Published in final edited form as:

Science. 2017 December 22; 358(6370): 1617–1622. doi:10.1126/science.aan6414.

Synthetic transcription elongation factors license transcription across repressive chromatin

Graham S. Erwin^{#1}, Matthew P. Grieshop^{#1}, Asfa Ali^{#1}, Qj Jun², Matthew Lawlor², Deepak Kumar^{3,4}, Istaq Ahmad^{3,4}, Anna McNally⁵, Natalia Teider⁵, Katie Worringer⁵, Rajeev Sivasankaran⁵, Deebe N. Syed⁶, Asuka Eguchi¹, Md. Ashraf¹, Justin Jeffery⁷, Xu Mousheng², Paul M. C. Park², Hasan Mukhtar⁶, Achal K. Srivastava³, Mohammed Faruq⁴, James E. Bradner^{2,5}, and Aseem Z. Ansari^{1,8,†}

¹Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA.

²Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.

³Department of Neurology, All India Institute of Medical Sciences, New Delhi, India.

⁴Genomics and Molecular Medicine, CSIR-Institute of Genomics and Integrative Biology, New Delhi, India.

⁵Novartis Institutes for BioMedical Research, Cambridge, MA 02139, USA.

⁶Department of Dermatology, University of Wisconsin-Madison, Madison, WI 53706, USA.

⁷Small Animal Imaging Facility, UW Carbone Cancer Center, Madison, WI 53792, USA.

⁸The Genome Center of Wisconsin, University of Wisconsin-Madison, Madison, WI 53706, USA.

[#] These authors contributed equally to this work.

Abstract

Releasing a paused RNA polymerase II into productive elongation is tightly-regulated, especially at genes that impact human development and disease. To exert control over this rate-limiting step, we designed sequence-specific synthetic transcription elongation factors (Syn-TEFs). These molecules are composed of programmable DNA-binding ligands flexibly tethered to a small molecule that engages the transcription elongation machinery. By limiting activity to targeted loci, Syn-TEFs convert constituent modules from broad-spectrum inhibitors of transcription into gene-specific stimulators. We present Syn-TEF1, a molecule that actively enables transcription across repressive GAA repeats that silence frataxin expression in Friedreich's ataxia, a terminal neurodegenerative disease with no effective therapy. Furthermore, the modular design of Syn-

[†] Corresponding author. azansari@wisc.edu.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.aan6414/DC1

Materials and Methods

Figs. S1 to S16

Tables S1 to S24

References (31–44)

Movie S1

TEF1 defines a general framework for developing a class of molecules that license transcription elongation at targeted genomic loci.

A longstanding challenge at the interface of chemistry, biology and precision medicine is to develop molecules that can be programmed to regulate the expression of targeted genes (1). It is increasingly evident that RNA polymerase II (Pol II) pauses during transcription (2, 3). Regulated release from the paused state into productive elongation is emerging as a critical step in gene expression. The number of diseases associated with proteins that play a role in implementing the pause or subsequent release into productive elongation is rapidly growing (4–6). In this context, we focused on creating molecules that enable Pol II to surmount barriers to productive elongation at targeted genomic loci. At their core, these synthetic transcription elongation factors (Syn-TEFs) incorporate two distinct chemical moieties: (i) programmable DNA binders that target desired genomic loci, and (ii) ligands that engage the transcription elongation machinery.

Pyrrole/imidazole-based polyamides have emerged as a class of synthetic molecules that can be programmed to bind specific DNA sequences using well-defined molecular recognition rules (7, 8). Recent examination of the genome-wide distribution of two polyamides designed to target different sequences revealed these molecules are primarily enriched at genomic loci bearing clusters of binding sites (9). A “summation of sites” (SOS) model that integrates the affinity of a given polyamide for all potential binding sites that occur within an ~400-base pair window best encapsulated the genome-wide binding preferences (9). Consistent with the SOS model, a polyamide previously designed to target a GAAGAAGAA site enriches at repressive GAA microsatellite repeats within the first intron of frataxin (*FXN*) in a cell line derived from a Friedreich’s ataxia (FRDA) patient (10).

In FRDA cells, expanded GAA repeats are enriched in repressive histone marks and can also adopt uncommon DNA conformations that impede transcription (11, 12). The number of repeats positively correlates with both the extent of repression and severity of disease (13, 14). The prevailing models in the field are that repressive chromatin and/or unusual DNA conformations present a barrier to the productive elongation of the *FXN* transcripts (11, 12, 15–18). Efforts to reverse repressive chromatin marks with freely diffusing histone deacetylase inhibitors or the utilization of a polyamide intended to drive uncommon structures toward canonical B-form DNA conformation did not elicit sufficient *FXN* expression (10, 19). Therefore, we reasoned that a synthetic molecule capable of binding repressive GAA repeats and *actively* assisting productive elongation would restore *FXN* expression to levels observed in normal cells. A pivotal step in the transition of a paused Pol II into productive elongation is the recruitment of the positive transcription elongation factor b (P-TEFb). This complex contains the cyclin-dependent kinase 9 (CDK9), which phosphorylates multiple proteins, including Pol II, to facilitate transcription elongation (2, 5, 20).

