HHS Public Access

Author manuscript

*Cell Genom*. Author manuscript; available in PMC 2022 August 12.

Published in final edited form as:

*Cell Genom*. 2022 April 13; 2(4): . doi:10.1016/j.xgen.2022.100119.

# Genome-wide functional perturbation of human microsatellite repeats using engineered zinc finger transcription factors

**Y. Esther Tak**[1,2,3,6], **Gaylor Boulay**[1,3,4,6], **Lukuo Lee**[1], **Sowmya Iyer**[1], **Nicholas T. Perry**[1,2], **Hayley T. Schultz**[1,2], **Sara P. Garcia**[1], **Liliane Broye**[5], **Joy E. Horng**[1,2], **Shruthi Rengarajan**[1], **Beverly Naigles**[1], **Angela Volorio**[1,5], **Jeffry D. Sander**[1,2,3], **Jingyi Gong**[1,2], **Nicolò Riggi**[5,*], **J. Keith Joung**[1,2,3,*], **Miguel N. Rivera**[1,3,4,7,*]

[1]Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA, USA

[2]Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA, USA

[3]Department of Pathology, Harvard Medical School, Boston, MA, USA

[4]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[5]Institute of Pathology, Department of Experimental Pathology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1011 Lausanne, Switzerland

## SUMMARY

Repeat elements can be dysregulated at a genome-wide scale in human diseases. For example, in Ewing sarcoma, hundreds of inert GGAA repeats can be converted into active enhancers when bound by EWS-FLI1. Here we show that fusions between EWS and GGAA-repeat-targeted engineered zinc finger arrays (ZFAs) can function at least as efficiently as EWS-FLI1 for converting hundreds of GGAA repeats into active enhancers in a Ewing sarcoma precursor cell



[*]**Correspondence**: nicolo.riggi@chuv.ch (N.R.), jjoung@mgh.harvard.edu (J.K.J.), mnrivera@mgh.harvard.edu (M.N.R.).

[6]These authors contributed equally

[7]Lead contact

AUTHOR CONTRIBUTIONS

Y.E.T., G.B., J.K.J., N.R., and M.N.R. conceived and designed experiments. Y.E.T., G.B., L.L., L.B., H.T.S., N.T.P., J.E.H., S.R., B.N., A.V., and J.G. performed experiments. S.I. and S.P.G. implemented the computational pipelines for the analysis of RNA- and ChIP-seq data. G.B., S.P.G., and S.I. analyzed ChIP- and RNA-seq data. J.D.S. and J.K.J. engineered the 2-ZF units and ZFAs used in this study. Y.E.T., G.B., J.K.J., N.R., and M.N.R. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

J.K.J. has, or had during the course of this research, financial interests in several companies developing gene-editing or epigenetic-editing technology: Beam Therapeutics, Blink Therapeutics, Chroma Medicine, Editas Medicine, EpiLogic Therapeutics, ETx Inc., Excelsior Genomics, Hera Biolabs, Monitor Biotechnologies, Pairwise Plants, Poseida Therapeutics, SeQure Dx Inc., and Verve Therapeutics. J.K.J.'s interests were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict of interest policies. J.K.J. is a consultant for Beam Therapeutics, Chroma Medicine, ETx Inc., Pairwise Plants, SeQure Dx Inc., and Verve Therapeutics and is a scientific co-founder of these companies and of Blink Therapeutics, Editas Medicine, EpiLogic Therapeutics, Excelsior Genomics, and Monitor Biotechnologies. G.B., J.K.J., M.N.R., and Y.E.T. are inventors on a patent application that encompasses work described in this paper. J.K.J. and Y.E.T. are inventors on patents or patent applications describing various gene- and epigenetic-editing technologies. M.N.R. receives research support from ACD and Merck Serono for work unrelated to this study. S.I. and S.P.G. are employees of Verve Therapeutics. J.K.J. is a member of the Advisory Board of *Cell Genomics*.

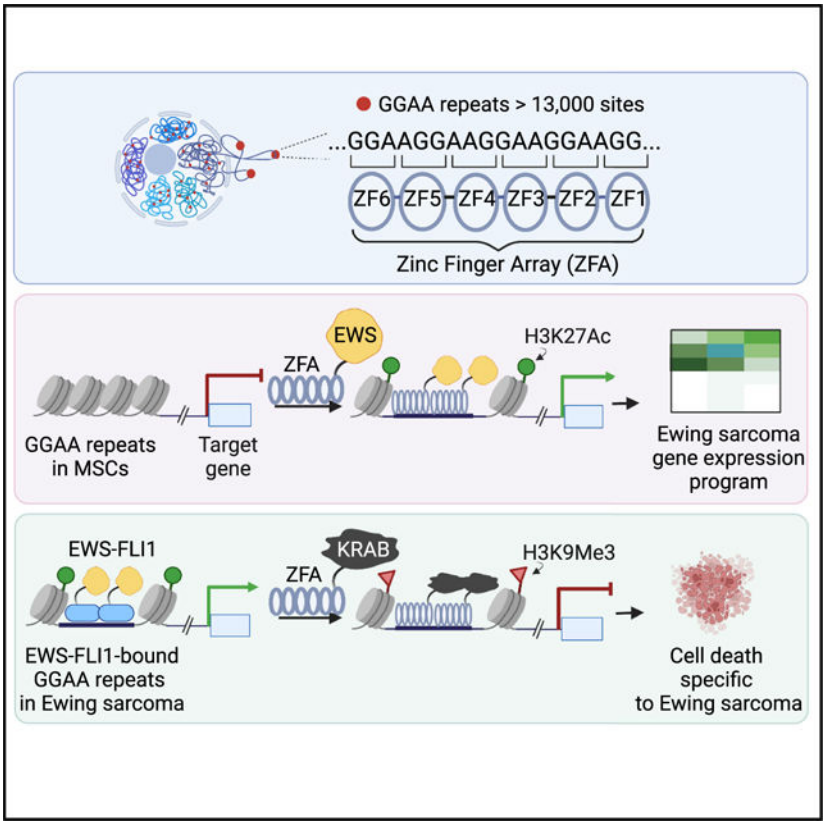
SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2022.100119.

model. Furthermore, a fusion of a KRAB domain to a ZFA can silence GGAA microsatellite enhancers genome wide in Ewing sarcoma cells, thereby reducing expression of EWS-FLI1-activated genes. Remarkably, this KRAB-ZFA fusion showed selective toxicity against Ewing sarcoma cells compared with non-Ewing cancer cells, consistent with its Ewing sarcoma-specific impact on the transcriptome. These findings demonstrate the value of ZFAs for functional annotation of repeats and illustrate how aberrant microsatellite activities might be regulated for potential therapeutic applications.

