow tilt.

e-f, Horizontal **(e)** and vertical **(f)** weight matrix calculations for examples peaks.

g, Velocity plots generated comparing test condition 1 to 2 or condition 2 to 1.

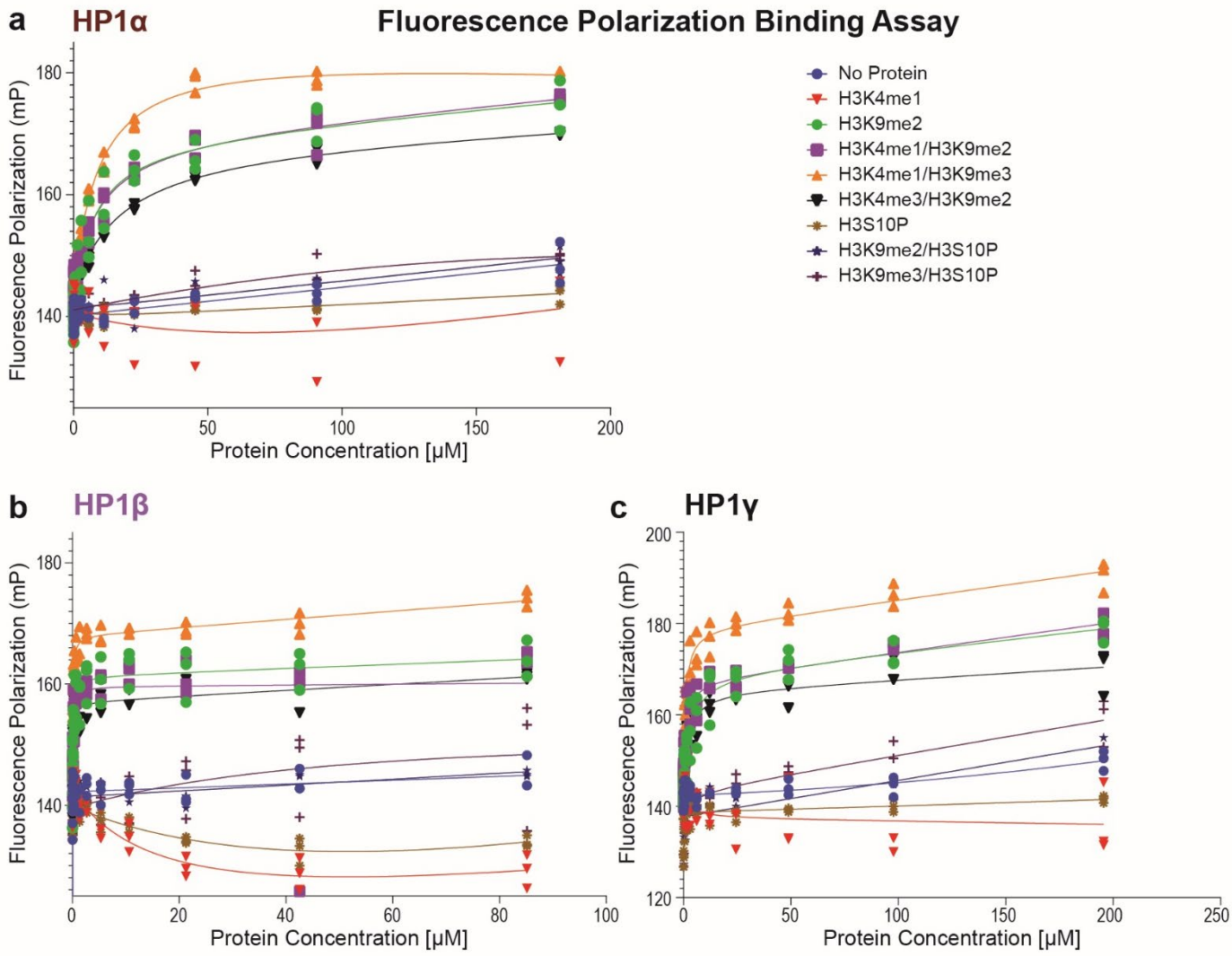

a
Histone H3 Tail Peptide
ARTKQTARKSTGGKAPRKQLATKA
5-FAM
b
Epifluorescence Trace
Peptide Mass
H3
2954.011
H3K4me1
2968.522
H3K4me3
2996.65
H3K9me2
2982.567
H3K9me3
2997.722
H3K4me1/H3K9me2
2996.836
H3K4me1/H3K9me3
3011.374
H3K4me3/H3K9me2
3025.077
H3K4me3/H3K9me3
3040.264
H3S10P
3034.795
H3K9me2/H3S10P
3062.775
H3K9me3/H3S10P
3077.685
EU
AU
Time (min)
m/z
c
Protein Purification Coomassie Gels
HP1α
HP1β
HP1γ
28 kDa
17 kDa

Extended Data Figure 2.3 | Labeled H3 modified peptide and protein generation

a, Figure showing sequence of H3 tail peptides created with C-Terminal FAM label that is used in Fluorescence Polarization studies.

b, HPLC epifluorescence traces and MALI-TOF spectra of modified H3 tail peptides after purification which shows the purity of the individual histone tails that were produced. Peptide purity was not measured but no large impurities were detected.

c, Coomassie gel of purified recombinant HP1α, HP1β, and HP1γ purified from *E. coli.*



d

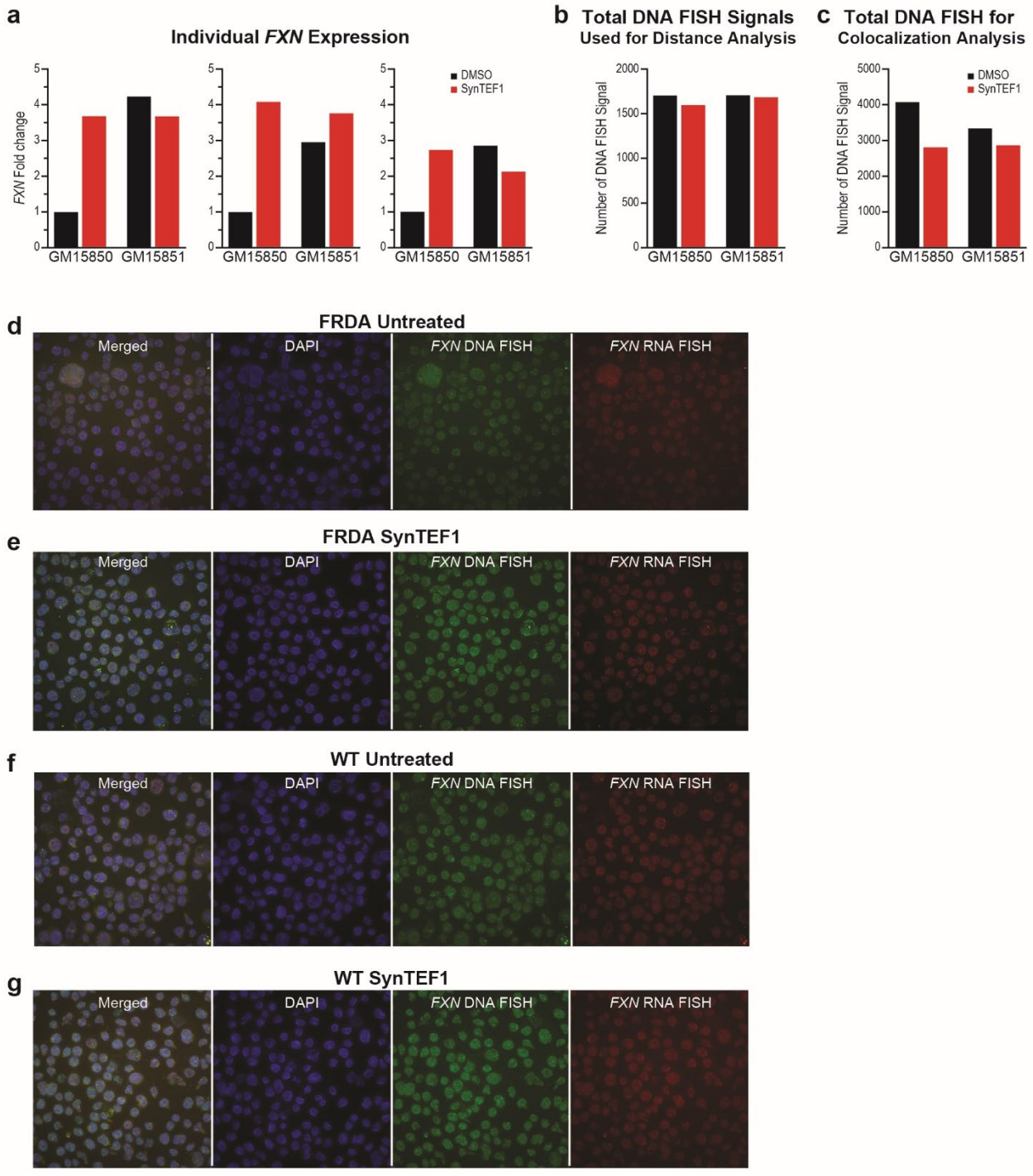
Fluorescence Polarization Kd Values (μM)

	H3K9me2	H3K4me1/H3K9me2	H3K4me1/H3K9me3	H3K4me3/H3K9me2
HP1α	6.657	9.367	9.725	17.09
HP1β	0.1336	0.1079	0.07343	0.1172
HP1γ	2.984	1.154	0.9241	2.337

Extended Data Figure 2.4 | HP1 binding of modified H3 tail peptides by fluorescence polarization

a-c, Fluorescent polarization binding assay of HP1α **(a)**, HP1β **(b)**, and HP1γ **(c)** to the indicated Histone H3 tail peptides, N=3. Data shows all three HP1 paralogues bind strongly to H3 di-modified peptides.

d, KD values of the specified recombinant HP1 protein with the specified fluorescein labeled peptide. Mono- and tri- methylation of H3K4 does not inhibit HP1 protein binding.


a
Individual FXN Expression
FXN Fold change
DMSO
SynTEF1
GM15850 GM15851
b Total DNA FISH Signals
Used for Distance Analysis
Number of DNA FISH Signal
c Total DNA FISH for
Colocalization Analysis
d
FRDA Untreated
e
FRDA SynTEF1
f
WT Untreated
g
WT SynTEF1
Merged
DAPI
FXN DNA FISH
FXN RNA FISH

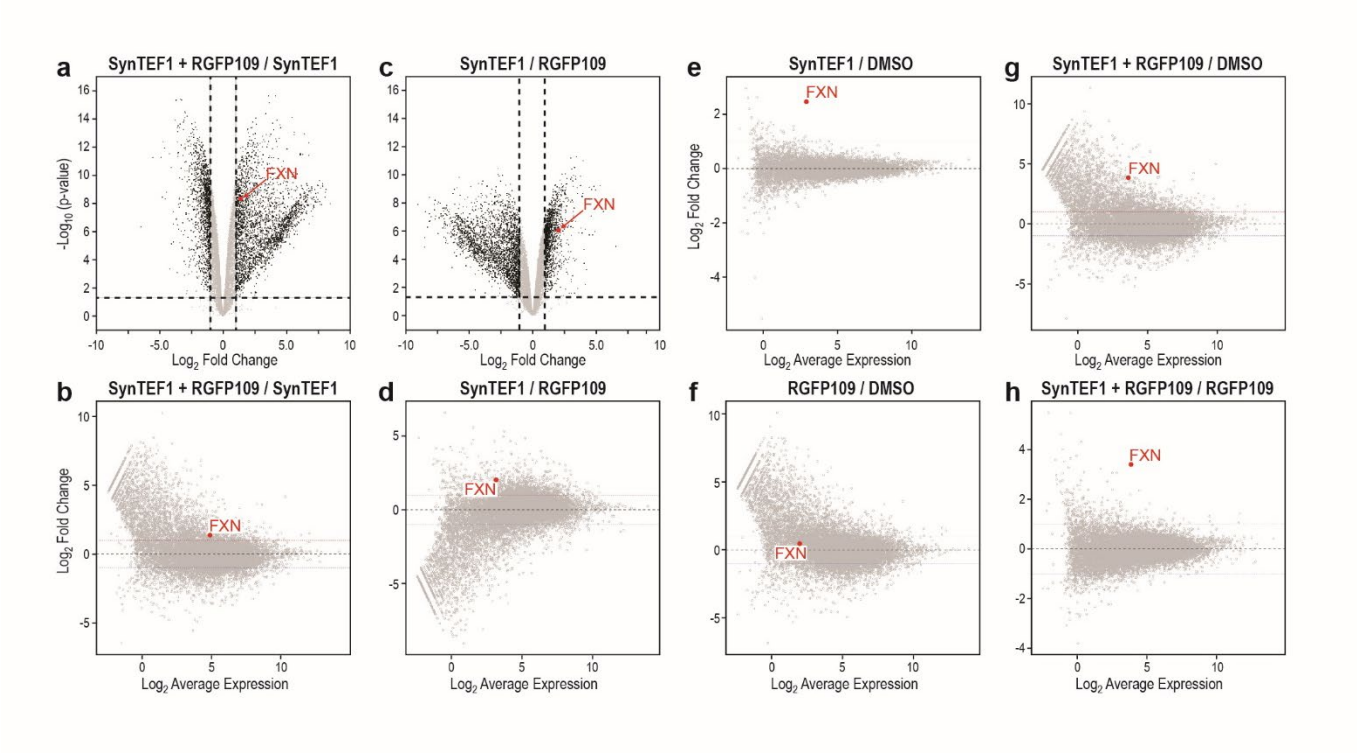
Extended Data Figure 2.5 | SynTEF1 stimulates Frataxin expression in a population of cells

a, *FXN* expression in GM15850 (FRDA) and GM15851 (Healthy) cells treated with DMSO or SynTEF1. Samples are normalized to GM15850 DMSO samples. SynTEF1 restored *FXN* expression in FRDA cells to levels similar to healthy controls.

b, Total DNA-FISH puncta used for the distance analysis for each treatment. Over 6,000 DNA puncta were used for distance analysis.

c, Total DNA-FISH puncta used for colocalization with RNA-FISH for each treatment. Over 12,000 puncta total were analyzed for RNA-DNA colocalization.