To avoid perturbing CDK9 kinase activity, we focused on ligands of BRD4, a protein that binds acetylated histones and engages active P-TEFb at transcribed genes (20). Among BRD4 ligands, JQ1 has been extensively characterized and shown to competitively displace BRD4 from regulatory regions of the genome (21). JQ1 therefore functions as a broad-

spectrum inhibitor of oncogene-stimulated transcription and a chemical derivative is currently in clinical trials (21). Based on its mechanism of action, we reasoned that tethering JQ1 to specific genomic loci would mitigate the global inhibitory properties and convert this molecule into a locus-specific *stimulator* of transcription. Moreover, rather than stimulating transcription initiation, we reasoned that JQ1-dependent recruitment of the elongation machinery across the length of the repressive GAA repeats, would enable Pol II to actively overcome the barrier to transcriptional elongation across the silenced *FXN* gene.

To design bifunctional Syn-TEFs, we examined the crystal structures of polyamide-nucleosome complex and JQ1-BRD4 bromodomain complex and identified optimal sites for chemical conjugation (Fig. 1A) (21, 22). Polyamides PA1 and PA2 were conjugated to JQ1 to generate Syn-TEF1 and Syn-TEF2, respectively (Fig. 1B, figs. S1 and S2, and table S1) (10, 23). Genome-wide binding profiles confirm that the linear polyamide, PA1 binds GAA repeats whereas the hairpin polyamide PA2 targets an unrelated sequence (9). The ability of Syn-TEFs to stimulate expression of endogenous *FXN* was examined in GM15850 cells, a FRDA patient-derived cell line (Fig. 1C). In this lymphoblastoid cell line, *FXN* levels are reduced by ~90% as compared to GM15851 cells from the patient's healthy sibling with fewer than 30 GAA repeats (Fig. 1C and fig. S4). In a dose-dependent manner, Syn-TEF1 restored *FXN* expression in FRDA cells to the levels observed in healthy cells (Fig. 2, A and B, and fig. S5B). Syn-TEF2, which does not target GAA repeats, did not activate *FXN* expression in either cell line, demonstrating the requirement for sequence-specific DNA targeting (Fig. 1C). *FXN* transcripts that are induced by Syn-TEF1 function, are spliced and translated to levels comparable to healthy cells (Fig. 1D).

To determine specificity of action in the context of the entire transcriptome, we performed RNA-seq in GM15850 and GM15851 cells (Fig. 1E, fig. S6, and tables S9 to S16). Remarkably, *FXN* was the transcript most significantly stimulated by Syn-TEF1 in the diseased cell line. The global transcriptome was minimally perturbed, with 11 genes differentially expressed more than 2-fold (Fig. 1E and tables S9 and S23). Milder Syn-TEF1 dependent perturbations, include 236 genes of which as many as 29 coincide with known FRDA expression networks (table S24) (24). Parallel treatment of the control GM15851 cells did not alter *FXN* expression while eliciting a comparably muted effect on the global transcriptome (Fig. 1E and table S10). This result starkly contrasts with the 4,091 genes whose levels are significantly perturbed by freely diffusing JQ1 (fig. S6E and table S11). Consistent with its anti-proliferative properties (21), freely diffusing JQ1 downregulates expression of the oncogene *c-MYC*. Since Syn-TEF1 targets JQ1 to specific genomic loci away from *c-MYC*, no change in *c-MYC* expression is observed in cells treated with this bifunctional molecule (Fig. 1E).

To determine whether Syn-TEF1 stimulates *FXN* by engaging the endogenous elongation machinery, we performed genome-wide chromatin immunoprecipitation analysis of BRD4, Pol II, and elongation-specific phospho-Ser2 marks on the largest subunit of Pol II. Given that polyamides bind to clustered sites in heterochromatin (9), we expect that Syn-TEF1 will localize to the first intron of *FXN* in disease cells that contain the GAA repeat expansion, but not in healthy cells (Fig. 2A and fig. S7). Consistent with the SOS profile of Syn-TEF1 binding, BRD4 levels are dramatically increased across the GAA repeat expansion in FRDA