## In brief

Functionally annotating repetitive sequences in biology and disease remains challenging. Tak et al. show that engineered zinc finger array-based fusions can target microsatellites across the genome and regulate their activities in a cancer model, thereby elucidating insights into genome-wide function of these microsatellites and suggesting potential therapeutic strategies.




## INTRODUCTION

Following the success of large-scale sequencing and mapping efforts, the functional annotation of assembled genomes has emerged as one of the most important current scientific challenges in biology and medicine. This task has been accelerated and facilitated by the development of targetable gene- and epigenetic-editing tools that make it possible to probe the endogenous function of individual genes and regulatory sequences.[1,2] While the characterization of coding sequences has advanced at a rapid pace, the functional annotation

of non-coding regions presents significant challenges, particularly for highly repetitive sequences that constitute up to two-thirds of the human genome and that have been proposed to play critical roles in normal development and disease.[3–7] Recent studies have adapted transcription activator-like effector (TALE) and RNA-guided catalytically inactive or "dead" (dCas9)-based effectors to test the function of subclasses of transposable elements,[8,9] but many other categories of highly repetitive sequences await characterization.

We sought to develop strategies to functionally characterize repeat elements in the genome by first focusing on microsatellite repeats, a class of simple tandem repeats that previous studies have shown can be dysregulated in multiple disease states.[4,10–12] For example, large-scale epigenetic dysregulation of microsatellite repeats has been observed in Ewing sarcoma, a pediatric bone tumor where the EWS-FLI1 translocation fusion protein operates as a transcriptional pioneer factor.[13,14] This fusion includes both the N-terminal transactivation domain of EWS and the C-terminal DNA-binding domain of FLI1. In contrast to FLI1, which stably binds to only non-repeat GGAA sites, EWS-FLI1 can bind to both non-repeat GGAA motifs and GGAA microsatellite repeats. Notably, binding by EWS-FLI1 converts hundreds of GGAA microsatellites present throughout the human genome into transcriptional enhancers, thereby inducing a tumor-specific gene-regulatory program.[14–17] This example, together with the dysregulated expression of other repeat classes in other tumor types,[6,18] illustrates how aberrant transcriptional programs in cancer and other diseases can be caused by the widespread activation of specific repeat categories and highlights the need for robust tools to conduct genome-wide studies and perturbation of these elements.

Here we provide a proof-of-concept demonstration for how engineered ZFAs can be used to efficiently target and alter the chromatin state of a class of microsatellite repeats in human cells. Using GGAA microsatellite repeats bound by EWS-FLI1 as a test case, we show that engineered EWS-ZFA fusion proteins targeted to these repeats can be over an order of magnitude more efficient than an EWS-dCas9-targeted fusion for activating a GGAA repeat previously shown to be converted into a *de novo* enhancer by EWS-FLI1. In addition, EWS-ZFA fusions can effectively phenocopy the pioneer function of EWS-FLI1 at GGAA microsatellites and recapitulate the GGAA-repeat-dependent chromatin landscape and gene expression profiles of Ewing sarcoma. Remarkably, coupling of a GGAA-repeat-targeted ZFA to a transcriptional repressor KRAB domain results in genome-wide silencing of GGAA microsatellites and cytotoxicity that is selective for Ewing sarcoma cells through the targeted inactivation of oncogenic gene expression programs. Our results validate the power and efficacy of engineered zinc finger (ZF) technology for targeting and altering the functional state of microsatellite repeats and illustrate how this platform can be deployed to interrogate the function of microsatellite repetitive elements at genome scale.

## RESULTS

### Engineering sequence-specific DNA-binding domains to target GGAA microsatellite repeats

Although multiple platforms are available to create DNA-binding modules that might be capable of recognizing GGAA microsatellite repeats bound by the EWS-FLI1 fusion

protein, we chose to focus on using engineered ZFAs and RNA-guided dCas9 from *Streptococcus pyogenes*. Cys2His2 ZFAs can be engineered to recognize novel DNA sequences of interest and have been used successfully to build artificial transcription factors capable of influencing gene expression in human cells.[19,20,21,22] Alternatively, dCas9 programmed by guide RNAs (gRNAs) have similarly been used to create gene-regulatory proteins that function efficiently in human cells[23–25] and offer the substantial additional advantage of simple targetability by altering the gRNA sequence. TALE repeats have also been used to build customized DNA-binding domains, utilizing assembled arrays of four TALE-repeat domains as "building blocks," with each recognizing one of the four different DNA bases.[26–28] However, because the NN TALE repeat typically used to recognize guanine (G) has also been reported to recognize adenine (A) (albeit with less efficiency),[29–31] we elected not to engineer TALE-repeat arrays designed to recognize GGAA microsatellite repeat sequences.