d-g, Representative full-sized images of merged and individual RNA and DNA FISH signals. GM15850 cells treated for 24 hours with DMSO vehicle **(d)** or 1 μM SynTEF1 **(e)**, and GM15851 healthy lymphoblasts treated with DMSO vehicle **(f)** or 1 μM SynTEF1 **(g)**.


a
SynTEF1 + RGFP109 / SynTEF1
-Log10 (p-value)
Log2 Fold Change
FXN
c
SynTEF1 / RGFP109
Log2 Fold Change
FXN
e
SynTEF1 / DMSO
Log2 Fold Change
Log2 Average Expression
FXN
g
SynTEF1 + RGFP109 / DMSO
Log2 Average Expression
FXN
b
SynTEF1 + RGFP109 / SynTEF1
Log2 Fold Change
Log2 Average Expression
FXN
d
SynTEF1 / RGFP109
Log2 Average Expression
FXN
f
RGFP109 / DMSO
Log2 Average Expression
FXN
h
SynTEF1 + RGFP109 / RGFP109
Log2 Average Expression
FXN

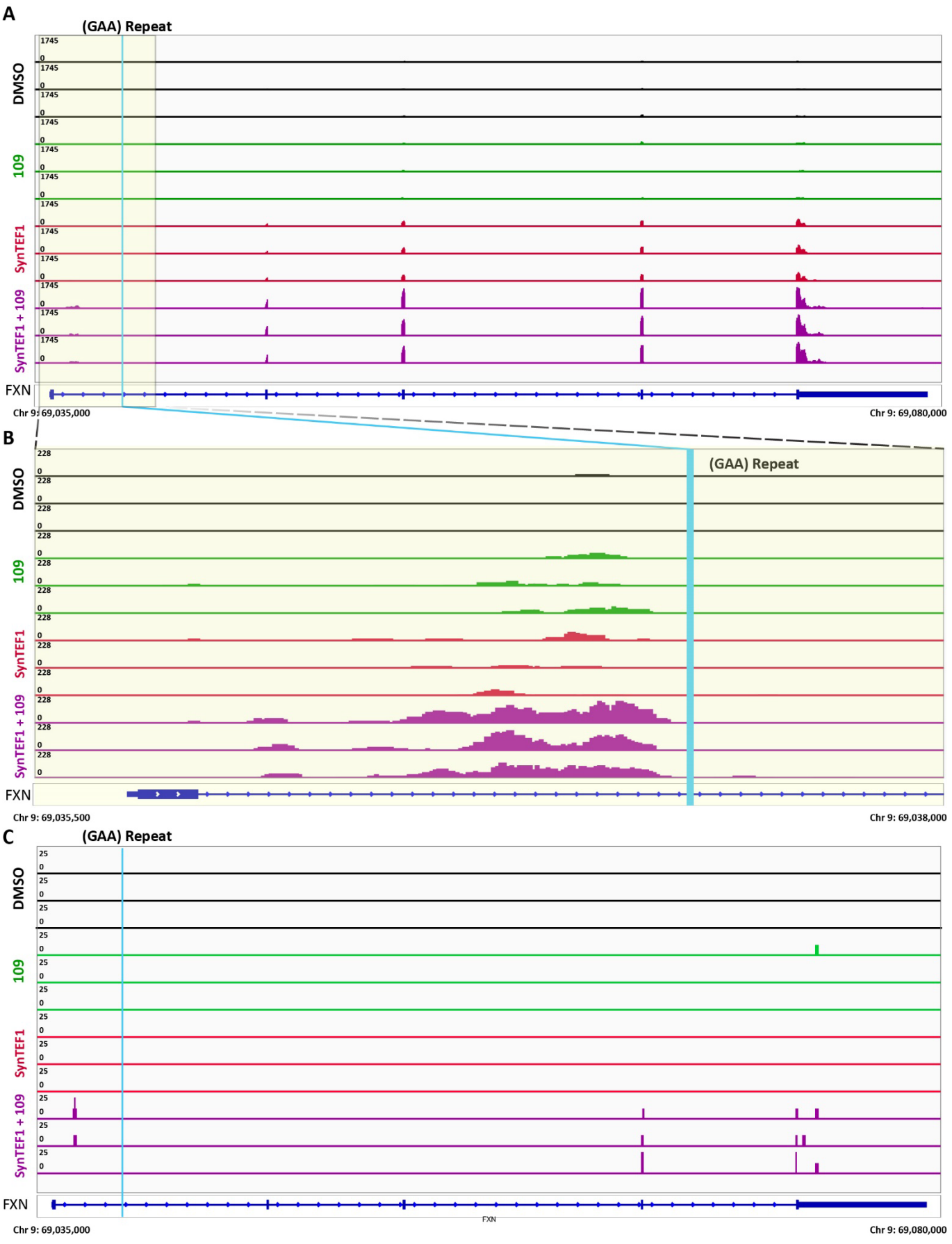
Extended Data Figure 2.6 | SynTEF1 and 109 work synergistically to activate *FXN* expression

a-b, Volcano plot comparing the Log_2 of the p-value on the y-axis vs. Log_2 of the fold change over DMSO on the x-axis **(a)**, and M&A plot showing the Log_2 of the Fold change on the y-axis vs. the Log_2 of the average expression on the x-axis **(b).** This shows how much of the transcriptome changes are associated with 109 treatment vs SynTEF1 resulting from the cotreatment of SynTEF1 and 109. 109 is responsible for larger changes in lowly transcribed genes and significant changes in actively transcribed genes.

c-d, Volcano (**c)** and M&A plot (**d)** showing RNAseq results of the treatment of GM15850 lymphoblasts with 1 µM SynTEF1 for 24 hours normalized to treatment with 50 µM 109. These data show that the individual treatments of SynTEF1 and 109 effect the transcriptome at different locations which, when you correct for 109, results in negative differences in fold change.

e-g, M&A plots showing for samples treated with SynTEF1 **(e)**, 109 **(f)**, SynTEF1 + 109 **(g)** compared to DMSO. Perturbations are more frequently seen in lowly transcribed genes in response to SynTEF1 than the more highly expressed genes. 109 and SynTEF1 + 109 treatments both effect gene transcription both positively and negatively more than SynTEF1 alone.

h, M&A plot showing RNAseq results from the cotreatment of SynTEF1 + 109 normalized to 109. The normalization to 109 decreases the transcriptome changes to mirror more closely those seen after SynTEF1 treatment.

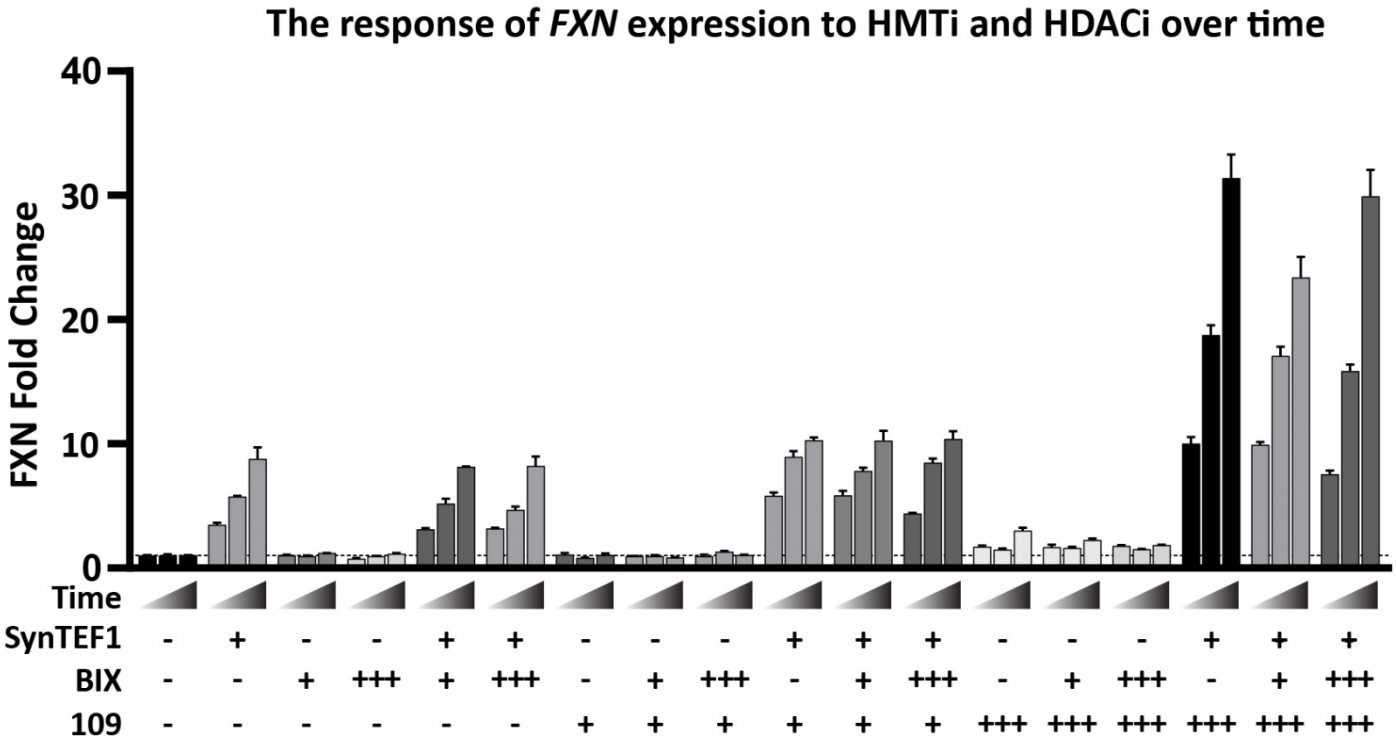

A
(GAA) Repeat
DMSO
109
SynTEF1
SynTEF1 + 109
FXN
Chr 9: 69,035,000
Chr 9: 69,080,000
B
(GAA) Repeat
DMSO
109
SynTEF1
SynTEF1 + 109
FXN
Chr 9: 69,035,500
Chr 9: 69,038,000
C
(GAA) Repeat
DMSO
109
SynTEF1
SynTEF1 + 109
FXN
Chr 9: 69,035,000
Chr 9: 69,080,000

Extended Data Figure 2.7 | *FXN* RNAseq peaks increase in response to small molecule treatment

a, RNAseq trace at *FXN* of FRDA GM15850 cells treated with DMSO vehicle (black), 50 μM 109 (green), 1 μM SynTEF1 (red), or cotreated with SynTEF1 and 109 (purple). (GAA) repeat is in blue. The level of transcription in vehicle treated cells is very low which does not change after treatment with 109. SynTEF1 increases RNA levels at introns 2 – 5, which is further increased with the co-treatment of SynTEF1 + 109.

b, (GAA) repeat region of *FXN* showing partially transcribed intronic region 5' of the (GAA) repeat region. Both 109 and SynTEF1 show low levels of transcription in the first intron 5' of the (GAA) repeat. Co-treatment with both SynTEF1 and 109 increases transcription in the first intron synergistically.

c, Antisense transcript levels at *FXN*. There is no observed antisense transcript present in vehicle, 109, or SynTEF1 treated cells. There are low levels of antisense transcript produced in response to the co-treatment with SynTEF1 + 109 primarily in the first intron, 4^{th} and 5^{th} exon, and the 3' UTR.


The response of *FXN* expression to HMTi and HDACi over time
FXN Fold Change
40
30
20
10
0
Time
SynTEF1 - + - - + + - - - + + + - - - + + +
BIX - - + +++ + +++ - + +++ - + +++ - + +++ - + +++
109 - - - - - - + + + + + + +++ +++ +++ +++ +++ +++

Extended Data Figure 2.8 | HDACi but not HMTi work synergistically with SynTEF1 to regulate *FXN* expression over time

FRDA GM15850 lymphoblasts were treated with the HMT inhibitor BIX at 0.1 µM and 0.5 µM, the HDACi 109 at 10 µM and 50 µM, or SynTEF1 at 1 µM for 24, 48, or 72 hours. Error bars are SEM. SynTEF1 shows the expected increase in *FXN* expression which continues to increase at each timepoint. BIX alone was unable to activate *FXN* expression and did not work synergistically with SynTEF1 or 109. 109 showed a ~2-fold increase after 72 hours at 50 µM. The cotreatment of SynTEF1 + 10 µM 109 showed synergism at all three timepoints irrespective of BIX treatment. SynTEF1 + 50 µM 109 showed synergistic activation at each timepoint.

2.8 Bibliography

1. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
2. Gershman, A., et al., *Epigenetic patterns in a complete human genome.* Science, 2022. **376**(6588): p. eabj5089.
3. Allis, C.D. and T. Jenuwein, *The molecular hallmarks of epigenetic control.* Nat Rev Genet, 2016. **17**(8): p. 487-500.
4. Allshire, R.C. and H.D. Madhani, *Ten principles of heterochromatin formation and function.* Nat Rev Mol Cell Biol, 2018. **19**(4): p. 229-244.
5. Nicetto, D., et al., *H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification.* Science, 2019. **363**(6424): p. 294-297.
6. Matsumura, Y., et al., *H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation.* Molecular Cell, 2015. **60**(4): p. 584-596.
7. Campuzano, V., et al., *Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion.* Science, 1996. **271**(5254): p. 1423-7.
8. Saveliev, A., et al., *DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing.* Nature, 2003. **422**(6934): p. 909-13.
9. Soragni, E., et al., *Epigenetic therapy for Friedreich ataxia.* Ann Neurol, 2014. **76**(4): p. 489-508.
10. Kim, E., M. Napierala, and S.Y. Dent, *Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia.* Nucleic Acids Res, 2011. **39**(19): p. 8366-77.

11. Herman, D., et al., *Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia.* Nat Chem Biol, 2006. **2**(10): p. 551-8.

12. Punga, T. and M. Buhler, *Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation.* EMBO Mol Med, 2010. **2**(4): p. 120-9.

13. Yandim, C., T. Natisvili, and R. Festenstein, *Gene regulation and epigenetics in Friedreich's ataxia.* J Neurochem, 2013. **126 Suppl 1**: p. 21-42.

14. Sherzai, M., et al., *HMTase Inhibitors as a Potential Epigenetic-Based Therapeutic Approach for Friedreich's Ataxia.* Front Genet, 2020. **11**: p. 584.

15. Rai, M., et al., *Two new pimelic diphenylamide HDAC inhibitors induce sustained frataxin upregulation in cells from Friedreich's ataxia patients and in a mouse model.* PLoS One, 2010. **5**(1): p. e8825.