cells (Fig. 2B). Since current algorithms remove sequencing reads that map to identical GAA repeats, the unannotated region (650 GAA repeats shown) is represented by a gap colored in blue (Fig. 2, A to D). Perhaps more striking is the profile of phospho-Ser2 marks placed on the productively elongating RNA Pol II (Fig. 2C and table S7). The peak of phospho-Ser2 enrichment is offset downstream of the BRD4 peaks, consistent with sequential action of P-TEFb and subsequent licensing of Pol II for productive transcription elongation. Due to the mechanistic coupling of transcriptional processes, phospho-Ser2 marks are retained until termination, well beyond the point of BRD4 recruitment by Syn-TEF1 (Fig. 2, B and C). Unexpectedly, a promoter-proximal BRD4 peak overlaps with the paused Pol II upstream of the GAA repeats (Fig. 2D and fig. S8). Upon treatment with Syn-TEF1, the decrease in paused Pol II coincides with the increase in Pol II levels within the body of *FXN*, thus furnishing evidence for licensing of productive elongation (Fig. 2D and table S6). Furthermore, tri- methylation of lysine 36 of histone H3 (H3K36me3), a signature of productive elongation, increases downstream of the GAA-repeats in Syn-TEF1 treated cells (Fig. 2E). In support of enhanced elongation, a downstream shift in tri-methylation of lysine 4 of histone H3 (H3K4me3), a promoter proximal chromatin mark, is also observed (Fig. 2E). In agreement with previous reports, we do not observe a dramatic increase of H3K4me3 marks at the promoter upon Syn-TEF1 treatment (11, 16). Consistent with a barrier to elongation, tethering proteins that stimulate transcriptional initiation fails to enhance *FXN* expression, whereas tethering VP16 or derivatives that can stimulate elongation, elicits modest *FXN* expression (25). As our results demonstrate, targeted recruitment of an elongation factor across the GAA repeats restores *FXN* expression in FRDA cells.

To further investigate the specificity of Syn-TEF1, we examined the enrichment of BRD4, Pol II and phospho-Ser2 at Syn-TEF1-targeted genomic loci. These loci are rank-ordered by their affinity for Syn-TEF1 (Fig. 3A and table S22). While BRD4 enrichment profile correlates with the polyamide binding profile (compare Fig. 3, A and B), neither phospho-Ser2 nor Pol II show any enrichment over a 10,000-base pair window centered on the polyamide-targeted genomic loci (Fig. 3B and fig. S10). Moreover, genome-wide binding of BRD4 is not perturbed by Syn-TEF1 whereas freely diffusing JQ1 dramatically reduces BRD4 occupancy en masse (Fig. 3C and fig. S10B). This observation is congruent with the minimal impact of Syn-TEF1 on global transcriptome profiles in healthy or diseased cells (Fig. 1E). Thus, Syn-TEF1 displays regulatory properties distinct from inhibitors of BRD4 that globally disrupt the transition to transcription elongation and elicit cell cycle arrest in cancer cells (21). We next mapped Syn-TEF1 targeted genomic loci to transcription start sites of proximal genes and compared SOS-scores with fold change in mRNA expression (Fig. 3D). Rather than proximity to TSS, the results suggest that genes with significant pausing of RNA Pol II (low licensing ratios -LRs) respond to Syn-TEF1 (Fig. 3E). Taken together, the results demonstrate that in the absence of paused or stalled Pol II, simply recruiting BRD4 to a genomic locus does not elicit transcriptional initiation. This conclusion is supported by a recent report that delivered JQ1 to two endogenous promoters and enriched BRD4 at those loci but did not report an increase in targeted gene expression from either locus (26). The dependence on paused Pol II, imposes mechanistic constraints on the

function of DNA tethered JQ1, and it serves as an invaluable specificity filter to limit Syn-TEF1 function to *FXN*, with minimal perturbation of the global transcriptome.

Next, we examined the impact of Syn-TEF1 on primary cells and cell lines derived from more than 20 FRDA patients with different genetic backgrounds and different GAA repeat expansions. In lymphoblastoid cells, fibroblasts and induced pluripotent cells derived from FRDA patients, Syn-TEF1 stimulated *FXN* expression while the control molecules or treatments did not (Fig. 4, A to E). To examine the ability of Syn-TEF1 to stimulate *FXN* expression in disease-relevant cell types, we differentiated GM23913 pluripotent cells to cardiomyocytes (27). Upon differentiation, cardiomyocytes expressed cardiac-specific markers and displayed rhythmic beating in culture (Fig. 4, B and D, figs. S11 and S12, and movie S1). Syn-TEF1 robustly stimulated *FXN* expression in these cells, whereas JQ1, with or without PA1, led to cytotoxicity (Fig. 4C). Similarly, neurons are particularly vulnerable to a reduction in *FXN* expression (28). Sensory neurons derived from three iPSC lines (Fig. 4F and fig. S13) displayed clear evidence for Syn-TEF1 responsive synthesis of processed mature FXN protein (Fig. 4E). In addition to cultured cells, primary peripheral blood mononuclear cells (PBMCs) obtained from 11 FRDA patients were genotyped and treated in parallel with Syn-TEF1. *FXN* expression was stimulated by Syn-TEF1 in all but one sample from the 11 FRDA patients (Fig. 4G).