We engineered ZFAs to recognize two 18 bp sequences that align within different registers of 4.5 GGAA repeats: 5′-GGA AGG AAG GAA GGA AGG and 5′-AAG GAA GGA AGG AAG GAA. A single ZF recognizes ~3 bp of DNA, and previous work has shown that highly active arrays of six ZFs that recognize 18 bp target sites can be assembled by using pre-selected 2-ZF units joined together by non-canonical (non-TGEKP) linkers such as TGSQKP or CGSQKP[32–40,22] (J.K.J.'s lab, unpublished data). Using this strategy and an archive of pre-selected 2-ZF units engineered to bind to various specific target sequences, we assembled eight different 6-ZF arrays for each of the two 18 bp target sites (Figure 1A) (STAR Methods and Table S1). To test the abilities of these 16 ZFAs to bind to GGAA microsatellite repeats, we fused the disordered prion-like N-terminal domain of EWSR1[17,41] (hereafter referred to as the EWS domain) to the N terminus or C terminus of each of the ZFAs (Figure 1B). We then assessed the abilities of each of these 32 fusions to activate the *UGT3A2* gene (an EWS-FLI1 target gene that has 11 GGAA repeats positioned ~2 kb upstream of its promoter) in human U2OS cells. We found that all of these ZF-based fusions activated *UGT3A2* with varying levels of efficiency (mean fold activation ranging from 14- to 190-fold) (Figure 1C). Because the ZF array ZFA7 exhibited approximately equivalent activity regardless of the position of the EWS domain (mean fold activation of 83- and 70-fold) and this level of activation was similar to that observed with EWS-FLI1 (mean fold activation of 121-fold) (Figure 1C), we selected the EWS-ZFA7 fusion (hereafter referred to as EWS-ZFA) for use in further experiments.

To enable binding of GGAA repeats by dCas9, we also designed a gRNA that would target a 23 bp sequence composed of 5.5 GGAA repeats: 5′-**AGG AAG GAA GGA AGG AAG GA**A GG, consisting of a 20 nt spacer (bold) and an NGG protospacer adjacent motif (underlined). To our surprise, expression of this gRNA together with a fusion protein in which the EWS domain was fused to the N-terminal or C-terminal end of dCas9 (hereafter referred to as EWS-dCas9 or dCas9-EWS, respectively) failed to activate the *UGT3A2* gene in U2OS cells (Figure 1D). To increase the number of EWS domains recruited by our dCas9-gRNA complex, we used single and multimerized configurations of two domains called DmrA and DmrC that only interact in the presence of a small-molecule A/C heterodimerizer. We co-expressed our gRNA and a DmrC-EWS domain fusion together

with a dCas9-DmrA fusion protein harboring one, two, three, or four DmrA domains in U2OS cells (Figure 1D). In the presence of heterodimerizer, we found that dCas9-DmrA fusions harboring two, three, or four DmrA domains could mediate modest activation of the *UGT3A2* gene (mean fold activation of 3.1, 3, or 1.1, respectively), at levels much lower compared with the activation observed using the EWS-ZFA fusion (Figure 1D). In contrast, these same dCas9-based EWS constructs were effective at activating various other genes when using different gRNAs directed to non-repetitive target sites within the promoters of those genes in U2OS cells and HEK293 cells (Figure S1). The ability of our dCas9-based EWS constructs to mediate activation from unique sites in the genome but not from GGAA repeats suggests that it may be challenging for these fusions to recognize and/or bind to these repeats that are present at over 13,000 loci in the human genome including the *UGT3A2* promoter. Taken together, our results show that an engineered EWS-ZFA fusion can more effectively activate an EWS-FLI1 target gene with upstream GGAA repeats than analogous dCas9-based fusions to the EWS domain.

## An EWS-ZFA fusion recapitulates genome-wide activation of microsatellite repeats observed in Ewing sarcoma

We next tested whether EWS-ZFA could target and activate GGAA microsatellites genome wide by comparing its activity to EWS-FLI1 in mesenchymal stem cells (MSCs). MSCs are a model for the cell of origin of Ewing sarcoma, and EWS-FLI1 has previously been shown to operate as a pioneer factor at GGAA repeats in these cells to induce a chromatin landscape and gene expression pattern similar to that of tumor cells.[14] Transduction of MSCs with lentiviral vectors expressing EWS-ZFA followed by chromatin immunoprecipitation sequencing (ChIP-seq) and unbiased sequence analysis identified GGAA repeats as the dominant motif found in EWS-ZFA peaks (more than 80% of EWS-ZFA-binding sites contained more than four consecutive GGAA units, Figures 2A and S2A). EWS-ZFA also bound nearly all of the GGAA repeats in the genome bound by EWS-FLI1 (Figures 2B and S2B). We categorized 13,029 GGAA microsatellites with more than four consecutive GGAA units based on their length (Figure S2C) and found that the EWS-ZFA binds a higher fraction of GGAA microsatellites (10%–20%) than EWS-FLI1 at each length interval (Figures 2C and S2D).

We further tested whether EWS-ZFA binding would lead to the induction of active chromatin states at GGAA repeats in MSCs. To this end, we used ChIP-seq to measure the active chromatin mark H3K27ac at 812 GGAA microsatellites that are consistently bound by endogenous EWS-FLI1 in Ewing sarcoma cell lines.[14] We observed strong binding of EWS-ZFA at these same sites and de novo deposition of H3K27ac, often at higher levels than that induced by EWS-FLI1 (Figures 2D, 2E, and S2E). This higher activity and the higher binding of GGAA repeats by EWS-ZFA compared with EWS-FLI1 (Figures 2C and S2D) may be due to higher protein expression levels observed upon lentiviral induction (Figure S2F). However, we cannot rule out other possible explanations such as structural and functional differences between the ZFA and FLI1 DNA-binding domains (which may provide distinct DNA stability profiles to each fusion protein) or differing numbers of fusion proteins recruited to a given GGAA repeat (which may result in variable recruitment of chromatin co-factors involved in H3K27ac deposition). By contrast, canonical non-repeat

GGAA sites bound by EWS-FLI1 showed no evidence of EWS-ZFA binding or chromatin state changes, thereby demonstrating the specificity of the engineered EWS-ZFA fusion for GGAA repeats relative to non-repeat GGAA sites as expected (Figures 2F, 2G, and S2G). In addition to changes in chromatin activity, we also measured transcriptional changes for genes in the vicinity of EWS-FLI1-bound GGAA repeats. Transcript analysis showed that 72% of the genes that are within 100 kb of EWS-FLI1-bound GGAA repeats (Figure 2H and Table S2) and are induced ≥2-fold by EWS-FLI1 were also upregulated to a similar degree by EWS-ZFA. Taken together, these data show that EWS-ZFA is able to phenocopy the chromatin and transcriptional activation observed in Ewing sarcoma, suggesting that localizing the N-terminal EWS domain to GGAA repeats via an engineered ZFA instead of the FLI1 DNA-binding domain is sufficient to initiate the recruitment of chromatin regulators required for pioneer function, enhancer activation, and target gene expression. These results provide an important proof of concept for how engineered ZFAs can be an effective tool to target and alter the functional state of GGAA microsatellites genome wide.