16. Erwin, G.S., et al., *Synthetic transcription elongation factors license transcription across repressive chromatin.* Science, 2017. **358**(6370): p. 1617-1622.

17. Burnett, R., et al., *DNA sequence-specific polyamides alleviate transcription inhibition associated with long GAA.TTC repeats in Friedreich's ataxia.* Proc Natl Acad Sci U S A, 2006. **103**(31): p. 11497-502.

18. Erwin, G.S., et al., *Synthetic genome readers target clustered binding sites across diverse chromatin states.* Proc Natl Acad Sci U S A, 2016. **113**(47): p. E7418-e7427.

19. Filippakopoulos, P., et al., *Selective inhibition of BET bromodomains.* Nature, 2010. **468**(7327): p. 1067-73.

20. Jang, M.K., et al., *The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription.* Mol Cell, 2005. **19**(4): p. 523-34.

21. Vakoc, C.R., et al., *Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin.* Mol Cell, 2005. **19**(3): p. 381-91.

22. Larson, A.G., et al., *Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin.* Nature, 2017. **547**(7662): p. 236-240.

23. Strom, A.R., et al., *Phase separation drives heterochromatin domain formation.* Nature, 2017. **547**(7662): p. 241-245.

24. Bannister, A.J., et al., *Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain.* Nature, 2001. **410**(6824): p. 120-4.

25. Lachner, M., et al., *Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.* Nature, 2001. **410**(6824): p. 116-20.

26. Nakayama, J., et al., *Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly.* Science, 2001. **292**(5514): p. 110-3.

27. De Biase, I., et al., *Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription.* PLoS One, 2009. **4**(11): p. e7914.

28. Blattes, R., et al., *Displacement of D1, HP1 and topoisomerase II from satellite heterochromatin by a specific polyamide.* EMBO J, 2006. **25**(11): p. 2397-408.

29. Zaidan, N.Z. and R. Sridharan, *HP1γ regulates H3K36 methylation and pluripotency in embryonic stem cells.* Nucleic Acids Research, 2020. **48**(22): p. 12660-12674.

30. Silva, A.M., et al., *Expanded GAA repeats impair FXN gene expression and reposition the FXN locus to the nuclear lamina in single cells.* Hum Mol Genet, 2015. **24**(12): p. 3457-71.

31. Gottesfeld, J.M., *Molecular Mechanisms and Therapeutics for the GAA·TTC Expansion Disease Friedreich Ataxia.* Neurotherapeutics, 2019. **16**(4): p. 1032-1049.

32. Zhang, S., M. Napierala, and J.S. Napierala, *Therapeutic Prospects for Friedreich's Ataxia.* Trends Pharmacol Sci, 2019. **40**(4): p. 229-233.

33. Delatycki, M.B. and S.I. Bidichandani, *Friedreich ataxia- pathogenesis and implications for therapies.* Neurobiol Dis, 2019. **132**: p. 104606.

2.8 Materials and Methods

Cell Culture and Treatment

GM15850 cells were obtained from the Coriell Institute and cultured under recommended conditions. Cells were cultured in RPMI (GIBCO) containing 15% FBS (GIBCO) and an antibiotic/antimycotic containing 100 U/mL penicillin, 100 U/mL streptomycin, and 0.25 U/mL Amphotericin B (GIBCO). Small molecules were dissolved in DMSO and were added to fresh culture in media.

Peptide Synthesis and Protein Production

The H3 peptide sequence ARTKQTARKSTGGKAPRKQLATKA labeled with a 3' 5-FAM was produced at the St. Jude Children's Research Hospital Peptide Production Facility using a Symphony X solid phase peptide synthesizer. The lysine and serine modifications were introduced during peptide synthesis using pre-modified amino acids. Peptide purity was validated using 220 nm and 492 nm HPLC and verified using mass spectrometry.

HP1α, HP1β, and HP1γ proteins were generated by the St. Jude Children's Research Hospital Protein Production Facility. Briefly, BL21(DE3) Rosetta *E. coli* cells were transformed with pBH4-6His-TEV-HP1α plasmid and selected using carbenicillin and kanamycin. Single colonies were grown and induced with IPTG. Pellets were lysed in lysis buffer (300 mM NaCl, 1xPBS, 10% glycerol, 7.5 mM imidazole pH 8) in the presence of PMSF, pepstatin A, aprotinin, and leupeptin. Supernatant from lysate was incubated with cobalt resin for at least 1 hour and then washed in

a gravity column. Protein was eluted in 20 mM HEPES pH 7.5, 150 mM KCl, 400 mM imidazole pH 7.5. 50 μL TEV protease at 2 mg/mL was used to cleave the tag, and the protein was then dialyzed overnight in 20 mM HEPES pH 7.2, 75 mM KCl, and 1 mM DTT. The protein was then passed through a MonoQ column using standard protocols and eluted with a 200 mM-1 M KCl gradient and collected. Protein was then passed through a 0.22 μm centrifugal filter and purified further on an S75 size exclusion column. Protein was concentrated using 10 kDa spin column concentrators as needed, and concentrations were verified using UV absorbance at 280 nm (Nanodrop).

A synthetic gene for BRD4s, codon-optimized for *E. coli*, was ordered from Genscript in a pET28a vector. A hexa-histidine tag followed by a TEV protease cleavage site was placed at the N-terminus of the open reading frame. The BRD4s plasmid was transformed into BL21-RIPL cells, cultures were grown in LB medium, and expression initiated by addition of IPTG. Cells were lysed in 30 mM imidazole pH 7.8, 1 M NaCl with a sonicator. The clarified lysate was loaded onto a 5 mL Fast Flow Chelating Sepharose gravity column and washed with the resuspension buffer. Protein was eluted in 300 mM imidazole pH 7.8, 300 mM NaCl and subsequently diluted three-fold before loading onto a HiTrap Heparin column for the removal of bound nucleic acid. The protein was eluted with a gradient of NaCl in 20 mM HEPES pH 7.5 and concentrated to 100 μM. Protein was then either dialyzed into 20 mM Tris pH 7.8, 150 mM NaCl, 5 mM DTT and flash frozen and stored at -80 °C.

Fluorescence Polarization Binding Assays

Fluorescence polarization binding assays were performed using 20 µL FP buffer (0.02% NP-40, 150 mM KCl, 20 mM HEPES pH 7.5, 1 mM DTT) in black flat bottom plates (Corning 3821BC). HP1 proteins (20 mM HEPES pH 7.5, 200 mM KCl, 10% glycerol, 2 mM DTT) were serially diluted into the plate. 30 nM final concentration of H3 peptides were added to each well with an Echo Liquid Handler. Plates were agitated on shaker for 5 minutes and incubated at RT for 90 minutes. Fluorescence polarization of all samples was determined using fluorescence excitation at 485nm and emission at 520nm.

Chromatin Immunoprecipitation

$2.5x10^7$ cells were used for each ChIP sample. Covaris truChIP Chromatin Shearing Kits were used for fixation and nuclear isolation using the standard protocol. Immunoprecipitation was performed as follows: 35 µL/sample of Protein A/G (Pierce 88803) beads were washed 2x with and resuspended in ChIP Lysis Buffer (50 mM Tris-HCl, 10 mM EDTA, 0.5% Empigen BB, 1% SDS) and used to preclear samples for 1 hr. 8 µL of sample was added to 292µL Elution Buffer (1% SDS, 100 mM NaHCO3) and stored as input at -20°C. 800 µL of sample was added to 2 mL of IP Buffer (2 mM EDTA, 150 mM NaCL, 20 mM Tris-HCL pH 8.0, 1% Triton X-100). 6µg antibody was added to each sample, and they were rotated at 4°C overnight. Antibodies used can be found in table S1. 100 µL magnetic beads per sample were washed 2x with equal volumes of IP buffer and then added to each sample and samples incubated for 1 hour. Samples were centrifuged at 13k RPM for 1 minute and then placed on magnet. Supernatant was discarded and beads were washed with 1 mL each of Wash Buffer 1 (2 mM EDTA, 20 mM Tris-HCl pH 8.0), 0.1% SDS, 1%

Triton X-100, 150 mM NaCl), Wash Buffer 2 (2mM EDTA, 20 mM Tris-HCl pH 8.0, 0.1% SDS, 1% Triton X-100, 500 mM NaCl), Wash Buffer 3 (1 mM EDTA, 10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1% Deoxycholate, 1% NP-40), and 2x with 1 mL TE Buffer (10 mM Tris-HCl, 1 mM EDTA). Samples were centrifuged at 13k RPM for 30s and residual TE buffer was removed. Magnetic beads were resuspended in 300 µL Elution Buffer. 12 µL 5 M NaCl was added to both samples and inputs and placed at 65°C for 18 hours. Samples were then spun at max speed at RT and supernatant was transferred to new tubes. RNase A (Fisher EN0531) was added at 0.2 µg/µL and samples incubated at 37 °C for 1 hour. Proteinase K (Fisher 25530049) was added at 0.2 ug/µL and samples incubated at 55 °C for 1 hour. DNA was isolated using Qiagen DNA cleanup buffers PB (19066) and PE (19065). Briefly 1.2 mL buffer PB was added to each sample and samples were added to DNA columns (Fisher NC0066803) and centrifuged for 60 seconds. Columns were washed 2x with 0.75 mL buffer PE and eluted in 50 µL nuclease-free water. Samples were sequenced using Next Generation Sequencing by NovaSeq6000 by the St. Jude Children's Research Hospital Genome Sequencing Facility.

ChIP-seq Data Analysis

All ChIP-seq data were sequenced for this study and deposited under accession number GSE196619 in the GEO database. Collection of all sequencing reads were processed with the Trim Galore tool (available on-line at https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), removing all potential adapter sequences and quality trimming reads with cutadapt using Q20 quality score cutoff[1].

Next, reads were aligned to the human reference genome GRCh38.p12 using bwa (v0.7.17-r1198)[2] and the output was converted to BAM format with samtools (v1.2)[3], followed by identification of duplicated reads with bamsormadup tool from biobambam2 program (v2.0.87)[4]. Subsequently, the SPP tool (v1.11)[5] was used to estimate fragment size with the relative strand cross-correlation analysis; and uniquely mapped reads were extracted from BAM files with samtools and extended with bedtools (v2.24.0)[6], using the fragment size value precalculated with cross-correlation analysis. Subsequently, the MACS2 program[7] was used to call peaks in narrow mode, with -nomodel -q 0.05 flags (high confidence peaks). In parallel, peaks were also called with more relaxed criteria, setting the -q flag to 0.5, which are here further referred to as low confidence peaks. The reproducible peaks of biological replicates were identified as those which either in both replicates had overlapping high confidence peaks or those, which in one replicate had a high confidence peak, which was supported by a low-confidence peak in the second replicate. Finally, reproducible peaks were annotated with genes if the peak overlapped the with gene promoter, defined as transcription start site (TSS) ± 2000 bp, and based on the reference annotation from Gencode[8]. For the purposes of visualization, the mapped reads' densities were converted to BigWig format and normalized to 15 million non-duplicated mapped reads. Next, whenever applicable, the average signal between biological replicates was calculated.

Differential peak binding analysis included calculation of the fragment counts per reproducible peaks based on bedtools (v2.24.0)[6], combined with in-house scripts, and was followed by limma-voom approach[9,10] to differentially binding peaks. Different levels of stringency were used to classify the peaks as differentially binding, including $p<0.05$.

Velocity Plots

To visualize the differences between two conditions, a new approach that utilize the vector (aka. Quiver) plots, was introduced, which for this study, covered region at coordinates chr9:69,035,000-69,039,600. These plots, which are here referred to as velocity plots, emphasize the differences of the shape and intensity of the enrichment signal between two conditions. Velocity plots are generated for a predefined genomic region of interest, and step-by-step guidelines on how the velocity plots are generated were depicted in the supplementary figure 2. The signal enrichment from two toy example conditions, both of which were pre-normalized to the same sequencing depth, and which both use the same bin size (in our case genome was separated in 50 bp long bins), are visualized on the figure 2A. The input format was bedGraph. Note that the values represented by entries of bedGraph files for each condition, which are 1-dimentional, were then converted to 2-Dimentional matrix, which dimensions depend on the number of bins that span the genomic region of interest, and the user-specified number of desired layers (in our case equal to 5 layers). Next, the algorithm identified the maximum enrichment value from both conditions, which was round up to integer. This max integer value was then divided by the number of layers, which was subsequently used to calculate m - the max enrichment value per layer. Next, for each bin, the algorithm calculated how many m values could be covered by the enrichment value from that bin, which in the final matrix were assigned the value of 1, and what was the ratio of the remaining signal to the m value (figure 2B). Finally, the values identified were ordered in ascending order, and filled in the values for the layers 1-5 of the corresponding bins. On the example of 12th bin of the condition 1 (figure 2B), which enrichment value is equal to 4.3, the m value is equal to 1, so the 4.3 = 4*m + 0.3; and the 0.3/m

= 0.3; therefore, the values placed in layers 1-5 were 0.3, 1, 1, 1, 1, as visualized on the figure 2B. Next, in order to compare the condition 1 to condition 2, the 2-D matrix of the condition 2 has to be subtracted from 2-D matrix of the condition 1 (figure 2C), and vice versa, to compare condition 2 to condition 1, 2-D matrix of the latter has to be subtracted from the 2-D matrix of the former. Next, the algorithm calculated the values required to plot regular quiver plot from matplotlib Python package (description available on-line at: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.quiver.html), that is X and Y coordinates, which correspond with the Bins and layers from the processed 2-D subtraction matrix. Moreover, the quiver plotting function requires the values of the horizontal and vertical arrow tilt to position the vector on the plot. Both were calculated by traverse through the 2-D subtraction matrix, with calculating the tilts using the weight pattern shown on figure 2D. I.e. in order to calculate the vertical arrow tilt, the value from current cell, as well as the cell corresponding to the previous layer, were taken under consideration (result values for vertical arrow tilt are shown on figure 2F); however, to calculate the horizontal arrow tilt, all the adjacent layers' values were taken under consideration, with the weights decreasing the further the distance from the cell of origin (for the example matrix shown in figure 2C, the horizontal tilt values are shown in figure 2E). Finally, the values for the vertical arrow tilt, which range between -1 and 1, were also used to assign color from the blue-white-red color map. The resulting vector-enrichment plots for the toy example, for both the condition 1 over condition 2, and condition 2 over condition 1, are shown on the supplementary figure 2G.

siRNA, RNA Isolation, and qPCR

On-target plus siRNA was purchased from Dharmacon and Dharmafect I was used for siRNA transfection and knockdown using the recommended protocol. RNA was isolated using Qiagen RNeasy isolation kit (74106) and eluted in 50µL nuclease free water. RNA concentration was determined using nanodrop and diluted to 100ng/uL in nuclease free water. 100ng of RNA was used to generate cDNA using iScript reverse transcription cDNA synthesis kit (Bio-Rad 1708891). The resulting cDNA concentration was determined using nanodrop and diluted to 100ng/uL. Quantitative polymerase chain reaction was performed using SYBR Green SsoAdvanced (Bio-Rad 1725275) with 100ng cDNA per sample. qPCR was performed on ABI7900.