To examine the future utility of Syn-TEF1 in restoring *FXN* levels in vivo, we transplanted human cells bearing a luciferase reporter fused in frame within the fifth exon of *FXN* into the flanks of immunocompromised mice (Fig. 4, H and I). Two HEK293 reporter cell lines were employed, the first containing only 6 GAA repeats and the second containing ~310 GAA repeats (29). Consistent with cell culture results (fig. S16) (29), we observe reduced *FXN-Luc* levels in transplanted cells bearing ~310 GAA repeats (Fig. 4, H and I). Upon Syn-TEF1 treatment, luciferase expression was stimulated in the cells with ~310 GAA repeats, nearly restoring levels to those seen in the reporter cell line with 6 repeats (Fig. 4, H and I). As a test of recovery of mitochondrial function, we observed ~90% recovery of aconitase activity in patient-derived lymphoblasts that were treated with Syn-TEF1 (Fig. 4J). Taken together, in multiple cell types from 20 FRDA patients with a broad range of repeat expansions and diverse genetic backgrounds, the prototype synthetic transcription elongation factor stimulated *FXN* expression and restored biological function.

Syn-TEF1 meets key design criteria, including the ability to: (i) target desired loci in the genome, (ii) access cognate sites in repressive chromatin, (iii) engage the endogenous elongation machinery, and (iv) license productive transcriptional elongation by a paused Pol II. In essence, our prototype Syn-TEF defines a general framework for the design of a class of molecules that could act on a diverse array of diseases caused by different stages of transcriptional dysfunction, especially at unstable microsatellite repeats (30). The growing mechanistic understanding of gene regulation, and how it goes awry in disease, offers new opportunities for intervention and remediation with precision-tailored synthetic molecules.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank members of the Ansari lab, especially Devesh Bhimsaria and members of the Raines lab, especially Thomas Smith, for helpful discussions. We thank Laura Vanderploeg and E. Nihal Korkmaz for help with figures, James Thomson, Ron Stewart, Jenifer Bolin, and Scott Swanson for help with RNA-seq, Anil Kumar, Shreyan Mohapatra, and Durdana Waseem for early experiments, Ricardo Dolmetsch and Ajamete Kaykas for facilitating the collaboration with the Neuroscience group at NIBR, and Samir Brahmachari and Anurag Agarwal for facilitating a collaboration with the Ataxia group at IGIB. The NIBR team thanks Marek Napierala (UAB), David Lynch (CHOP) and FARA (Friedreich's Ataxia Research Alliance) for making the UAB FA patient lines available for use. FR patient fibroblasts and derivative IPS, sensory neurons were obtained under a material transfer agreement with CHOP. We thank Richard Wade-Martins (Oxford) for the luciferase reporter cell lines. This work was supported by NIH grants CA133508, GM117362, HL099773 and the W. M. Keck Medical Research Award to A.Z.A. and NIH grant P30 AR066524 to D.N.S. G.S.E. was supported by NIH GM07215 and a Peterson Fellowship. M.P.G. was supported by a Hilldale scholarship. A.A. Is supported by an Indo-US postdoctoral fellowship from SERB, government of India. RNA-seq and ChIP-seq data are deposited in NCBI GEO GSE99403. A.Z.A., G.S.E., and M.P.G. have filed patent applications relating to the work in this manuscript, including U.S. provisional patent applications 62/478291 and 15/472852, filed 30 March 2016 by the Wisconsin Alumni Research Foundation (WARF).