## KRAB-ZFAs can selectively silence the microsatellitedriven Ewing sarcoma gene expression program

Given that EWS-ZFA can efficiently target and activate GGAA microsatellites in MSCs, we hypothesized that a fusion of our engineered ZFA to a repressive KRAB domain[42,43] might conversely silence active GGAA microsatellites bound by endogenous EWS-FLI1 in Ewing sarcoma cells, thereby inactivating its downstream oncogenic gene expression program. This approach offers the possibility to delineate the precise functional role of GGAA repeats in Ewing sarcoma cells, in isolation from the non-repeat GGAA target sites of EWS-FLI1. We expressed a KRAB-ZFA fusion protein and found that it bound efficiently to GGAA microsatellites in two Ewing sarcoma cell lines (SKNMC and A673). Interestingly, KRAB-ZFA binding was followed by EWS-FLI1 eviction from the same genomic sites, as assessed by FLI1 ChIP-seq performed in SKNMC and A673 cells (FLI1 ChIP-seq can be used to detect the binding of EWS-FLI1 because these two cell lines do not express endogenous wildtype FLI1) (Figures 3A, 3B, S3A, and S3B). KRAB-ZFA binding was also associated with striking changes in chromatin states and the induction of repressive marks with increased H3K9me3 and decreased H3K27ac signals at GGAA microsatellites (Figures 3A, 3C, 3D, S3A, and S3C). As expected, these changes were observed uniquely at GGAA repeats and not at non-repeat GGAA EWS-FLI1-binding sites, confirming the specificity of KRAB-ZFA (Figures S3D–S3F). Among the genes located within 100 kb of EWS-FLI1-bound GGAA repeats that showed ≥2-fold decreases in EWS-FLI1-depleted cell lines, 49% and 47% showed a similar decrease due to KRAB-ZFA expression in SKNMC and A673 cells, respectively (Figures 3G and S4A; Table S3). Genes involved in specific functional categories (e.g., cell-cycle regulation and neurogenesis) that have previously been identified after EWS-FLI1 knockdown[14] and are linked to Ewing sarcoma cell survival were enriched among the genes downregulated by KRAB-ZFA in SKNMC and A673 cells (Figure S4B; Tables S4 and S5).

Because the KRAB-ZFA fusion would only be expected to alter the function of GGAA repeats in Ewing sarcoma cells in which the EWS-FLI1 is expressed (and in which these repeats function as enhancers), we were interested in evaluating the effects of KRAB-ZFA

expression in non-Ewing sarcoma cells. To this end, we analyzed genome-wide chromatin state changes in HEK293T cells upon expression of KRAB-ZFA. Similar to what has been observed in most non-Ewing cell types previously examined,[14] we found that HEK293T cells were largely devoid of active chromatin marks at GGAA repeats and that there were no major changes in H3K27ac signals induced with KRAB-ZFA expression (Figure S4C). However, GGAA repeats in HEK293T cells accumulated strong repressive H3K9me3 signals after expression of KRAB-ZFA in the same manner as Ewing sarcoma cells (Figures 3E and 3F). In contrast to Ewing sarcoma cells, HEK293T transduced with KRAB-ZFA displayed minimal transcriptional changes, which only included a handful of genes with GGAA repeats located within their promoters (Figure S4D and Table S5).

Finally, we tested whether the selective antagonistic effect exerted by the KRAB-ZFA fusion on the EWS-FLI1-induced transcriptional program in Ewing sarcoma cells might also translate into a cell-type-specific impact on cell viability. To this end, we quantitatively compared the viability of four different Ewing sarcoma cell lines with four non-Ewing sarcoma control lines upon the expression of KRAB-ZFA or GFP (as a negative control) (Figure 3H). Strikingly, despite similar KRAB-ZFA protein expression levels (Figure S4E), only the viability of Ewing sarcoma cells was affected by KRAB-ZFA, with a reduction exceeding 80%, whereas minimal toxicity was observed in all negative control non-Ewing sarcoma cell lines (Figure 3H).

## DISCUSSION

Our results show that engineered ZFAs are highly effective and specific tools for targeting widely distributed repetitive elements and altering their chromatin states. Engineered ZFAs have distinct advantages for this purpose given their high DNA-binding affinities, small size, and similarities to endogenous transcription factors.[44] Our findings further demonstrate that engineered ZFAs can greatly facilitate the functional assessment of the important but challenging-to-study repetitive elements of the human genome and may provide a strategy for therapeutically modifying the non-coding function of these repeats. Targeting repeat elements outside of non-coding regions with ZFAs may also have therapeutic value, as suggested by a recent study targeting a single CAG repeat expansion in the coding sequence of the Huntington disease gene, *HTT*.[45] Future studies may thus develop ZFAs that target other repeats, whose expansion or dysfunction may alter a number of biological processes including gene regulation, RNA stability, and altered protein function.[11] In addition, our approach provides tools to investigate the function of other microsatellite elements that are widely distributed such as telomeric repeats, interstitial telomeric sequences implicated in genome stability,[46] or specific classes of simple repeats that are enriched at promoters and regulatory elements.[47]

In the case of GGAA microsatellites that are activated genome wide in Ewing sarcoma, the high degree of specificity conferred by ZFAs allowed us to isolate their function and determine that these elements are in fact responsible for large-scale gene activation in this tumor type. By recruiting specific regulatory domains without the involvement of endogenous DNA-binding proteins, engineered ZFAs also make it possible to study the contribution of poorly understood proteins, as shown by our finding that the N terminus of

EWSR1 is sufficient to activate GGAA microsatellites in the absence of the erythroblast transformation-specific (ETS) DNA-binding domain contained in EWS-FLI1.