RNA-seq data analysis

Raw RNA-seq reads were quality filtered and trimmed with Trim Galore tool (available on-line at https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Reads were then aligned to the human reference genome (hg38 / GRCh38.p12) using STAR[11]. Subsequently, based on the reference annotation from Gencode (Release 31)[8], read counts per gene were calculated using RSEM[12], including only level 1 and 2 protein-coding genes, with at least 10 reads per condition. Next, to identify differentially expressed genes (DEGs), the remaining genes were processed using limma-voom approach[9,10]. Based on the values of log2(fold-change), p-value and false discovery rate (FDR), genes were then classified into various categories of differential expression representing different levels of stringency: $|\log2(FC)| > 1$ and $FDR < 0.05$.

RNA/DNA FISH

GM15850 and GM15851 cells were cultured according to Coriell's protocol using RPMI +15% FBS, seeded at $2.5x10^5$ cells/ml, and treated for 24 hours with either 0.1% DMSO or 1uM Syn-TEF1 in 0.1% DMSO prior to collection. $2.5x10^5$ cells/sample were used for PCR analysis of *FXN* and *GAPDH* expression. $1\text{-}5x10^6$ cells/treated were used for DNA/RNA FISH using an *FXN* probe from WI2-1857D19 fosmid. This experiment had 3 biological replicates with 2-4 technical replicates. Samples were sequentially stained and imaged for RNA and then DNA FISH using DAPI, FITC, and TRITC channels for both acquisitions. DNA FITC and RNA TRITC channels contain peaks of interest; DNA TRITC and RNA FITC channels contain background/bleed through. DNA and RNA z-stacks were trimmed to the same length using points of interest to determine z-offset and then merged using DAPI stains to align samples in Nikon Elements. Peak picking of RNA (RNA TRITC) and DNA (DNA FITC) spots in merged z-stacks were conducted by generating peak masks by interactively thresholding a Laplacian of Gaussian (sigma: 0.15) and peak positions were defined as the center of gravity of the masked image pixels. At each peak, the intensity value of each image channel was read (following a gaussian blur with sigma 0.15). Additionally, the closest neighbors across all image channels were determined using the KDTree algorithm. The thresholds used were: 1200 (DNA FITC), 600 (DNA TRITC and RNA FITC), 1000 (RNA TRITC). Peak intensity values and xyz location were filtered first to see if they were nuclear (DAPI > 200) and far from a bleed through signal (DNA FITC peak further than 0.75um from any DNA TRITC peak and RNA TRITC peak further than 0.75um from any RNA FITC peak). DNA spots were counted for any DNA FITC peak that met the above filters. A RNA spot was counted if the RNA TRITC peak met the above filters and was within 0.75µm of a DNA spot.

Image Analysis

FISH images were generated using a merged z-stack with LUTS adjustments made with ImageJ. Distance to nuclear periphery analysis was done using Imaris Image Analysis Software using DAPI to generate a 3D ROI representing the nucleus and measuring the distance of the DNA FISH puncta to the closest edge of the ROI. Cells containing more than 2 DNA FISH signals were excluded.

601 DNA FWD

CTGGAGAATCCCGGTCTGCAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTA
CGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATA
CATCCTGT

601 DNA REV

ACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGG
GGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTGCAGACCG
GGATTCTCCAG

GAA DNA FWD

GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA

AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG

AAGAAGAAGAAGAAGAA

GAA DNA REV

TTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT

CTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC

Table S1.

Target	Antibody
H3K4me3	Abcam Ab8540
H3K9me3	Abcam Ab8898
HP1α	Cell Signaling 2616S
HP1β	Cell Signaling 8676S
HP1γ	Millipore 05-690

Table S2.

Target	Sequence
FXN Forward	CAGAGGAAACGCTGGACTCT
FXN Reverse	AGCCAGATTTGCTTGTTTGG
GAPDH Forward	CTGAGCTCATTTCCTGGTATGA
GAPDH Reverse	CTTCCTCTTGTGCTCTTGCTG
HDAC1 Forward	GCTGGCAAAGGCAAGTATTATG
HDAC1 Reverse	CTAGGCTGGAACATCTCCATTAC
HDAC2 Forward	GGAGAAGGAGGTCGAAGAAATG
HDAC2 Reverse	ACCACTGTTGTCCTTGGATTTA
HDAC3 Forward	TGATCGATTGGGCTGCTTTA
HDAC3 Reverse	CAGCACGAGTAGAGGGATATTG
HP1β Forward	GGAGAGCTCATGTTCCTGATG
HP1β Reverse	CGTCAGCCTTTCCTCATAGAAG

HP1γ Forward	GGCCTCCAACAAAACTACATTG
HP1γ Reverse	TCCACTTTCCCATTCACTACAC
HP1α Forward	TTGCCCTGAGCTAATTTCTGAAT
HP1α Reverse	GATGTCATCGGCACTGTTTGA
FXN A Forward	CCCCACATACCCAACTGCTG
FXN A Reverse	GCCCGCCGCTTCTAAAATTC
FXN B Forward	AAACTGACCCGACCTTTATTCCA
FXN B Reverse	GGAATCCCCCAAGGTCACA
FXN C Forward	GAAACCCAAAGAATGGCTGTG
FXN C Reverse	TTCCCTCCTCGTGAAACACC
FXN D Forward	CTGGAAAAATAGGCAAGTGTGG
FXN D Reverse	CAGGGGTGGAAGCCCAATAC

Acknowledgements: We thank Geoff Neale and Scott Olsen of the Hartwell Center Sequencing Facility, Patrick Rodrigues of the Hartwell Center Macromolecular Synthesis, Youming Shao and Richard Heath of the Protein Production Facility and Gang Wu for advice, materials, and services that enabled this study. HP1 plasmids and purification protocols were generously provided by Geeta Narlikar and Emily Wong. We thank Ines Chen and members of the Ansari lab for invaluable discussions, and Laura Vanderploeg for the artwork. Funding for this work was provided by the NIH (NS108376 and P30 CA021765), NSF (CEE-EFRI: #1933402), Friedreich's ataxia research alliance (FARA) and the American Lebanese Syrian Associated Charities at St. Jude (ALSAC) to AZA and TM.

Author contributions: Author contributions: CJB, TM, WR, TC, WL and AZA designed research; CJB, WR, MC, SR, MK, MV performed research; AM contributed new reagents; CJB, WR, WR, SR, BH, MC, KK, SJK, BX, TM and AZA analyzed the data; and CJB with AZA wrote the paper.

Competing interests

A.Z.A. is co-founder and scientific advisor of Design Therapeutics Inc, the sole member of VistaMotif, LLC and founder of the educational US nonprofit WINStep Forward (501 (c)(3)).

Materials & Correspondence

Please contact Aseem Ansari for material requests.

Methods References

1 MARTIN, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal,* **17**, 10-12, doi:https://doi.org/10.14806/ej.17.1.200. (2011).

2 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

3 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

4 Tischler, G. & Leonard, S. biobambam: tools for read pair collation based algorithms on BAM files. *Source Code for Biology and Medicine* **9**, 13, doi:10.1186/1751-0473-9-13 (2014).

5 Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature biotechnology* **26**, 1351-1359, doi:10.1038/nbt.1508 (2008).

6 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

7 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

8 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**, D766-d773, doi:10.1093/nar/gky955 (2019).

9 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).

10 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29, doi:10.1186/gb-2014-15-2-r29 (2014).

11 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

12 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).

Chapter 3 | SynTEF1 facilitates BRD4 entry into HP1α phase separated condensates

Portions of this chapter has been adapted from “Partitioning of BRD4 into HP1 condensates overrides persistent repressive chromatin at microsatellite repeats in Friedreich’s Ataxia”. Christopher J. Brandon, Wojciech Rosikiewicz, Matthew J. Cuneo, Sam Rider, Ashraf Mohammed, Marcus Valentine, Mangesh Kaulage, Sandra J. Kietlinska, Burkhard Hoeckendorf, Wenwei Lin, Khaled Khairy, Taosheng Chen, Beisi Xu, Tanja Mittag, Aseem Z. Ansari. *Submitted Nature 2022*

Author contributions: CJB, TM, WR, MC, TC, WL and AZA designed research; CJB, WR, MC, SR, MK, MV performed research; AM contributed new reagents; CJB, WR, WR, SR, BH, MC, KK, SJK, BX, TM and AZA analyzed the data; and CJB with AZA wrote the paper.

Acknowledgements

I was taught how to perform phase separation experiments by Matt Cuneo and Tanja Mittag, who both contributed to this work. I would like to thank them for their guidance and assistance in these experiments, the data they contributed, and the reagents they shared. I would also like to thank Aseem Ansari for his constant guidance, and Sandra Kietlinska for her helpful input and discussions.

3.1 Introduction

Throughout the life of a cell the proper partitioning of proteins to specific subcellular compartments is required to maintain homeostatic balance. Correct spatial localization is necessary for protein function, and mutations that changes protein trafficking results in a bevy of diseases from Parkinson's to Type-2 Diabetes to various types of cancer [1]. Historically these compartments were either membrane bound organelles separated from the cellular milieu by a phospholipid bilayer, or the membrane-less nucleolus which maintains a liquid-like state of condensed DNA and proteins. Phospholipid bilayers provide a physical barrier to protein and small molecule transport which restricts access to everything that cannot be trafficked or won't freely diffuse across. The liquid—like state found in the nucleolus instead clusters specific proteins, RNA, and DNA to the same region to facilitate ribosome biosynthesis, but does so without physically separating the region with a membranous barrier. Instead, these proteins cluster in a liquid—like state and form liquid—liquid phase separated (LLPS) condensates, which exchange protein—water interactions with protein—protein interactions [2]. This results in two different phases, a protein rich dense phase and a protein poor dilute phase.

At their most fundamental state, proteins are simply linear chains of amino-acids. Through entropically favorable intramolecular interactions proteins form stable secondary and tertiary structures which restricts solvent access to hydrophobic regions, resulting in a 3-dimensional protein structure. A protein can be subdivided into domains which either form stable secondary structures or are intrinsically disordered in solution. Folded protein domains that form multivalent interactions with other molecules can facilitate protein liquid—liquid phase

separation [3]. Similarly, intrinsically disordered regions drive phase separation through weak multivalent interactions [2]. Since the central mechanism for both structured and unstructured protein LLPS is multivalent interactions, phase separated condensates frequently incorporate nucleic acids or other macromolecular scaffolds which provide multiple potential interaction sites [4].

Sub-cellular compartmentalization of proteins into LLPS condensates provides the cell a mechanism to locally concentrate proteins with similar functions into a membraneless organelle at a specific locus. This results in an increase in the local concentration of a small group of interacting proteins which results in increased function and cooperativity [4]. Biological processes generally require the collective interaction of interdependent proteins which we are learning is frequently catalyzed through LLPS. Condensate formation has been implicated in diverse biological processes such as heterochromatin formation, active transcription, RNA degradation and splicing, DNA damage response, and RNAi processing [4-8].

SynTEF1 recruits BRD4 to the *FXN* locus in FRDA patient cells, which further recruits other components necessary for active transcription such as pTEFb and CDK9 [9]. We have shown in this work that this occurs despite the establishment of functional epigenetic repression. Considering LLPS has been involved in both heterochromatin formation and active transcription, SynTEF1 treatment must either regulate the interaction of these LLPS condensates, or bypass it and dissolve them entirely.

3.2 HP1 phase separation

The placement of H3K9me3 at gene poor and pericentromeric repeat regions results in the recruitment of the HP1 family of proteins and the subsequent condensation of heterochromatin, as has been discussed previously. HP1 will also stably associate with free DNA and RNA through direct interactions with two lysine/arginine rich motifs in the unstructured hinge region [10, 11]. The DNA wrapped and H3K9me3 modified nucleosomes provides a large, multivalent substrate not unlike those observed to catalyze LLPS.

Notably the multiple complex interactions between HP1, DNA, and modified histones are not required for LLPS. HP1α will form liquid droplets on its own when phosphorylated on the N-terminal domain or hinge region, and is maintained highly phosphorylated in the cell [8, 12, 13]. This provides a mechanism for HP1α to be stabilized in a region of heterochromatin without directly binding DNA or nucleosomes, which alone results in an increased unbound local concentration at regions of DNA that HP1 is bound. Despite the sequence similarity between the HP1 paralogues, these observations are specific to HP1α; HP1β and HP1γ do not form higher order oligomers on their own [8].

In addition to the independent phase separation displayed by HP1α, it will also phase separate with DNA sequences as short as 147 bp and as long as 48.5 kb [8]. The length of the available DNA affects LLPS, where longer DNA sequences allow HP1α to phase separate at lower concentrations. HP1γ will also associate with longer DNA sequences and form LLPS condensates [13]. HP1β instead prefers to condense on H3K9me3 modified mono-nucleosomes [14]. These paralogue specific differences may help to account for the differences that we observed at the *FXN* gene. HP1α and HP1β were recruited after SynTEF1 treatment around the GAA repeat

region, where HP1γ was located primarily at the TSS. The disparity between these proteins may arise from the differences in DNA binding and their individual phase separation properties.