REFERENCES AND NOTES

1. Eguchi A, Lee GO, Wan F, Erwin GS, Ans AZ. , Controlling gene networks and cell fate with precision-targeted DNA-binding proteins and small-molecule-based genome readers *Biochem. J.* 2014; 462:397–413. 10.1042/BJ20140400 Medline. [PubMed: 25145439]
2. Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: Emerging roles In metazoans *Nat. Rev. Genet.* 2012; 13:720–731. 10.1038/nrg.3293 Medline. [PubMed: 22986266]
3. Mayer A, Landry HM, Churchman LS. Pause & go: From the discovery of RNA polymerase pausing to Its functional Implications *Curr. Opin. Cell Biol.* 2017; 46:72–80. 10.1016/j.ceb.2017.03.002 Medline. [PubMed: 28363125]
4. Lee TI, Young RA. Transcriptional regulation and its misregulation In disease *Cell.* 2013; 152:1237–1251. 10.1016/j.cell.2013.02.014 Medline. [PubMed: 23498934]
5. Zhou Q, Li T, Price DH. RNA polymerase II elongation control *Annu. Rev. Biochem.* 2012; 81:119–143. 10.1146/annurev-biochem-052610-095910 Medline. [PubMed: 22404626]
6. Luo Z, Lin C, Shilatifard A. The super elongation complex (SEC) family In transcriptional control. *Nat. Rev. Mol. Cell Biol.* 2012; 13:543–547. 10.1038/nrm3417 Medline. [PubMed: 22895430]
7. Dervan PB. Molecular recognition of DNA by small molecules *Bioorg. Med. Chem.* 2001; 9:2215–2235. 10.1016/S0968-0896(01)00262-0 Medline. [PubMed: 11553460]
8. Carlson CD, Warren CL, Flauschild KE, Ozers MS, Qadir N, Bhimsaria D, Lee Y, Cerrina F, Ansari AZ. Specificity landscapes of DNA binding molecules elucidate biological function *Proc. Natl. Acad. Sci. U.S.A.* 2010; 107:4544–4549. 10.1073/pnas.0914023107 Medline. [PubMed: 20176964]
9. Erwin GS, Grieshop MP, Bhimsaria D, Do TJ, Rodríguez-Martínez JA, Mehta C, Khanna K, Swanson SA, Stewart R, Thomson JA, Ramanathan P, Ansari AZ. Synthetic genome readers target clustered binding sites across diverse chromatin states *Proc. Natl. Acad. Sci. U.S.A.* 2016; 113:E7418–E7427. 10.1073/pnas.1604847113 Medline. [PubMed: 27830652]
10. Burnett R, Melander C, Puckett JW, Son LS, Wells RD, Dervan PB, Gottesfeld JM. 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; 103:11497–11502. 10.1073/pnas.0604939103 Medline. [PubMed: 16857735]
11. Li Y, Lu Y, Polak U, Lin K, Shen J, Farmer J, Seyer L, Bhalla AD, Rozwadowska N, Lynch DR, Butler JS, Napierala M. Expanded GAA repeats impede transcription elongation through the FXN gene and induce transcriptional silencing that is restricted to the FXN locus *Hum. Mol. Genet.* 2015; 24:6932–6943. Medline. [PubMed: 26401053]
12. Yandim C, Natisvili T, Festenstein R. Gene regulation and epigenetics in Friedreich's ataxia *J. Neurochem.* 2013; 126:21–42. 10.1111/jnc.12254 Medline. [PubMed: 23859339]
13. Schulz JB, Boesch S, Bürk K, Dürr A, Giunti P, Mariotti C, Pousset F, Schöls L, Vankan P, Pandolfo M. Diagnosis and treatment of Friedreich ataxia: A European perspective *Nat. Rev. Neurol.* 2009; 5:222–234. 10.1038/nrneurol.2009.26 Medline. [PubMed: 19347027]