Intriguingly, we observed large differences between ZFA and RNA-guided dCas9 approaches for targeting GGAA repeats. Although future studies are needed to investigate the underlying causes of this, we speculate that these differences might be linked to: (1) higher binding affinity of EWS-ZFA for GGAA repeats and/or higher expression of EWS-ZFA compared with dCas9-EWS (perhaps owing to protein size differences), which in turn lead to higher occupancy of GGAA-repeat sites; (2) formation of R loops by dCas9-EWS that might interfere with the recruitment of other endogenous transcription factors;[48] and (3) potentially higher binding steric hindrance constraints imposed by nucleosomes at closed GGAA repeats on dCas9 fusions compared with their ZF counterparts.[49] Taken together, these observations suggest that ZFAs may have advantages over dCas9 for studying the function of tandem repeats that occur at a high number of different locations in the human genome.

Finally, the significant toxicity observed when silencing GGAA repeats in Ewing sarcoma cells compared with other cell types indicates that the function of repeat elements may be highly cell-type dependent. This points to the importance of systematically evaluating the role of repeats in normal and pathogenic states and their potential as targets for altering cell behavior in a robust and specific manner.

### Limitations of the study

Our study presents proof-of-concept experiments showing that ZFAs can be used to alter the chromatin states of microsatellite repeats genome wide. While the methodologies utilized can in principle be applied more generally to other repetitive elements, further experiments will be required to test the effectiveness of this approach in other contexts. Our results are also notable for showing significant advantages of ZFAs compared with dCas9-based technologies. Further studies will be needed to determine whether these differences hold true for other genomic repeats and to elucidate the mechanistic basis for these observations. Finally, we demonstrated that repression of GGAA repeats by KRAB-ZFA leads to selective toxicity in Ewing sarcoma cell lines, suggesting the possibility of a therapeutic application for our findings. Further development of this concept will require extensive *in vivo* validation and the development of effective approaches for the delivery of KRAB-ZFA to solid tumors.

## STAR★METHODS

### RESOURCE AVAILABILITY

**Lead contact—**Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Miguel N. Rivera (mnrivera@mgh.harvard.edu).

**Materials availability—**Plasmids generated in this study have been deposited to Addgene.

**Data and code availability—**Raw western blot data from Figures S2 and S4 were deposited on Mendeley at http://doi.org/10.17632/zbrd8fgdn5.1. Datasets from ChIP-seq and RNA-seq experiments have been deposited in the Gene Expression Omnibus (GEO) repository with the accession number GSE163886. This paper does not report original code. Scripts for ChIP-seq and RNA-seq analysis will be made available upon request. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

**Cell lines—**Primary bone marrow derived-MSCs were collected with approval from the Institutional Review Board of the Center Hospitalier Universitaire Vaudois. Samples were de-identified prior to our analysis. MSCs were cultured in IMDM (Life Technologies) containing 10% fetal calf serum (FCS) and 10 ng/mL platelet-derived growth factor BB (PeproTech). U2OS were obtained from Toni Cathomen (Freiburg). All other cell lines were obtained from ATCC and media from Life Technologies. Ewing sarcoma cell lines SKNMC, A673, EW7 were grown in RPMI 1640 and CHP100 in McCoy's 5a Medium. HEK293T, Hela and U2OS were grown in DMEM and MRC5 in EMEM. All media were supplemented with 1% penicillin and streptomycin (Life Technologies). McCoy's 5a medium was supplemented with 15% FBS and all other media were supplemented with 10% FBS. Cells were cultured at 37°C with 5%$CO_2$. Media supernatant was analysed biweekly for the presence of Mycoplasma using MycoAlertTM PLUS (Lonza). Cell lines were authenticated by ATCC STR profiling.

## METHODS DETAILS

**Plasmids and oligonucleotides—**Each of the 16 different ZFAs that recognize ~4.5 GGAA tandem repeats was generated by assembling pre-selected 2-ZF units from an unpublished Joung lab archive. Although we used an unpublished archive of engineered zinc finger modules to provide the various 2-ZF units for constructing our ZFAs, there are other published public sources of zinc finger units as well as protocols that can be used to create customized zinc finger arrays.[32–37] The assembled ZFAs were inserted into the pENTR3C vector and EWS N-terminus[14] or KRAB (from BPK1407) was cloned into pENTR3C-ZFAs by Gibson assembly. The EWS-ZFA or KRAB-ZFA fusions thus generated were transferred to lentiviral pLIV vector containing EF1-alpha promoter via LR reactions using Gateway LR clonase II Enzyme Mix (Invitrogen). dCas9-EWS (NP173) was constructed by cloning EWS into BPK1179 digested with *XhoI* and Not1 by Gibson assembly, and EWS-dCas9 (YET3486) was constructed by cloning EWS into dCas9 (pSQT847) digested with *AgeI* and BstZ17i by Gibson assembly. DmrC-EWS was generated by inserting EWS into DmrC entry vector digested with *NruI*, using Gibson assembly. Sequences of gRNAs used in this study are provided in Table S5.

**Transfection—**For EWS-ZFA experiments in U2OS cells, $2 \times 10^5$ cells were transfected with 1ug of plasmids by nucleofection using the DN-100 program on a Lonza 4-D Nucleofector with the SE Cell Line Kit (Lonza) and transfected cells were plated in 24-well plates. For dCas9-based EWS constructs, we used the nucleofection method described in detail previously.[50]

**Lentiviral generation—**Lentivirus was produced in HEK293T LentiX cells (Clontech) by LT1 (Mirus Bio) transfection with gene delivery vector and packaging vectors GAG/POL and VSV plasmids.[17] Viral supernatants were collected 72 h after transfection and concentrated using the LentiX concentrator (Clontech). Virus containing pellets were resuspended in PBS and added dropwise on cells in presence of growth media supplemented with six ug/ml polybrene. Cells infected with lentivirus were selected using puromycin (Invivogen) at a concentration of one ug/ml for SKNMC, EW7, CHP100, HEK293T, HeLa and U2OS or two ug/ml for A673 and MRC5 in the growth medium. MSCs were selected with 0.75 ug/ml puromycin. Overexpression efficiency was determined by immunoblot analysis.