3.3 BRD4 phase separation

The assembly of the various transcription factors at the enhancer region that are necessary for active transcription occurs in euchromatic regions generally marked with H3K4me1 and HeK27Ac as discussed previously [15]. This includes BRD4 recognition of the H3K27Ac modification by its bromodomains, and its association with MED1. Multiple enhancers can cluster together to form super-enhancers (SE) with high densities of transcriptional machinery coalescing in one area. These regions are greatly enriched for MED1 and H3K27Ac at transcribed genes crucial for cellular identity [16, 17]. SEs bound by MED1 and BRD4 display liquid-like properties, and the intrinsically disordered regions (IDR) from BRD4 and MED1 phase separate together [18]. The MED1 and BRD4 phase separated enhancers concentrate the necessary general transcription factors and other components into a small, nuclear, membraneless organelle, specifically targeting this machinery to active genes. These MED1 bound enhancers also from condensates with RNA Polymerase which further supports a direct relationship between LLPS and active transcription [19]. BRD4 recruitment to these loci through its interaction with H3K27ac epigenetic modifications is paramount. BRD4 and MED1 LLPS condensates are dispersed by JQ1 treatment, which binds to the BRD4 bromodomains and directly inhibits its binding to acetylated chromatin [19]. Additional *in vitro* studies show unmodified nucleosomal arrays will form LLPS condensates, which are dispersed when acetylated at H3K27 [6]. When BRD4 is added to the system the LLPS condensates reform, pointing to the importance of BRD4

to the system. This suggests a biphasic system where the BRD4 binding of H3K27Ac nucleates the recruitment of MED1 and RNA Polymerase assembly at enhancers and super-enhancers.

BRD4 is expressed as both a long and a short isoform. The short isoform lacks a large ~600 amino acid disordered region on the C-terminus, but preferentially forms liquid like phase separated condensates *in vivo* and *in vitro* [5]. Similar to HP1α, BRD4s will also bind and phase separate with dsDNA through interactions with an extremely basic, lysine rich region [5]. But while HP1α phase separates more readily when phosphorylated, BRD4s fails to form LLPS condensates *in vitro* or in cell lysates after phosphorylation. The importance of BRD4 phosphorylation in transcriptional regulation is becoming more apparent, with its hyperphosphorylation correlating with poor survival in triple-negative breast cancer [20, 21]. Interestingly both BRD4s and HP1 are substrates for casein kinase 2 (CK2) which may provide a transitional switch from active transcription to repression at co-occupied regions [5, 12].

3.4 BRD4s and HP1α coacervate into mixed phase separated condensates

H3K9me3 and HP1α, HP1β, and HP1γ are all enriched at *FXN* by SynTEF1, and these transcriptional repressors coexist with the recruited transcriptional coactivator BRD4. Critically, HP1 recruitment to *FXN* is functional, and both the induced increase in histone acetylation by 109 as well as siRNA knockdown of HP1α and HP1γ further amplify the ability of SynTEF1 to activate transcription. These findings raised a conundrum: how does SynTEF1 binding, BRD4 recruitment, and Pol II transcription through GAA repeats occur without reduction in HP1 levels or erasure of H3K9me3 marks? We hypothesized that HP1α forms repressive condensates at GAA repeat

expansions, which SynTEF1 can enter and grant access to BRD4 and Pol II without dispersing the HP1a condensate. Due to its rapid on/off DNA binding kinetics [13], HP1α could locally release and re-engage GAA repeats allowing passage to an elongating Pol II.

To test this possibility, we first examined whether HP1α forms condensates with a 147 base pair DNA fragment bearing 49 GAA repeats which can adopt unusual conformations [22]. When compared with the well-characterized 601 DNA sequence of identical length [13], HP1α phase separated with GAA repeats at comparable saturation concentrations and formed condensates of similar dimensions (Figs. 3.1 a-c). Next, to examine if SynTEF1 can access cognate DNA binding sites in HP1a condensates, we synthesized a GAA-binding polyamide (PA1) conjugated to JF646, a fluorophore that emits at 646 nm (Fig. 3.1 d and Extended Data Figs. 1b-c). Consistent with its specificity for GAA repeats [23], PA1-JF646 enriches in the HP1α condensates bearing 49 contiguous GAA-repeats but not in condensates formed with identically sized 601 DNA, which lack PA1/SynTEF1 binding sites (Figs. 3e-g).

While other polyamides have been reported to displace HP1α [24], PA1 or PA1-JF646 do not disperse the GAA-HP1α condensates even at saturating concentrations (Fig. 3.1 e-g). However, it was unclear if SynTEF1 would partition into HP1α condensates because JQ1, a key component of SynTEF1, was reported to not enrich in HP1α condensates [25]. Moreover, HP1α and BRD4 form functionally and physically distinct condensates and their ability to co-condense at physiological concentrations was unexpected [2, 18, 26, 27]. Remarkably, not only does SynTEF1 enable JQ1 to enter HP1α condensates it also facilitates the partitioning of BRD4 into these repressive condensates at physiological concentrations (Figs. 3i and 3j). Intriguingly, at higher concentrations (>250nM), BRD4 co-condensates with HP1α without the assistance of

SynTEF1 (Fig. 3j). When incubated with pre-formed HP1α condensates, SynTEF1 recruits BRD4 within seconds (Figs. 3k-l and Extended Data Fig. 3.1e). This unexpected result points to a more generally employed mechanism by which genes embedded in repressive heterochromatin might permit transient access to transcriptional machinery.

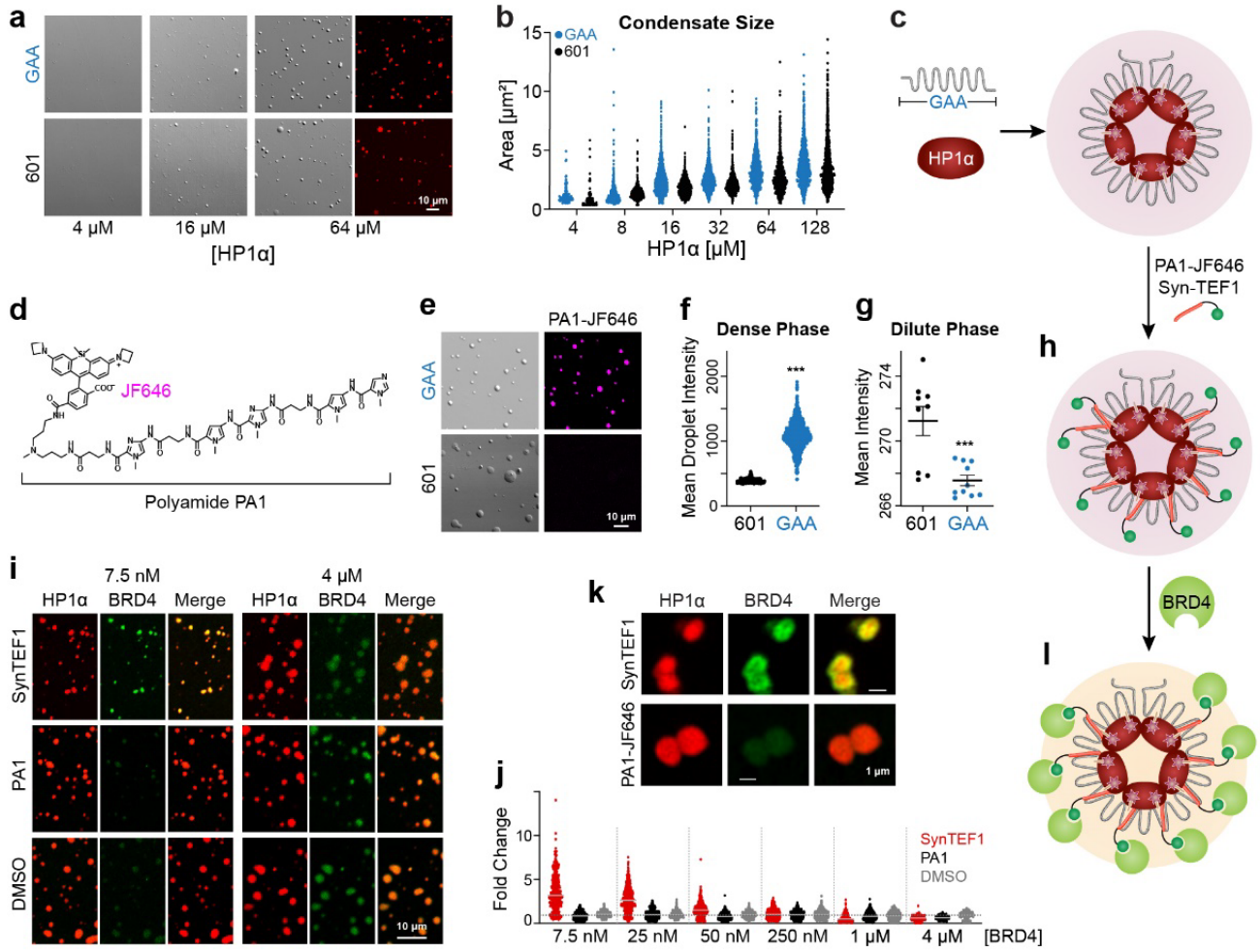

a
GAA
601
4 μM
16 μM
64 μM
[HP1α]
10 μm
b
Condensate Size
GAA
601
Area [μm²]
HP1α [μM]
c
GAA
HP1α
PA1-JF646
Syn-TEF1
d
JF646
Polyamide PA1
e
PA1-JF646
f
Dense Phase
Mean Droplet Intensity
601
GAA
g
Dilute Phase
Mean Intensity
h
BRD4
i
7.5 nM
4 μM
HP1α
BRD4
Merge
SynTEF1
PA1
DMSO
j
Fold Change
SynTEF1
PA1
DMSO
7.5 nM
25 nM
50 nM
250 nM
1 μM
4 μM
[BRD4]
k
HP1α
BRD4
Merge
SynTEF1
PA1-JF646
1 μm
l

Fig. 3.1 | BRD4 can access HP1 condensates

a, Titration of 601 DNA and GAA-repeat DNA with increasing concentrations of HP1α. Samples imaged by DIC microscopy. Fluorescence micrographs of samples at 64 μM HP1α show its enrichment in condensates.

b, The change in HP1α condensate area with increasing concentration with 147 bp 601 or $(GAA)_{49}$ repeat DNA. A total of over 28,000 total droplets and 4 replicates were analyzed.

c, Schematic showing HP1α can form condensates with GAA repeat DNA.

d, Chemical structure of PA1-JF646. The polyamide found in SynTEF1 conjugated to fluorophore JF646 instead of JQ1.

e, PA1-JF646 enters HP1α condensates formed with GAA-repeat DNA, but not 601 DNA.

f, PA1-JF646 fluorescence intensity in condensates formed with $(GAA)_{49}$ or 601 DNA. PA1-JF646 preferentially recognizes GAA repeat DNA sequences.

g, Depletion of PA1-JF646 fluorescence intensity from the dilute phase in presence of HP1α-DNA condensates.

h, Schematic showing PA1-JF646 and SynTEF entering HP1α-GAA DNA condensates.

i, Access to BRD4 into HP1α-GAA DNA condensates in the presence of SynTEF1, PA1 or DMSO. BRD4 preferentially binds SynTEF1 saturated HP1α-GAA DNA condensates at low concentrations.

j, Quantification of (i) over 15,000 droplets and 2 replicate samples. Fold change is fluorescence intensity normalized to the average intensity in condensates treated with DMSO.

k, Super-resolution micrographs showing SynTEF1-dependent entry of BRD4 into pre-formed HP1α-DNA condensates.

l, Schematic showing SynTEF1-mediated BRD4 recruitment to HP1α condensates.

3.5 Polyamides change the properties of HP1α condensates in a DNA sequence specific manner

Since HP1α readily associates with 601 and GAA DNA sequences and form similar sized condensates, we wanted to know whether a similar amount of HP1α exists in those condensates. To test this, we titrated HP1α and measured the droplet intensity, and found that HP1α formed denser droplets when bound with 601 than GAA DNA at lower protein concentrations (Fig 3.2 a-b). Next, to understand how DNA availability changes in response to polyamide addition, we measured HP1α droplet size and intensity when formed in the presence of PA1-JF646 (Fig 3.2 c-e). HP1α bound to GAA DNA was found to be much more compressed than 601 bound droplets. This result may be caused by PA1-JF646 binding to and bridging across two GAA repeat DNA sequences which would provide a less motile and larger scaffold for HP1α to associate with. HP1α is sensitive to DNA size differences and will produce droplets with different tensile properties [13].

Since BRD4 can form LLPS condensates with HP1α, we wanted to understand if HP1α can restrict BRD4 entry by occupying the DNA available to BRD4, and whether SynTEF1 can still facilitate BRD4 entry. To test this, we preincubated either SynTEF1 or PA1-JF646 with DNA and HP1α, and then added BRD4 to the pre-formed droplets (Figs 3.2 f-h). SynTEF1 was again able to greatly increase BRD4 entry into HP1α droplets, increasing both BRD4 entry and the overall droplet size (Figs 3.2 f-g). To better understand the sub-structure of these droplets we imaged them under super-resolution microscopy. These condensates showed homogenous HP1α throughout the droplets, but an enrichment of PA-JF646 at the periphery, while BRD4 entered at low levels (Fig 3.2 h-i). SynTEF1 bound DNA resulted in the enrichment of BRD4, primarily to the

periphery of the HP1α droplet in a similar manner as PA1-JF646. These unanticipated results show that the interior of pre-formed HP1α droplets are non-uniform, and SynTEF1 mediates BRD4 recruitment at primarily at the boundary of the dense phase.