14. Filla A, De Michele G, Cavalcanti F, Pianese L, Monticelli A, Campanella G, Coccozza S. The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia *Am. J. Hum. Genet.* 1996; 59:554–560. Medline. [PubMed: 8751856]
15. Greene E, Mahishi L, Entezam A, Kumarl D, Usdin K. Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia *Nucleic Acids Res.* 2007; 35:3383–3390. 10.1093/nar/gkm271 Medline. [PubMed: 17478498]
16. Punga T, Bühler M. Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation *EMBO Mol. Med.* 2010; 2:120–129. 10.1002/emmm.201000064 Medline. [PubMed: 20373285]
17. Groh M, Lufino MMP, Wade-Martins R, Gromak N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome *PLOS Genet.* 2014; 10:e1004318. 10.1371/journal.pgen.1004318 Medline.
18. Perdomini M, Hick A, Puccio H, Pook MA. Animal and cellular models of Friedreich ataxia *J. Neurochem.* 2013; 126:65–79. 10.1111/jnc.12219 Medline. [PubMed: 23859342]
19. Soragni E, Miao W, Iudicello M, Jacoby D, De Mercanti S, Clerico M, Longo F, Piga A, Ku S, Campau E, Du J, Penalver P, Rai M, Madara JC, Nazor K, O'Connor M, Maximov A, Loring JF, Pandolfo M, Durelli L, Gottesfeld JM, Rusche JR. Epigenetic therapy for Friedreich ataxia *Ann. Neurol.* 2014; 76:489–508. 10.1002/ana.24260 Medline. [PubMed: 25159818]
20. Yang Z, Yik JHN, Chen R, He N, Jang MK, Ozato K, Zhou Q. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4 *Mol. Cell.* 2005; 19:535–545. 10.1016/j.molcel.2005.06.029 Medline. [PubMed: 16109377]
21. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, Morse EM, Keates T, Hickman TT, Felletar I, Philpott M, Munro S, McKeown MR, Wang Y, Christie AL, West N, Cameron MJ, Schwartz B, Heightman TD, La Thangue N, French CA, Wiest O, Kung AL, Knapp S, Bradner JE. Selective Inhibition of BET bromodomains *Nature.* 2010; 468:1067–1073. 10.1038/nature09504 Medline. [PubMed: 20871596]
22. Suto RK, Edayathumangalam RS, White CL, Melander C, Gottesfeld JM, Dervan PB, Luger K. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands *J. Mol. Biol.* 2003; 326:371–380. 10.1016/S0022-2836(02)01407-9 Medline. [PubMed: 12559907]
23. Viger A, Dervan PB. Exploring the limits of benzimidazole DNA-binding oligomers for the hypoxia Inducible factor (HIF) site *Bioorg. Med. Chem.* 2006; 14:8539–8549. 10.1016/j.bmc.2006.08.028 Medline. [PubMed: 16971129]
24. Coppola G, Burnett R, Perlman S, Versano R, Gao F, Plasterer H, Rai M, Saccá F, Filla A, Lynch DR, Rusche JR, Gottesfeld JM, Pandolfo M, Geschwind DH. A gene expression phenotype in lymphocytes from Friedreich ataxia patients *Ann. Neurol.* 2011; 70:790–804. 10.1002/ana.22526 Medline. [PubMed: 22162061]
25. Chapdelaine P, Coulombe Z, Chikh A, Gérard C, Tremblay JP. A Potential new therapeutic approach for Friedreich ataxia: Induction of frataxin expression with TALE proteins *Mol. Ther. Nucleic Acids.* 2013; 2:e119. 10.1038/mtna.2013.41 Medline.
26. Liszczak GP, Brown ZZ, Kim SH, Oslund RC, David Y, Muir TW. Genomic targeting of epigenetic probes using a chemically tailored Cas9 system *Proc. Natl. Acad. Sci. U.S.A.* 2017; 114:681–686. 10.1073/pnas.1615723114 Medline. [PubMed: 28069948]
27. Eguchi A, Wlekinski MJ, Spurgat MC, Heiderscheid EA, Kropornicka AS, Vu CK, Bhimsaria D, Swanson SA, Stewart R, Ramanathan P, Kamp TJ, Slukvin I, Thomson JA, Dutton JR, Ansari AZ. Reprogramming cell fate with a genome-scale library of artificial transcription factors *Proc. Natl. Acad. Sci. U.S.A.* 2016; 113:E8257–E8266. 10.1073/pnas.1611142114 Medline. [PubMed: 27930301]
28. Koeppen AH, Mazurkiewicz JE. Friedreich ataxia: Neuropathology revised *J. Neuropathol. Exp. Neurol.* 2013; 72:78–90. 10.1097/NEN.0b013e31827e5762 Medline. [PubMed: 23334592]
29. Lufino MMP, Silva AM, Németh AH, Alegre-Abarrategui J, Russell AJ, Wade-Martins R. A GAA repeat expansion reporter model of Friedreich's ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds *Hum. Mol. Genet.* 2013; 22:5173–5187. 10.1093/hmg/ddt370 Medline. [PubMed: 23943791]

30. Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, Daly MJ, Price AL, Pritchard JK, Sharp AJ, Erlich Y. Abundant contribution of short tandem repeats to gene expression variation in humans *Nat. Genet.* 2016; 48:22–29. 10.1038/ng.3461 Medline. [PubMed: 26642241]
31. Erwin GS, Bhimsaria D, Eguchi A, Ansari AZ. Mapping polyamide-DNA interactions in human cells reveals a new design strategy for effective targeting of genomic sites *Angew. Chem. Int. Ed.* 2014; 53:10124–10128. 10.1002/anie.201405497 Medline.
32. Baird EE, Dervan PB. Solid phase synthesis of polyamides containing imidazole and pyrrolic amino acids *J. Am. Chem. Soc.* 1996; 118:6141–6146. DOI: 10.1021/ja960720z
33. Frisan, T., Levitsky, V., Masucci, M., in *Epstein-Barr Virus Protocols*, Wilson, JB.; May, GHW., Eds. (Humana Press, 2001), pp. 125–127.
34. Roberts JM, Bradner JE. A bead-based proximity assay for BRD4 ligand discovery *Curr. Protoc. Chem. Biol.* 2015; 7:263–278. 10.1002/9780470559277.ch150024 Medline. [PubMed: 26629616]
35. Hou Z, Jiang P, Swanson SA, Elwell AL, Nguyen BKS, Bolin JM, Stewart R, Thomson JA. A cost-effective RNA sequencing protocol for large-scale gene expression studies *Sci. Rep.* 2015; 5:9570. 10.1038/srep09570 Medline. [PubMed: 25831155]
36. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome *Genome Biol.* 2009; 10:R25. 10.1186/gb-2009-10-3-r25 Medline. [PubMed: 19261174]
37. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome *BMC Bioinformatics.* 2011; 12:323. 10.1186/1471-2105-12-323 Medline. [PubMed: 21816040]
38. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, Haag JD, Gould MN, Stewart RM, Kendziorski C. EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments *Bioinformatics.* 2013; 29:1035–1043. 10.1093/bioinformatics/btt337 Medline. [PubMed: 23428641]
39. Chen EY, Tan CM, Kou Y, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool *BMC Bioinformatics.* 2013; 14:128. 10.1186/1471-2105-14-128 Medline. [PubMed: 23586463]
40. Orlando DA, Chen MW, Brown VE, Solanki S, Choi YJ, Olson ER, Fritz CC, Bradner JE, Guenther MG. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome *Cell Rep.* 2014; 9:1163–1170. 10.1016/j.celrep.2014.10.018 Medline. [PubMed: 25437568]
41. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities *Mol. Cell.* 2010; 38:576–589. 10.1016/j.molcel.2010.05.004 Medline. [PubMed: 20513432]
42. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: A flexible platform for exploring deep-sequencing data *Nucleic Acids Res.* 2014; 42:W187–W191. 10.1093/nar/gku365 Medline. [PubMed: 24799436]
43. Lian X, Hsiao C, Wilson G, Zhu K, Hazeltine LB, Azarin SM, Raval KK, Zhang J, Kamp TJ, Palecek SP. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling *Proc. Natl. Acad. Sci. U.S.A.* 2012; 109:E1848–E1857. 10.1073/pnas.1200250109 Medline. [PubMed: 22645348]
44. Blanchard JW, Eade KT, Szcs A, Lo Sardo V, Tsunemoto RK, Williams D, Sanna PP, Baldwin KK. Selective conversion of fibroblasts into peripheral sensory neurons *Nat. Neurosci.* 2015; 18:25–35. 10.1038/nn.3887 Medline. [PubMed: 25420069]

Fig. 1. Synthetic transcription elongation factors (Syn-TEFs) selectively activate *FXN* expression. (A) Cocystal structures of JQ1 bound to BRD4 (PDB 3MXF) and polyamide bound to nucleosomal DNA (PDB 1M1A). The distance allowing interaction of these complexes is estimated. (B) Linear PA1 and Syn-TEF1 target the DNA sequence 5'-AAGAAGAAG-3'. Hairpin PA2 and Syn-TEF2 target 5'-WTACGTW-3', where W = A or T. *N*-methylimidazole is bolded for clarity. *N*-methylpyrrole (open circle), *N*-methylimidazole (filled circle), 3-chlorothophene (square), and β -alanine (diamond) are represented in a ball and stick format. The structure of JQ1 linked to polyethylene glycol (PEG₆) is represented as a blue circle. (C) Relative expression of *FXN* mRNA in GM15850 (left panel) and GM15851 (right panel) cell lines by quantitative RT-PCR. Results are mean $\pm$ SEM ($n = 4$), normalized to relative expression of *FXN* in GM15851 cells (see also fig. S4). All treatments are 24 hours with 1 μ M of the indicated molecule, except DMSO (0.1%) and Syn-TEF1 (0.1, 0.5, or 1 μ M). * $P < 0.05$; ** $P < 0.01$. (D) Immunoblot of *FXN* and α -Tubulin (TUB) with treated GM15850 (left panel) and GM15851 (right panel) cells. Cells were treated as in (C). (E) Volcano plots of RNA-seq data display the change in global gene expression after 24 hours treatment of GM15850 (left panel) and GM15851 (right panel) cells with 1 μ M Syn-TEF1 ($n = 4$). Values represent the posterior probability of equal expression (PPEE) versus

fold-change in expression normalized to DMSO treated samples ($n = 4$). *FXN* and *c-Myc* are labeled red and blue, respectively.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 2.
Syn-TEF1 recruits BRD4 to its target sites and licenses productive Pol II elongation at *FXN*. All data are from GM15850 cells treated 24 hours with the indicated molecules. Signal traces are in reference-adjusted reads per million reads per base pair (rrpm/bp). **(A)** Summation of sites (SOS) profile of PA1 and Syn-TEF1 across the *FXN* locus. **(B)** BRD4 occupancy at the *FXN* locus. **(C)** Occupancy of phosphorylated serine 2 (phospho-Ser2) of the C-terminal domain of RNA Pol II at the *FXN* locus. **(D)** Occupancy of RNA Pol II at the 5' region of *FXN*. The grey bar identifies the location of the repressive GAA repeats and blue/cyan filled regions highlight unannotated reads but do not have defined quantitative values. **(E)** Occupancy of H3K4me3 and H3K36me3 measured at several locations within the *FXN* locus. Results are mean ($n = 3$) $\pm$ SEM. **(F)** A model of the cascade of interactions and reactions initiated by Syn-TEF1 at *FXN*.