**Immunoblot analysis—**Immunoblot analyses were performed using standard protocols.[17] Primary antibodies were used at the following concentrations: rat anti-HA (Roche, 1ug/ml), rabbit anti-FLI1 (abcam, 1ug/ml), and mouse anti-GAPDH (Millipore, 0.1 ug/ml). Secondary antibodies were goat anti-rabbit, goat anti-rat, and goat anti-mouse IgG respectively conjugated with horseradish peroxidase (Bio Rad, 1: 10,000 dilution). Membranes were developed using Western Lightning Plus-ECL enhanced chemiluminescence substrate (PerkinElmer) and visualized using photographic film.

**Real-time quantitative reverse transcription PCR—**Total RNA was extracted from the transfected cells 72 h post-transfection using the NucleoSpin RNA Plus (Clontech), and 250 ng of purified RNA was used for cDNA synthesis in 20ul of total reaction using High-Capacity RNA-cDNA kit (ThermoFisher). cDNA was diluted 1:20 and 3 μL of cDNA was used for qPCR using SYBR Green Real-Time PCR Master Mix (ThermoFisher), and primers specific for the target transcript (Table S5). qPCR was performed using Roche LightCycler480 with the following cycling protocols: initial denaturation at 95°C for 20 s (s) followed by 45 cycles of 95°C for 3 s and 60°C for 30 s. Ct values over 35, were considered as 35. Relative quantification of each target, normalized to an endogenous control (*GAPDH or HPRT1*), was performed using the comparative Ct method (Applied Biosystems).

**Cell viability assays—**Cells that were transduced with lentiviral KRAB-ZFA plasmid or GFP control plasmid were grown for 8 days and cell viability was measured using the CellTiter-Glo luminescent assay (Promega) as described by the manufacturer. Endpoint luminescence was measured on a SpectraMax M5 plate reader (Molecular Devices).

**Definition of target genes associated with EWS-FLI1 bound GGAA-repeats—**In Figure 2H, 126 GGAA repeat-associated genes were selected based on a maximum distance of 100 kb from EWS-FLI1 bound GGAA repeats (n = 812) and upregulation upon EWS-FLI1 induction in MSCs (greater than 2-fold). In Figure 3H, 235 GGAA repeat-associated genes were selected based on a maximum distance of 100 kb from EWS-FLI1 bound GGAA repeats (n = 812) and downregulation upon EWS-FLI1 knockdown in both SKNMC and A673 Ewing sarcoma cell lines (greater than 2-fold).[14]

**ChIP-seq—**ChIP assays of MSCs, SKNMC, A673 and HEK293T cells were carried out using two to five x $10^6$ cells per sample and per epitope, following the procedures described previously.[51] In brief, chromatin from formaldehyde-fixed cells were fragmented to 200–700

bp with a Branson 250 sonifier. Solubilized chromatin was immunoprecipitated overnight at 4C with 3 mg of target specific antibodies (rat anti-HA (Roche), rabbit anti-FLI1 (Abcam), rabbit anti-H3K27ac (Active Motif), and rabbit anti-H3K9me3 (Abcam)). Antibody-chromatin complexes were pulled down with protein G-Dynabeads (Life Technologies), washed, and then eluted. After crosslink reversal, RNase A, and proteinase K treatment, immunoprecipitated DNA was extracted with AMP Pure beads (Beckman Coulter). ChIP DNA was quantified with Qubit. Sequencing libraries were prepared with 1–5 ng of ChIP DNA samples and input samples using the Ovation Ultralow System V2 kit (Nugen). Libraries were sequenced with single-end (SE) 50–75 cycles on an Illumina Nextseq 500 Illumina genome analyzer.

**ChIP-seq bioinformatic analysis—**Reads were aligned to human reference genome hg19 using bwa.[52] Aligned reads were then filtered to exclude PCR duplicates and were extended to 200 bp to approximate fragment sizes. Density maps were generated by counting the number of fragments overlapping each position using igvtools, and normalized to 10 million reads. We used MACS2[53] to call peaks using matching input controls with a q-value threshold of 0.01. Peaks were filtered to exclude blacklisted regions as defined by the ENCODE consortium.[54] Peaks within 200 bp of each other were merged. Genomewide GGAA microsatellite repeats were previously annotated.[17,55] Peak intersections were identified using bedtools.[56] Average ChIP-seq signals across intervals were calculated using bwtool.[57] findMotifsGenome.pl was used to identify *de novo* DNA motifs between 8 and 20 bp from all sites bound by EWS-ZFA with the Homer suite of tools.[58] Signals shown in heatmaps (100 bp windows) and composite plots (10 bp window) were calculated using bwtool.[57] Heatmap signals are in log2 scale, centered around EWS-FLI1 binding sites[14] and are capped at the 99th percentile.

**RNA-seq—**Total RNA was isolated from cells using NucleoSpin RNA Plus (Clontech). For Figure 2H, RNA libraries were prepared from 500 ng of total RNA treated with Ribogold zero to remove rRNA using TruSeq Stranded Total RNA Library Prep Gold kit (Illumina, 20,020,599) and TruSeq RNA Single Indexes. The RNA libraries were sequenced with PE 32 cycles on an Illumina Nextseq500 system. For Figure 3G, RNA samples were sent to Novogene Corporation for mRNA sequencing. RNA libraries were sequenced with PE150 cycles on an Illumina NovaSeq 6000 system. Reads were aligned to hg19 using STAR.[59] Mapped reads were filtered to exclude PCR duplicates and reads mapping to known rRNA coordinates, obtained from the rmsk table in the UCSC database (http://genome.ucsc.edu). Gene expression was calculated using featureCounts.[60] Only primary alignments with mapping quality of 10 or more were counted. Counts were then normalized to one million reads. Signal tracks were generated using bedtools.[56] Differential expression was calculated using DESeq2.[61]

**GSEA analysis—**Gene set overlaps were computed using Gene Set Enrichment Analysis (GSEA, https://www.gsea-msigdb.org/gsea/msigdb/annotate.jsp). Genes lists for GSEA analysis were selected using a log2 fold change of 0.6 for upregulated genes and −0.6 for downregulated genes. An adjusted p value threshold of <0.1 was also applied. Gene

lists were then analyzed for overlaps with C2 (curated gene sets) and BP (GO biological process), with an FDR q-value < 0.05.