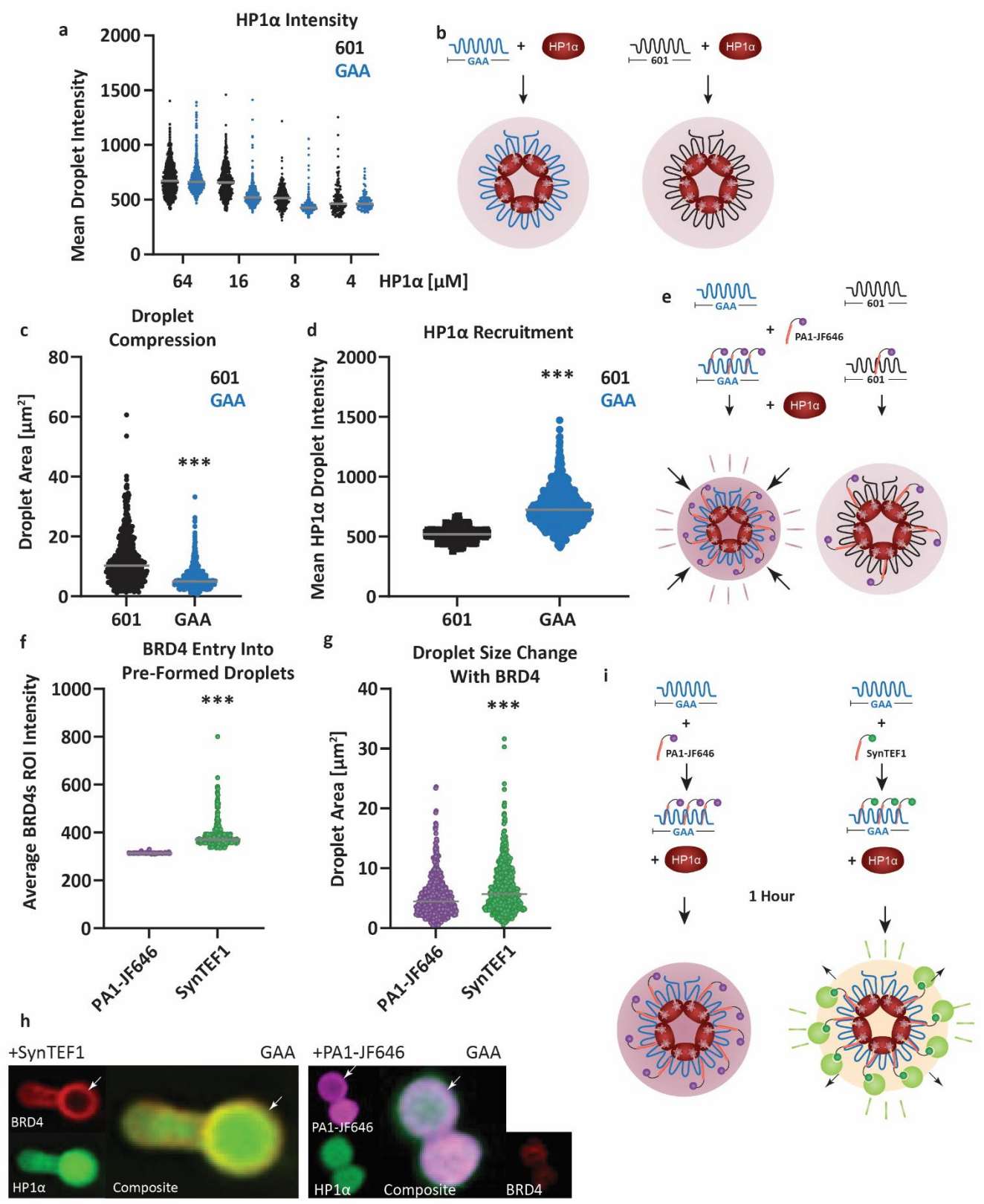

a
HP1α Intensity
Mean Droplet Intensity
2000
1500
1000
500
0
601
GAA
64
16
8
4
HP1α [μM]
b
GAA
HP1α
601
HP1α
c
Droplet Compression
Droplet Area [μm²]
80
60
40
20
0
601
GAA
***
601
GAA
d
HP1α Recruitment
Mean HP1α Droplet Intensity
2000
1500
1000
500
0
***
601
GAA
601
GAA
e
GAA
601
PA1-JF646
GAA
601
HP1α
f
BRD4 Entry Into Pre-Formed Droplets
Average BRD4s ROI Intensity
1000
800
600
400
200
0
***
PA1-JF646
SynTEF1
g
Droplet Size Change With BRD4
Droplet Area [μm²]
40
30
20
10
0
***
PA1-JF646
SynTEF1
i
GAA
PA1-JF646
GAA
HP1α
GAA
SynTEF1
GAA
HP1α
1 Hour
h
+SynTEF1
GAA
BRD4
HP1α
Composite
+PA1-JF646
GAA
PA1-JF646
HP1α
Composite
BRD4

Fig 3.2 | HP1α droplet structure is modulated by the addition of polyamides and BRD4

a, HP1α droplet intensity decreases with protein concentration. HP1α droplets formed with GAA DNA has less HP1α recruitment than 601 DNA as the concentration decreases.

b, Diagram showing the formation of HP1α droplets with GAA or 601 DNA

c-d, HP1α droplets formed in the presence of PA-JF646 bound to GAA. Droplet size was measured in **(c),** and droplets formed in the presence of 601 DNA were found to be larger than those formed with GAA DNA. HP1α droplet intensity was measured **(d)** and found to be more highly recruited by GAA DNA. Droplets were formed with μM HP1α, 0.1 μM DNA, and 1μM PA1-JF464.

e, Diagram showing the effect of DNA sequence on HP1α droplet formation in the presence of polyamide.

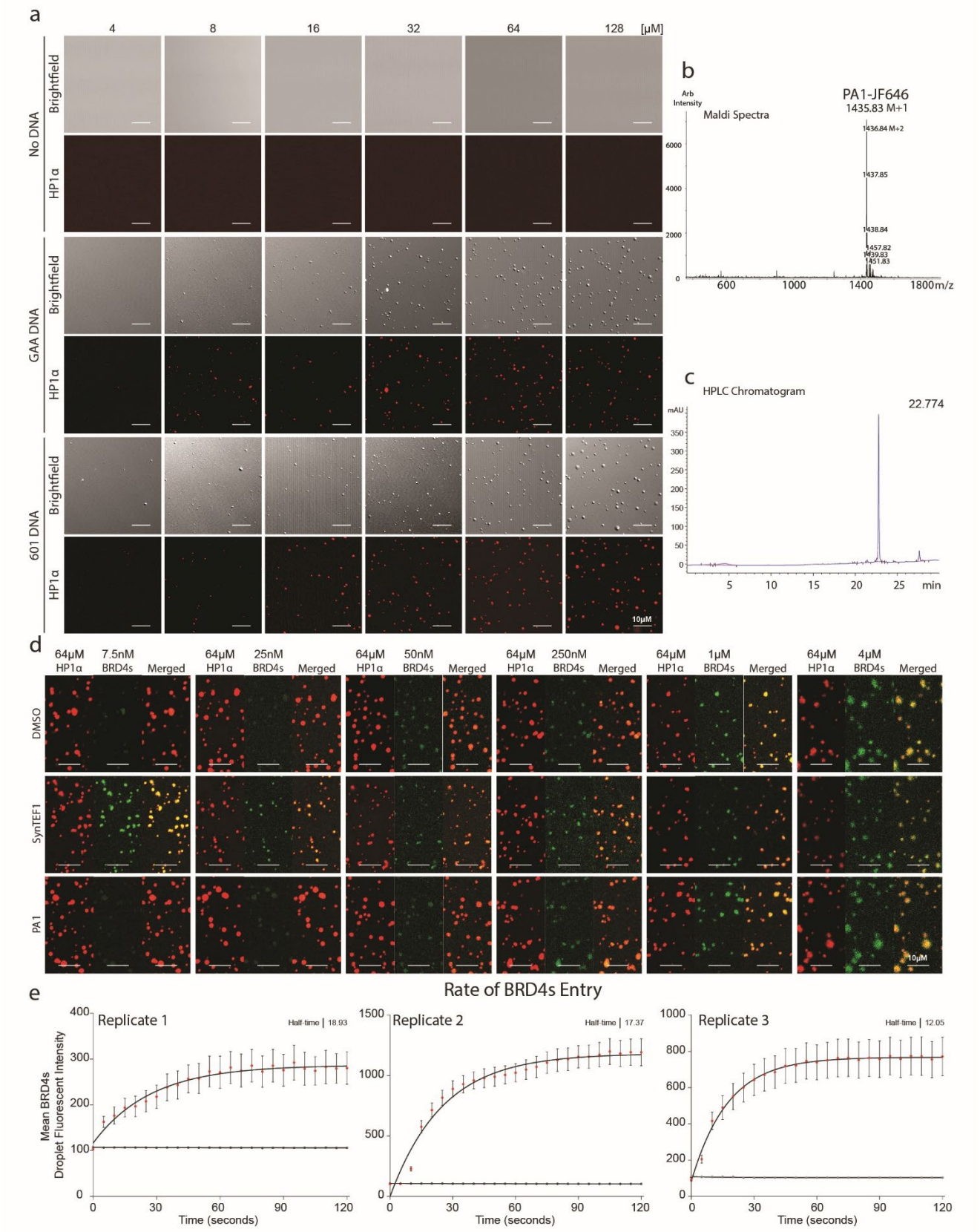
f, BRD4 entry into mature HP1α droplets. BRD4 prefers to enter droplets formed with SynTEF1. Droplets were formed with 64 μM HP1α, 0.1 μM GAA DNA, and 1 μM PA1-JF646 or SynTEF1.

g, HP1α droplet size in response to BRD4 treatment. BRD4 entry into SynTEF1 treated condensates results in an increase in condensate size.

h, Super-resolution microscopy of HP1α condensates formed with GAA DNA in the presence of SynTEF1 or PA1-JF646. BRD4 forms a pronounced ring at the periphery of the HP1α droplet. PA1-JF646 also forms a ring but less pronounced in the HP1α condensate.

3.6 Conclusions

The calibrated restoration of *FXN* levels upon SynTEF1 treatment may well benefit from the ability of HP1α to dynamically engage a segment of DNA within the repressive condensate. The HP1α-DNA condensates are resistant to rapid or repeated force, but because of rapid binding and dissociation kinetics, the condensates are susceptible to sustained force of the scale that is exerted by slowly elongating Pol II [13]. Such viscoelastic properties of the HP1α condensates at GAA-repeats would rebuff high flux transcription bursts but would permit passage to a slowly elongating Pol II that is assisted by synthetically recruited elongation machinery (Fig. 3.1i). This model agrees with previous observations that condensates have viscoelastic properties, and their response to perturbations depends on the frequency of the perturbation [13, 28, 29]. Beyond disease-causing repeat expansion in *FXN*, the ability of SynTEF1 to partition BRD4 into HP1 condensates and license transcription has direct implications for the mechanisms by which natural transcription factors elicit gene expression within heterochromatin. In a negative-feedback loop, the resulting transcripts often recruit repressive epigenetic machinery to silence expression [30-32]. Similar cellular surveillance may function as a rheostat to moderate levels of *FXN* expression in SynTEF1 treated cells.


a
4
8
16
32
64
128
[μM]
No DNA
GAA DNA
601 DNA
Brightfield
HP1α
10μM
b
Arb Intensity
Maldi Spectra
PA1-JF646
1435.83 M+1
1436.84 M+2
1437.85
1438.84
1457.82
1439.83
1451.83
600
1000
1400
1800 m/z
c
HPLC Chromatogram
22.774
mAU
min
d
64μM HP1α
7.5nM BRD4s
25nM BRD4s
50nM BRD4s
250nM BRD4s
1μM BRD4s
4μM BRD4s
Merged
DMSO
SynTEF1
PA1
10μM
e
Rate of BRD4s Entry
Replicate 1
Replicate 2
Replicate 3
Half-time | 18.93
Half-time | 17.37
Half-time | 12.05
Mean BRD4s Droplet Fluorescent Intensity
Time (seconds)

Extended Data Figure 3.1 | BRD4 is recruited to HP1α phase separated condensates by SynTEF1

a, Titration of HP1α without DNA, with $(GAA)_{49}$ DNA, or with 147bp 601 DNA. HP1α phase separates with both $(GAA)_{49}$ and 601 DNA with increasing concentration but fails to phase separate in the absence of DNA.

b-c, MALDI TOF spectra **(b)** and HPLC chromatogram of PA1-JF646 showing compound purity.

d, Titration of BRD4s with 64 μM HP1α either in the presence of DMSO vehicle, SynTEF1, or PA1. SynTEF1 facilitates BRD4 recruitment into HP1α droplets at low concentrations.

e, Individual replicates of the timed entry of 50 nM BRD4s into 64 μM HP1α condensates preformed with SynTEF1 (red) or vehicle DMSO (black). SEM between droplets in each replicate is shown. Half-time calculated by one-phase association.

3.7 Bibliography

1. Yarwood, R., et al., *Membrane trafficking in health and disease.* Disease Models & Mechanisms, 2020. **13**(4).
2. Alberti, S., A. Gladfelter, and T. Mittag, *Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates.* Cell, 2019. **176**(3): p. 419-434.
3. Li, P., et al., *Phase transitions in the assembly of multivalent signalling proteins.* Nature, 2012. **483**(7389): p. 336-340.
4. Lyon, A.S., W.B. Peeples, and M.K. Rosen, *A framework for understanding the functions of biomolecular condensates across scales.* Nature Reviews Molecular Cell Biology, 2021. **22**(3): p. 215-235.
5. Han, X., et al., *Roles of the BRD4 short isoform in phase separation and active gene transcription.* Nat Struct Mol Biol, 2020. **27**(4): p. 333-341.
6. Gibson, B.A., et al., *Organization of Chromatin by Intrinsic and Regulated Phase Separation.* Cell, 2019. **179**(2): p. 470-484 e21.
7. Strom, A.R., et al., *Phase separation drives heterochromatin domain formation.* Nature, 2017. **547**(7662): p. 241-245.
8. Larson, A.G., et al., *Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin.* Nature, 2017. **547**(7662): p. 236-240.
9. Erwin, G.S., et al., *Synthetic transcription elongation factors license transcription across repressive chromatin.* Science, 2017. **358**(6370): p. 1617-1622.
10. Zhao, T., et al., *Heterochromatin Protein 1 Binds to Nucleosomes and DNA in Vitro *.* Journal of Biological Chemistry, 2000. **275**(36): p. 28332-28338.

11. Muchardt, C., et al., *Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha.* EMBO Rep, 2002. **3**(10): p. 975-81.

12. Nishibuchi, G., et al., *N-terminal phosphorylation of HP1α increases its nucleosome-binding specificity.* Nucleic Acids Res, 2014. **42**(20): p. 12498-511.

13. Keenen, M.M., et al., *HP1 proteins compact DNA into mechanically and positionally stable phase separated domains.* Elife, 2021. **10**.

14. Qin, W., et al., *HP1beta carries an acidic linker domain and requires H3K9me3 for phase separation.* Nucleus, 2021. **12**(1): p. 44-57.

15. Calo, E. and J. Wysocka, *Modification of Enhancer Chromatin: What, How, and Why?* Molecular Cell, 2013. **49**(5): p. 825-837.