Fig. 3.

Syn-TEF1 recruits BRD4 to its target sites and selectively activates *FXN*. All data are from GM15850 cells treated 24 hours with the indicated molecules (1 μ M). (A) Heatmap of the SOS profile of PA1 and Syn-TEF1 across the top 250 predicted binding sites of PA1 (9, 31). (B) Heatmaps of BRD4, Pol II, and phospho-Ser2 occupancy across the same genomic loci as in panel A. (C) Occupancy of BRD4 at BRD4 binding sites across the genome following treatment with Syn-TEF1 or PA1 + JQ1. (D) Scatterplot of the SOS score versus distance of the predicted Syn-TEF1 binding site to the transcription start site (TSS) for the top 500 genes predicted to be targeted by Syn-TEF1. Each gene is shaded according to the change in gene expression following Syn-TEF1 treatment. (E) Scatterplot of the SOS score, change in gene expression (Syn-TEF1 treatment), and licensing ratio (LR) of RNA Pol II for the top 500 genes predicted to be targeted by Syn-TEF1.

Fig. 4. Syn-TEFs activate *FXN* expression in primary patient cells and patient derived fibroblasts, iPSCs, cardiomyocytes, sensory neurons, and mouse xenografts. All treatments were 24 hours except where specified. **(A)** Relative expression of *FXN* mRNA, normalized to *GAPDH* in three lymphoblastoid cell lines (LCLs) derived from three different FRDA patients. All treatments were 1 μ M (results are mean $\pm$ SEM, $n = 3$). * $P < 0.05$; ** $P < 0.01$. **(B)** Expression of cell-type specific markers in GM23913 iPSCs or the iPSC-derived cardiomyocytes following treatment with 0.1% DMSO (results are mean $n =$

2). (C) Syn-TEF1 dependent induction of *FXN* mRNA in GM23913-derived cardiomyocytes (60 hours treatment, results are mean $n = 2$). (D) Immunohistochemistry of GM23913 iPSCs and the iPSC-derived cardiomyocytes. iPSCs were fixed and stained with OCT4 and SOX2. iPSC-derived cardiomyocytes were fixed and stained with TNNT2 and MYL2. Scale bars represent 100 μm . (E) Immunoblot of FXN and β -actin (β -act) following treatment of three different primary FRDA fibroblasts, fibroblast-derived iPSCs, and sensory neurons with the indicated molecules. Fibroblasts were collected from patients UAB4259 (550/1000), UAB4230 (1000/1200), and UAB66 (90/1025). Cells were treated 72–96 hours. (F) Immunohistochemistry of FRDA patient-derived iPSCs and iPSC-derived sensory neurons. iPSCs were fixed and stained with OCT4 and SSEA-4. Sensory neurons were fixed and stained with neuronal markers CGRP and MAP-2. Scale bars represent 100 μm . (G) (i) Genotyped repeats and (ii) relative expression of *FXN* mRNA normalized to *GAPDH* in peripheral blood mononuclear cells (PBMCs) from 11 patients following 24 hours treatment with 1 μM Syn-TEF1. (H) Bioluminescent images of two representative mice harboring xenografts (HEK293 *FXN-Luc* with 6 and ~310 GAA repeats in left and right flanks, respectively) (29). Mice were treated with either vehicle (DMSO) or 0.5 nmol Syn-TEF1 subcutaneously into each tumor (1 nmol total per mouse). Mice were imaged 22 hours after treatment. (I) Relative expression of *FXN-Luc* with 6 or ~310 GAA repeats following Syn-TEF1 treatment of mice as described in (H). Data are mean $\pm$ SEM ($n = 4$ and $n = 3$, Syn-TEF1- and DMSO-treated mice, respectively). $*P < 0.05$. (J) Aconitase activity in GM16214 lymphoblastoid cells following 72 hours treatment with DMSO (0.1%), PA1 (125 nM), or Syn-TEF1 (62.5 or 125 nM). Aconitase activity was normalized to GM16215 cells. Results are mean $\pm$ SEM ($n = 3$).