### QUANTIFICATION AND STATISTICAL ANALYSIS

Information on the number of biological replicates, statistical tests and p values is provided in the figure legends.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

J.K.J. was supported by grants from the National Institutes of Health (R35 GM118158, R01 CA211707, R01 CA204954, DP1 OD006862, and DP1 GM105378), a Massachusetts General Hospital (MGH) Collaborative Center for X-Linked Dystonia-Parkinsonism grant, the Jim and Ann Orr MGH Research Scholar Award, and the Desmond and Ann Heathwood MGH Research Scholar Award. N.R. is supported by the Swiss National Science Foundation Professorship grant (PP00P3-157468/1 and PP00P3_183724), the Swiss Cancer League grant (KFS-3973-08-2016 and KFS4859-08-2019), the Fond'Action Contre le Cancer grant, the FORCE grant, and the Fondation Muschamp grant. M.N.R. is supported by NCI/NIH (U54 CA231637) and the Thomas F. and Diana L. Ryan MGH Research Scholar Award. We thank Elizabeth Dahlborg for help with ZF engineering. We thank Ligi Paul Pottenplackel, Bradley E. Bernstein, Martin Aryee, and Rui Dong for comments on the manuscript. Graphical abstract was created with BioRender.com.

## REFERENCES

1. Sander JD, and Joung JK (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol 32, 347–355. [PubMed: 24584096]
2. Holtzman L, and Gersbach CA (2018). Editing the epigenome: reshaping the genomic landscape. Annu. Rev. Genom. Hum. Genet 19, 43–71.
3. Payer LM, and Burns KH (2019). Transposable elements in human genetic disease. Nat. Rev. Genet 20, 760–772. [PubMed: 31515540]
4. Usdin K (2008). The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. Genome Res. 18, 1011–1019. [PubMed: 18593815]
5. Pehrsson EC, Choudhary MNK, Sundaram V, and Wang T (2019). The epigenomic landscape of transposable elements across normal human development and anatomy. Nat. Commun 10, 5640. [PubMed: 31822674]
6. Ting DT, Lipson D, Paul S, Brannigan BW, Akhavanfard S, Coffman EJ, Contino G, Deshpande V, Iafrate AJ, Letovsky S, et al. (2011). Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science 331, 593–596. [PubMed: 21233348]
7. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860–921. [PubMed: 11237011]
8. Jachowicz JW, Bing X, Pontabry J, Boskovic A, Rando OJ, and Torres-Padilla M-E (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat. Genet 49, 1502–1510. [PubMed: 28846101]
9. Fuentes DR, Swigut T, and Wysocka J (2018). Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. Elife 7, e35989.
10. Subramanian S, Mishra RK, and Singh L (2003). Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biol. 4, R13. [PubMed: 12620123]
11. Malik I, Kelley CP, Wang ET, and Todd PK (2021). Molecular mechanisms underlying nucleotide repeat expansion disorders. Nat. Rev. Mol. Cell Biol 22, 589–607. [PubMed: 34140671]

12. Trost B, Engchuan W, Nguyen CM, Thiruvahindrapuram B, Dolzhenko E, Backstrom I, Mirceta M, Mojarad BA, Yin Y, Dov A, et al. (2020). Genome-wide detection of tandem DNA repeats that are expanded in autism. Nature 586, 80–86. [PubMed: 32717741]
13. Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, Kovar H, Joubert I, de Jong P, and Rouleau G (1992). Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature 359, 162–165. [PubMed: 1522903]
14. Riggi N, Knoechel B, Gillespie SM, Rheinbay E, Boulay G, Suvà ML, Rossetti NE, Boonseng WE, Oksuz O, Cook EB, et al. (2014). EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. Cancer Cell 26, 668–681. [PubMed: 25453903]
15. Gangwal K, Sankar S, Hollenhorst PC, Kinsey M, Haroldsen SC, Shah AA, Boucher KM, Watkins WS, Jorde LB, Graves BJ, et al. (2008). Microsatellites as EWS/FLI response elements in Ewing's sarcoma. Proc. Natl. Acad. Sci. U S A 105, 10149–10154. [PubMed: 18626011]
16. Guillon N, Tirode F, Boeva V, Zynovyev A, Barillot E, and Delattre O (2009). The oncogenic EWS-FLI1 protein binds in vivo GGAA microsatellite sequences with potential transcriptional activation function. PLoS One 4, e4932. [PubMed: 19305498]
17. Boulay G, Sandoval GJ, Riggi N, Iyer S, Buisson R, Naigles B, Awad ME, Rengarajan S, Volorio A, McBride MJ, et al. (2017). Cancer-specific retargeting of BAF complexes by a prion-like domain. Cell 171, 163–178.e19. [PubMed: 28844694]
18. Burns KH (2017). Transposable elements in cancer. Nat. Rev. Cancer 17, 415–424. [PubMed: 28642606]
19. Gräslund T, Li X, Magnenat L, Popkov M, and Barbas CF 3rd. (2005). Exploring strategies for the design of artificial transcription factors: targeting sites proximal to known regulatory regions for the induction of gamma-globin expression and the treatment of sickle cell disease. J. Biol. Chem 280, 3707–3714. [PubMed: 15537646]
20. Beerli RR, Segal DJ, Dreier B, and Barbas CF (1998). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc. Natl. Acad. Sci. U S A 95, 14628–14633. [PubMed: 9843940]
21. Liu PQ, Rebar EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, et al. (2001). Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. J Biol Chem 276, 11323–11334. [PubMed: 11145970]
22. Tan S, Guschin D, Davalos A, Lee YL, Snowden AW, Jouvenot Y, et al. (2003). Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. Proc Natl Acad Sci U S A 100, 11997–12002. [PubMed: 14514889]
23. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, and Lim WA (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183. [PubMed: 23452860]
24. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, and Joung JK (2013). CRISPR RNA-guided activation of endogenous human genes. Nat. Methods 10, 977–979. [PubMed: 23892898]
25. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods 10, 973–976. [PubMed: 23892895]
26. Scholze H, and Boch J (2011). TAL effectors are remote controls for gene activation. Curr. Opin. Microbiol 14, 47–53. [PubMed: 21215685]
27. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, and Bonas U (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326, 1509–1512. [PubMed: 19933107]
28. Moscou MJ, and Bogdanove AJ (2009). A simple cipher governs DNA recognition by TAL effectors. Science 326, 1501. [PubMed: 19933106]
29. Streubel J, Blücher C, Landgraf A, and Boch J (2012). TAL effector RVD specificities and efficiencies. Nat. Biotechnol 30, 593–595. [PubMed: 22781676]
30. Deng D, Yan C, Pan X, Mahfouz M, and Wang J (2012). Structural Basis for Sequence-specific Recognition of DNA by TAL Effectors (Science).