16. Whyte, Warren A., et al., *Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes.* Cell, 2013. **153**(2): p. 307-319.

17. Hnisz, D., et al., *Super-Enhancers in the Control of Cell Identity and Disease.* Cell, 2013. **155**(4): p. 934-947.

18. Sabari, B.R., et al., *Coactivator condensation at super-enhancers links phase separation and gene control.* Science, 2018. **361**(6400).

19. Cho, W.-K., et al., *Mediator and RNA polymerase II clusters associate in transcription-dependent condensates.* Science, 2018. **361**(6400): p. 412-415.

20. Sanz-Álvarez, M., et al., *Expression of Phosphorylated BRD4 Is Markedly Associated with the Activation Status of the PP2A Pathway and Shows a Strong Prognostic Value in Triple Negative Breast Cancer Patients.* Cancers (Basel), 2021. **13**(6).

21. Pedregal, M., et al., *Linking pBRD4 and PP2A inhibition in triple negative breast cancer.* Journal of Clinical Oncology, 2018. **36**(15_suppl): p. e13011-e13011.

22. Mirkin, S.M., *Expandable DNA repeats and human disease.* Nature, 2007. **447**(7147): p. 932-940.

23. Burnett, R., et al., *DNA sequence-specific polyamides alleviate transcription inhibition associated with long GAA.TTC repeats in Friedreich's ataxia.* Proceedings of the National Academy of Sciences, 2006. **103**(31): p. 11497-11502.

24. Blattes, R., et al., *Displacement of D1, HP1 and topoisomerase II from satellite heterochromatin by a specific polyamide.* EMBO J, 2006. **25**(11): p. 2397-408.

25. Klein, I.A., et al., *Partitioning of cancer therapeutics in nuclear condensates.* Science, 2020. **368**(6497): p. 1386-1392.

26. Gibson, B.A., et al., *Organization of Chromatin by Intrinsic and Regulated Phase Separation.* Cell, 2019. **179**(2): p. 470-484.e21.

27. Shin, Y., et al., *Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome.* Cell, 2018. **175**(6): p. 1481-1491.e13.

28. Alshareedah, I., et al., *Programmable viscoelasticity in protein-RNA condensates with disordered sticker-spacer polypeptides.* Nat Commun, 2021. **12**(1): p. 6620.

29. Jawerth, L., et al., *Protein condensates as aging Maxwell fluids.* Science, 2020. **370**(6522): p. 1317-1323.

30. Allshire, R.C. and H.D. Madhani, *Ten principles of heterochromatin formation and function.* Nat Rev Mol Cell Biol, 2018. **19**(4): p. 229-244.

31. Eimer, H., et al., *RNA-Dependent Epigenetic Silencing Directs Transcriptional Downregulation Caused by Intronic Repeat Expansions.* Cell, 2018. **174**(5): p. 1095-1105.e11.

32. Yu, R., X. Wang, and D. Moazed, *Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation.* Nature, 2018. **558**(7711): p. 615-619.

3.8 Materials and Methods

Protein Labeling

Protein was dialyzed into 10 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT. HP1α was labeled with Oregon Green (Fisher O6147) and BRD4s was labeled with Rhodamine Red (Thermofisher R6160). Briefly the dye was dissolved at 100 mM in DMSO and added to protein in a 20:1 dye-to-protein molar ratio and incubated on a shaking platform for 5 hours. The reaction was quenched with 2 M excess Tris. Protein was then dialyzed into 20 mM Tris pH 7.8, 150 mM NaCl, and 5 mM DTT.

Polyamide Synthesis

Polyamide (PA1) was synthesized using solid phase peptide synthesis and then conjugated to activated JF646-NHS ester by using DIPEA and DMF. The reaction was carried out at room temperature for 6h. After completion of the reaction, diluted in 15% of acetonitrile in H2O and injected in Prep-HPLC to purify the compound. The pure fractions were collected and lyophilized to obtain pure compound.

Phase Separation Assays

DNA oligos were purchased from IDTDNA, resuspended in 50 mM NaCl and annealed at 95 °C for 5 minutes and ramped to 4 °C at 0.1 degree/s. DNA was re-annealed before each experiment and incubated with polyamide for at least 3 h at 4 °C. HP1α and BRD4s were incubated with DNA/polyamide in dialysis buffer for 1 hour before imaging on a Nikon C2. Time course

experiments were performed by pre-forming HP1α condensates in PCR tubes for 1hr, transferring the solution to a coverslip affixed to a perforated 35 mm glass plate, and then adding BRD4s at 50 nM. The samples were imaged on the Marianas 2 and images were taken every 5 seconds. Super-resolution microscopy was performed on a Zeiss LSM980 Airyscan microscope on a solution in the 2-phase regime equilibrated for 90 minutes.

Image Analysis

Protein colocalization in phase-separated condensates was performed by image analysis using Nikon Elements. An ROI was generated using the HP1α FITC fluorescence channel and average fluorescence intensity measurements were taken from all channels. The kinetics of BRD4s entry into HP1α condensates was determined by analysis of movies in ImageJ using a line ROI at default width covering the HP1α condensate. BRD4s channel intensity was recorded at each timepoint.

Chapter 4 | Future Directions

4.1 Introduction

In this work I have shown that FRDA patient derived cells contain functional HP1 protein at the *FXN* gene that works to down-regulate gene expression. SynTEF1 restores transcription by recruiting BRD4 and pTEFb without removing this epigenetic repression, which instead unexpectedly increases after treatment in order to regulate transcript levels. I then showed that the combination treatment of SynTEF1 and HDACi-109 works synergistically to change the epigenetic state and further increase *FXN* expression. This occurs because SynTEF1 facilitates the entry of BRD4 into HP1α phase separated condensates, producing a state in which both proteins are present and functional at a specific genetic locus. This bi-functional state generates *FXN* transcript whose expression levels are "tuned" by the continued placement of H3K9me3 and the recruitment of HP1 proteins. HDAC-109 treatment further overcomes this epigenetic repression and provides additional H3K9Ac modifications for BRD4 to bind. This work provides insight into the mechanisms that govern gene regulation at regions where these standard active and repressive epigenetic reader proteins interact.

As a therapeutic, SynTEF1 is a novel and modular compound which can be rearranged to target many different types of repetitive regions and redesigned to deliver specific effector "warheads" to different locations in the genome. The development of this tool has the potential to treat other repeat-containing diseases and can additionally be redesigned to genes in the future.

The effectiveness of SynTEF1 at overcoming the heterochromatic epigenetic regulation at *FXN* capitalizes both on the presence of stalled polymerase and the interactions of BRD4 and HP1 proteins. The inability of HDACi-109 to fully restore *FXN* transcription suggests that

targeting just the active epigenetic state of a gene that is homeostatically regulated is not sufficient to overcome heterochromatic silencing. This occurs because, while BRD4 can enter HP1 droplets, it does so much less effectively then when SynTEF1 drives its recruitment.

4.2 Cellular Co-localization of HP1 and BRD4

We have shown in this work that SynTEF1 treatment restores *FXN* nascent transcription at a similar number of genomic loci as found in healthy cells. We also know show that all three HP1 paralogues colocalize at the *FXN* locus by ChIP-seq, and that BRD4 can enter HP1α phase separated condensates. While these experiments do provide a basic understanding of how BRD4 and HP1 interact, the visualization of these proteins colocalizing *in vivo* would provide additional direct evidence of this bivalent state. Immunofluorescence of HP1α, HP1β, HP1γ along with BRD4 would allow us to directly observe the change in the overlapping regions of these disparate proteins. When paired with DNA and RNA FISH, or with PA1-JF646, we could visualize BRD4 at actively transcribing *FXN* loci.

Visualizing these bound proteins at similar loci by immunohistochemistry will provide information on colocalization, but to determine the state of these proteins and whether they form liquid like states in the cell we would need to understand their turnover in the droplet. This can be explored by overexpressing fluorescently labeled HP1 and BRD4 proteins in healthy and FRDA model fibroblasts. These cells can be grown on coverslips and the change in protein colocalization can be directly observed. Fluorescence recovery after photobleaching (FRAP) experiments then can be performed on BRD4 and HP1 puncta to determine how quickly the

areas recover which provides the rate of protein turnover at the locus. When used in conjunction with PA1-JF646, FRAP measurements can be taken individually of HP1α, HP1β, or HP1γ. Additional fluorescently labeled compounds are under development which can also recruit BRD4 to the HP1 bound FXN locus. FRAP experiments at these loci will provide additional information on the changes in LLPS condensates at *FXN*.

Unfortunately, smFISH and smiFISH experiments have been unsuccessful at locating the *FXN* transcript in the cell, which has instead has required the use of a larger FOSMID construct (see Chapter 2). Another mechanism to visualize the change in *FXN* transcript after SynTEF1 treatment is through inserting an MS2 loop into the 3' end of the gene, which provides a handle for an MS2-GFP protein to bind and can be done in live cells [177, 178]. A CRISPR knock-in of multiple MS2 loops in the *FXN* gene along with the stable overexpression of MS2-bound reporter protein (frequently GFP) would provide another way to visualize the change in *FXN* transcription. Since these cells do not need to be fixed, the overexpression of HP1 or BRD4 proteins may provide additional insights into their recruitment and interaction at the gene.

A different protocol called multi-cut & tag has been recently developed, which builds on the familiar Cut & Tag protocol but instead probes genomic loci bound by multiple specific proteins [179, 180]. It uses transposases pre-loaded with two different adapter sequences and complexed with distinct antibodies. This generates cuts in regions of DNA that have one or both target proteins, depending on if one or both adapters are sequenced. Adaption of this method to SynTEF1 treated FRDA cells would provide orthogonal support for SynTEF1 mediated BRD4 colocalization with HP1 proteins, and provide a genome-wide analysis of where these proteins interact. HP1γ can be found in inactive regions of chromatin dense with H3K9me3

modifications, but can also bind actively transcribed euchromatin which does not contain these marks [61]. It may do this by entering and other condensate types but not actually interacting with the DNA or nucleosomes, and instead regulating condensate formation or structure. Multi-cut & tag would provide insight into how HP1γ is associating with these disparate regions, and whether it is in fact binding DNA independent of H3K9me3. Additionally, since all three HP1 paralogues do not have entirely complementary coverage along the genome, it would provide further insight into how they bind to and restrict transcription.

4.3 HP1 DNA specificity

HP1 proteins associate with satellite DNA and can recognize both DNA and RNA. In this work we have shown that HP1α readily forms phase separated condensates with both 601 and GAA 147bp DNA sequences. Others have shown that HP1α forms phase separated condensates with short and long DNA, HP1γ forms condensates with long DNA only, and HP1β doesn't phase separate with DNA at all [64]. HP1γ does show some nucleotide specificity as it recognizes hexameric repeats, associates with SINE elements, and can help regulate splicing [181]. Since HP1γ displays nucleotide specificity, it logically follows that HP1α and HP1β may also have a preferred nucleotide binding sequence. Because HP1 proteins form large multimeric structures, these DNA sequences would be difficult to observe with standard ChIP methods. And since improper silencing of actively transcribed genes frequently results in disease, it would be very beneficial to understand what sequences the HP1 proteins prefer to bind, and whether those sequences facilitate phase separation.

Sequence specific Protein-DNA interactions can be enriched for using cognate-site identification (CSI) [182]. This *in vitro* method uses a randomized DNA sequence surrounded on both ends with barcoded DNA as a bait for protein binding. The preferred sequences are enriched through a series of protein pull-downs and DNA amplifications, which results in pool of DNA that can be sequenced to determine the preferred protein binding motifs. CSI provides a method to understand if any of the HP1 proteins display sequence specificity, or if their recruitment to DNA is facilitated through H3K9me3 binding and its affinity for DNA is only a mechanism to maintain the heterochromatic state. Since HP1 is readily found in alpha-satellite repeats, I hypothesize similar DNA sequences would be observed by CSI. Once these preferred sequences are determined, fluorescence polarization (FP) and gel shift assays can be used to determine relative binding affinities.

Phase separation in-part relies on the presence of multiple binding sites or interaction motifs for a protein to associate with [63]. The reason HP1β and HP1γ show less of a propensity to phase separate with DNA than HP1α may be due to a higher, or different sequence specificity [64]. CSI experiments with the HP1 paralogues will uncover these motif differences. When presented with their preferred binding sequence, HP1β and HP1γ may phase separate more readily instead of the generic 601 sequence. Additionally, these specific sequences may facilitate phase separation at different protein concentrations, or change the way other HP1 proteins co-condense. If one HP1 paralogue binds more tightly to a specific DNA sequence, it may more readily restrict access to the DNA, which in turn would make it more difficult for other proteins such as BRD4 to access. This would provide evidence of a novel, sequence specific mechanism of HP1 transcriptional regulation.

4.4 The regulation of co-condensate formation

Perhaps the most unexpected finding presented in this work is that BRD4 can enter HP1α phase separated condensates despite their opposing function in the cell. HP1β does not readily form condensates *in vitro* with short DNA sequences, and it is hypothesized to regulate and possibly even dissolve HP1α condensates, in turn making the DNA more accessible [64]. HP1γ can, on the other hand, enter and partially displace HP1α from the condensate without causing its complete dispersion. Because BRD4 and HP1α can co-condense, it would be beneficial to understand how HP1β and HP1γ interface with that condensate. I hypothesize that BRD4 would more readily mix with HP1α condensates in the presence of the other paralogues since they have been shown to disperse the droplets on their own.

To test this hypothesis, we would need to pre-form HP1α droplets with standard 601 DNA 147bp sequences in the presence and absence of BRD4 and titrate in HP1β or HP1γ. This experiment can be further enhanced if a preferred DNA sequence is identified by CSI. While HP1β destabilizes HP1α condensates, HP1α:HP1γ droplets are stable for short periods and can be pre-formed [64]. BRD4 can then be titrated in to determine if the presence of HP1γ affects the rate or concentration at which BRD4 can enter, or if the droplet further destabilizes. How the different condensates regulate BRD4 entry will provide insight into how these proteins interact, and how their presence at active genes regulates transcription.

4.5 SynTEF1 and cellular function

To this point we have shown that SynTEF1 overcomes the epigenetic transcriptional repression and increases *FXN* expression in FRDA patient-derived cellular models. The increase in transcript has been shown previously to occur concomitantly with an increase in FXN protein [136]. I hypothesize that the restoration of FXN protein will result in a restoration of cellular homeostasis including increased mitochondrial iron-sulfur cluster biogenesis, resulting in both increased metabolic function and decreased levels of ROS generation.

The treatment of FRDA cells with SynTEF1 should produce an increase in mitochondrial respiration. This can be measured using a Seahorse instrument, which is designed to look for changes in respiratory function by measuring pH and O_2 changes in the media. After SynTEF1 restores FXN protein levels, we can measure the relative restoration of mitochondrial function in FRDA patient cells.

An orthogonal method to probe metabolic function is measuring the change in metabolites in the cell. The restoration of mitochondrial function by SynTEF1 should result in large changes to the cellular metabolome, which can be observed by mass-spectrometry before and after SynTEF1 treatment. Metabolomic analysis will allow us to identify the specific pathways that are most affected during FRDA, which can be targeted for future drug development.

4.6 Small molecule development

SynTEF1 is a singular example of how polyamides can be leveraged to target endogenous proteins to a repeat-based disease. The design of SynTEF1 is modular in both the sequence of the polyamide and its attached warhead. The polyamide can be altered to target different repeat regions, and the warhead can be exchanged to recruit specific proteins. This design flexibility allows researchers to directly target specific regions of DNA and affect the proteins present there. While the modularity of polyamide-based drugs offers many possibilities, there are many other concerns that must be addressed when considering this therapeutic design.

The pharmaceutical industry has spent many years screening for and identifying small molecules that target transcription, many of which have the potential to act as warheads for polyamides. This large pool of compounds, with known therapeutic targets, provides a good starting point to identify compounds to complex with a polyamide. The proteins targeted by the small molecule must be bound tightly but retain their function in order to be useful. Therefore, the small molecule must not target the active site of the protein, so that when the protein is recruited to the locus it is still active. It must also have a chemically modifiable region that is not utilized for protein recognition to use as an attachment point for the polyamide. Therefore, such as with the JQ1 moiety in SynTEF1, molecules that function as inhibitors by strongly associating at protein binding and recognition sites can be complexed with a polyamide and instead be used to recruit those proteins to the target region.

Since there are specific requirements for warheads and the available pool of potential compounds is limited to what has already been identified, additional warhead design will be

useful. The rational design of compounds using new docking programs such as DOCK3.7 allows for the identification of new, potentially useful small molecules that contain chemical features that are necessary to complex to a polyamide [183]. These *in silico* programs use pre-existing protein structures and model small molecules to fit in those structures. Therefore, rationally designing compounds to target the protein's recognition site allow you to both visualize and select for compounds that will be more likely to be functional when complexed later to a polyamide. This type of screen can identify novel warheads that can be validated using subsequent binding assays. The strongest binding compounds can then be complexed to polyamides and used to create new classes of SynTEFs.

Since SynTEF1 is effective at recruiting BRD4 to the HP1 bound regions at *FXN*, it follows that directing BRD4 to other disease-causing genes that contain repeat expansions may positively affect local transcription. To do this the polyamide sequence must be altered to target the specific region of DNA [163]. While this is a logical step, the mechanism of inhibition should be considered to determine if BRD4 recruitment will be beneficial. We have shown in this work that SynTEF1 has high specificity to the *FXN* gene. If the transcriptional repression of the target gene is not a result of heterochromatic silencing similar to that found in FRDA, the recruitment of BRD4 and subsequently pTEFb will likely not overcome the epigenetic blockage. Instead, it will likely be beneficial to recruit specific proteins that target the mechanism of repression specific for that disorder.

Other diseases have a repeat expansion in the coding region of a gene which results in a large amino-acid insertion into the translated protein which affects its function. This occurs in many repeat expansion diseases (Figure 1.1) including Huntington's Disease and specific

Spinocerebellar ataxias. In these cases, inhibiting transcription at the disease-causing locus is beneficial for the patient. Therefore, the targeted recruitment of the heterochromatic HP1 proteins or the PRC2 complex to the repeat may suffice downregulate transcription at the disease locus. To target these proteins directly, I hypothesize that polyamides bound with a small H3K9me3 or H3K27me3 modified histone peptide would be effective in recruiting these proteins to the disease-causing region, resulting in the localized repression of the diseased gene while maintaining activity in the healthy copy.

To rationally design future polyamides, we need to understand both the benefits and limitations of these compounds. They are large and difficult to manufacture, both of which are barriers that must be overcome for large compound screens. Instead, intelligently designing compounds based on the rules of polyamide-DNA recognition as well as identifying new, specific warheads will be an effective method to continue the development of polyamide-based therapeutics for repeat based diseases.

4.7 Bibliography

1. Muramoto, T., et al., *Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation.* Proceedings of the National Academy of Sciences, 2012. **109**(19): p. 7350-7355.
2. Alexander, J.M., et al., *Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity.* eLife, 2019. **8**: p. e41769.
3. Gopalan, S., et al., *Simultaneous profiling of multiple chromatin proteins in the same cells.* Molecular Cell, 2021.

4. Skene, P.J. and S. Henikoff, *An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites.* eLife, 2017. **6**: p. e21856.
5. Vakoc, C.R., et al., *Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin.* Mol Cell, 2005. **19**(3): p. 381-91.
6. Keenen, M.M., et al., *HP1 proteins compact DNA into mechanically and positionally stable phase separated domains.* Elife, 2021. **10**.
7. Rachez, C., et al., *An impact of HP1γ on the fidelity of pre-mRNA splicing arises from its ability to bind RNA via intronic repeated sequences.* bioRxiv, 2019: p. 686790.
8. Carlson, C.D., et al., *Specificity landscapes of DNA binding molecules elucidate biological function.* Proceedings of the National Academy of Sciences, 2010. **107**(10): p. 4544-4549.
9. Alberti, S., A. Gladfelter, and T. Mittag, *Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates.* Cell, 2019. **176**(3): p. 419-434.
10. Erwin, G.S., et al., *Synthetic transcription elongation factors license transcription across repressive chromatin.* Science, 2017. **358**(6370): p. 1617-1622.
11. Bender, B.J., et al., *A practical guide to large-scale docking.* Nature Protocols, 2021. **16**(10): p. 4799-4832.
12. Dervan, P.B. and R.W. Bürli, *Sequence-specific DNA recognition by polyamides.* Current Opinion in Chemical Biology, 1999. **3**(6): p. 688-693.

Appendix 1 | A more effective method to prepare pancreatic islets for immunohistochemistry

Pancreatic islets are small, spheroid, clumps of cells that are isolated directly from the pancreas using a series of enzymatic digestions and purifications. These complex spherical organoids are highly vascularized and consist primarily of glucagon secreting alpha cells and insulin secreting beta cells, interspersed with other less common cell types. Typical islets contain approximately 1000 cells and are typically 50 – 150 μm in diameter [1, 2]. Islets isolated from a mouse or human pancreas can be isolated and treated *ex vivo* with potential therapeutics to directly stimulate and measure insulin secretion. Isolated islets can also be used to visualize the effect of small molecule treatments directly on islet tissue using immunohistochemistry. This was done by minimally suspending cells in dilute molten agarose along with colored agarose beads and taking slices through the cooled agarose block. This block was paraffin embedded and sliced using a microtome and mounted on a slide by the University of Wisconsin School of Veterinary Medicine Histology Department (Fig 1a-b). Due to the small diameter of the islet and the relatively large size of the agarose block, visualizing islets was successful ~50-75% of the time, and there was very little chance of obtaining more than one longitudinal slice through an islet (Fig 1b).

It was necessary to suspend the islets in molten agarose because layering molten and solid agarose resulted in splitting and general instability of the agarose block. To resolve the issue a microscope slide was layered on the ends with a thin portion of Whatman paper and standard labeling tape to create a 0.5 mm raised end on either side (Fig 1c). The islets were mixed with a minimal amount of liquid containing the blue agarose beads, and then resuspended in a minimal amount of molten agarose. This was compressed by the addition of a coverslip and rapidly cooled on a heat block submerged in ice, which stabilized the islets in a thin “pancake” of agarose (Fig 1d-e). This was then embedded in paraffin and sliced using the same method described above. This resulted in an increase in the number of islets that were able to be visualized in one field of view, the total number of islets that could be visualized on one microscope slide, and provided

the additional benefit of being able to reliably generate slides that contained cells from the same islet (Fig 1f).

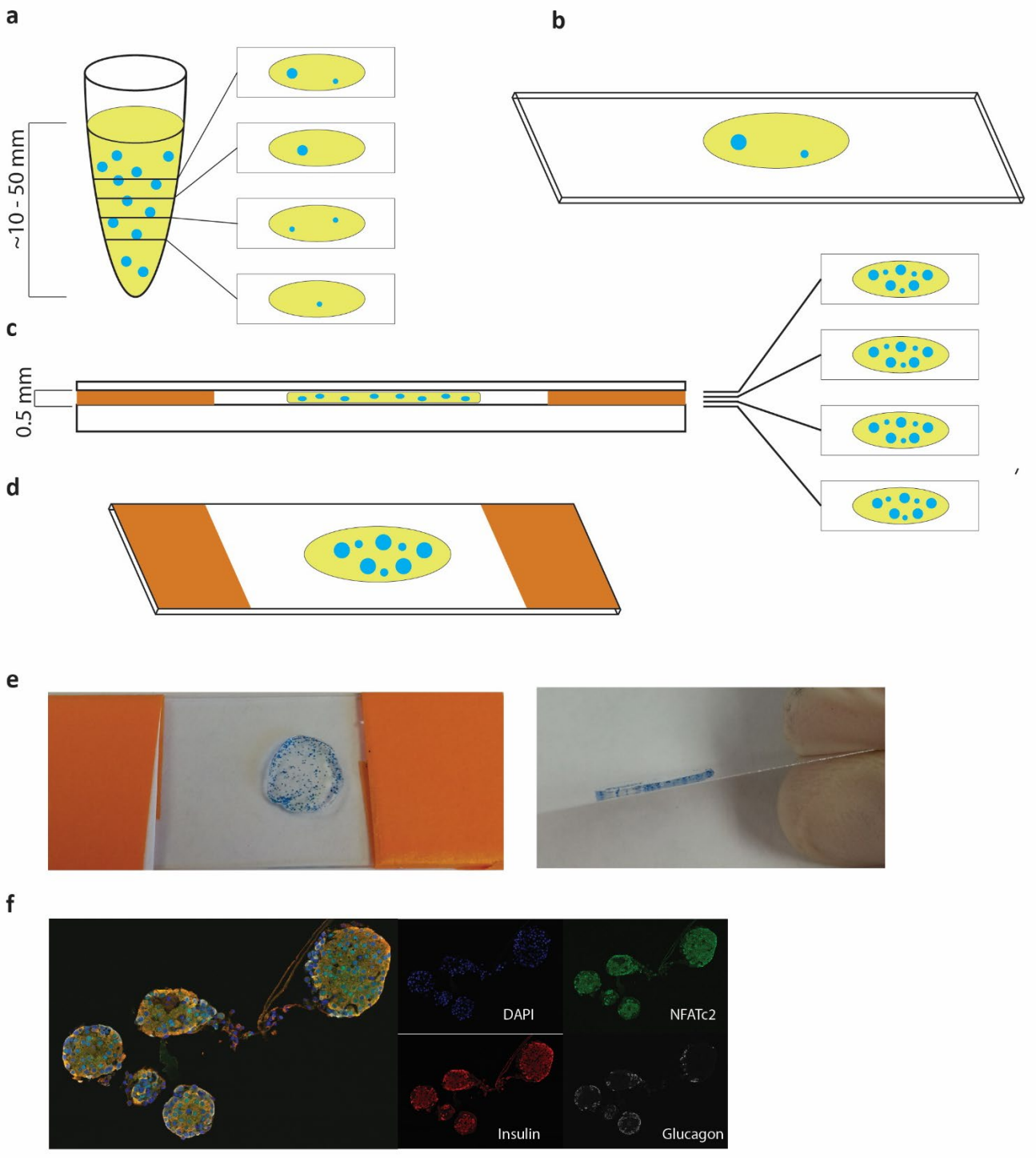

a
~10 - 50 mm
b
c
0.5 mm
d
e
f
DAPI
NFATc2
Insulin
Glucagon

Figure 1 | Preparation of pancreatic islets for immunohistochemistry

a-b, Representation of the standard method to embed pancreatic islets into agarose. The dispersion of islets in the Z axis leads to limited islet inclusion in individual slices.

c-e, The addition of spacers and the reduction of the Z axis as well as the relatively large surface area of the pancake provides more available surface area available for the microtome to slice through. This increases the likelihood of multiple islets being captured in a localized area.

f, Representative image of 5 islets captured in one 60x FOV. Blue is DAPI, red is insulin, white is glucagon, and green is the transcription factor NFATC2.

Methods

Setup

Warm heat pad

Fill ice bucket and put heat block on ice to cool

Keep blue agarose beads on ice

Heat 2% agarose + 1% Formalin (diluted in ddH_2O) until boiling

Buffer:

1mL PBS + 0.2% BSA / sample

Procedure

1. Pull PBS off of islets leaving ~100uL in tube.
2. Re-suspend islets in 1mL PBS + 0.2% BSA
3. Centrifuge at 180xg for 1 minute to pellet islets
4. Remove supernatant leaving ~30uL PBS + BSA
5. Use yellow tip to remove islets + supernatant and put onto mounting microscope slide with spacers
6. Add 5uL blue agarose beads
7. Use P10 to draw off as much liquid as possible
8. Put slide on heat pad for ~30s to warm glass
9. Using a yellow tip with ~1cm cut off from end, add 75uL of molten but near cooled 2% agarose + 1% Formalin and quickly re-suspend islets
10. Put coverslip on top of agarose to flatten into pancake.
11. Put microscope slide + coverslip onto heat block in ice to cool for 30s. Flip onto coverslip side and cool an additional 30s.
12. Gently remove agarose from microscope slide and put in labeled cassette for paraffin embedding and slicing.

Bibliography

1. Komatsu, H., et al., *Oxygen environment and islet size are the primary limiting factors of isolated pancreatic islet survival.* PLOS ONE, 2017. **12**(8): p. e0183780.
2. Pisania, A., et al., *Quantitative analysis of cell composition and purity of human pancreatic islet preparations.* Lab Invest, 2010. **90**(11): p. 1661-75.