31. Christian ML, Demorest ZL, Starker CG, Osborn MJ, Nyquist MD, Zhang Y, Carlson DF, Bradley P, Bogdanove AJ, and Voytas DF (2012). Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. PLoS One 7, e45383.
32. Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, and Dobbs D (2010). ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res. 38, W462–W468. [PubMed: 20435679]
33. Fu F, Sander JD, Maeder M, Thibodeau-Beganny S, Joung JK, Dobbs D, Miller L, and Voytas DF (2009). Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. Nucleic Acids Res. 37, D279–D283. [PubMed: 18812396]
34. Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, Fu F, Porteus MH, Dobbs D, Voytas DF, et al. (2006). Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nat. Protoc 1, 1637–1652. [PubMed: 17406455]
35. Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y, et al. (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat. Methods 8, 67–69. [PubMed: 21151135]
36. Maeder ML, Thibodeau-Beganny S, Sander JD, Voytas DF, and Joung JK (2009). Oligomerized pool engineering (OPEN): an “open-source” protocol for making customized zinc-finger arrays. Nat. Protoc 4, 1471–1501. [PubMed: 19798082]
37. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, et al. (2008). Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol. Cell 31, 294–301. [PubMed: 18657511]
38. Joung JK, Voytas DF, and Kamens J (2015). Accelerating research through reagent repositories: the genome editing example. Genome Biol. 16, 255. [PubMed: 26585970]
39. Pearson H (2008). Protein engineering: the fate of fingers. Nature 455, 160–164. [PubMed: 18784697]
40. Moore M, Klug A, and Choo Y (2001). Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. Proc. Natl. Acad. Sci. U S A 98, 1437–1441. [PubMed: 11171969]
41. Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, Darzacq X, et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science 361, eaar2555.
42. Margolin JF, Friedman JR, Meyer WK, Vissing H, Thiesen HJ, and Rauscher FJ 3rd. (1994). Krüppel-associated boxes are potent transcriptional repression domains. Proc. Natl. Acad. Sci. U S A 91, 4509–4513. [PubMed: 8183939]
43. Groner AC, Meylan S, Ciuffi A, Zangger N, Ambrosini G, Dénervaud N, Bucher P, and Trono D (2010). KRAB–Zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS Genet. 6, e1000869.
44. Gersbach CA, Gaj T, and Barbas CF 3rd. (2014). Synthetic zinc finger proteins: the advent of targeted gene regulation and genome modification technologies. Acc. Chem. Res 47, 2309–2318. [PubMed: 24877793]
45. Zeitler B, Froelich S, Marlen K, Shivak DA, Yu Q, Li D, Pearl JR, Miller JC, Zhang L, Paschon DE, et al. (2019). Allele-selective transcriptional repression of mutant HTT for the treatment of Huntington’s disease. Nat. Med 25, 1131–1142. [PubMed: 31263285]
46. Aksenova AY, and Mirkin SM (2019). At the beginning of the end and in the middle of the beginning: structure and maintenance of telomeric DNA repeats and interstitial telomeric sequences. Genes 10, 118.
47. Sawaya S, Bagshaw A, Buschiazzo E, Kumar P, Chowdhury S, Black MA, and Gemmell N (2013). Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. PLoS One 8, e54710.
48. Gao X, Tsang JCH, Gaba F, Wu D, Lu L, and Liu P (2014). Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. Nucleic Acids Res. 42, e155. [PubMed: 25223790]

49. Yarrington RM, Verma S, Schwartz S, Trautman JK, and Carroll D (2018). Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. Proc. Natl. Acad. Sci. U S A 115, 9351–9358. [PubMed: 30201707]
50. Tak YE, Kleinstiver BP, Nuñez JK, Hsu JY, Horng JE, Gong J, Weissman JS, and Joung JK (2017). Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. Nat. Methods 14, 1163–1166. [PubMed: 29083402]
51. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP, et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560. [PubMed: 17603471]
52. Li H, and Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. [PubMed: 19451168]
53. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. 9, R137. [PubMed: 18798982]
54. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. [PubMed: 22955616]
55. Boulay G, Volorio A, Iyer S, Broye LC, Stamenkovic I, Riggi N, and Rivera MN (2018). Epigenome editing of microsatellite repeats defines tumor-specific enhancer functions and dependencies. Genes Dev. 32, 1008–1019. [PubMed: 30042132]
56. Quinlan AR, Clark RA, Sokolova S, Leibowitz ML, Zhang Y, Hurles ME, Mell JC, and Hall IM (2010). Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Res. 20, 623–635. [PubMed: 20308636]
57. Pohl A, and Beato M (2014). bwtool: a tool for bigWig files. Bioinformatics 30, 1618–1619. [PubMed: 24489365]
58. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. [PubMed: 20513432]
59. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]
60. Liao Y, Smyth GK, and Shi W (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. [PubMed: 24227677]
61. Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. [PubMed: 25516281]

## Highlights

- Engineered ZFA fusions can efficiently target microsatellites across the human genome
- EWS-ZFA recapitulates the genome-wide activation of GGAA repeats seen in Ewing sarcoma
- KRAB-ZFA silences GGAA repeats and induces selective toxicity in Ewing sarcoma cells
- ZFA fusions are tools for functional annotation and therapeutic targeting of repeats

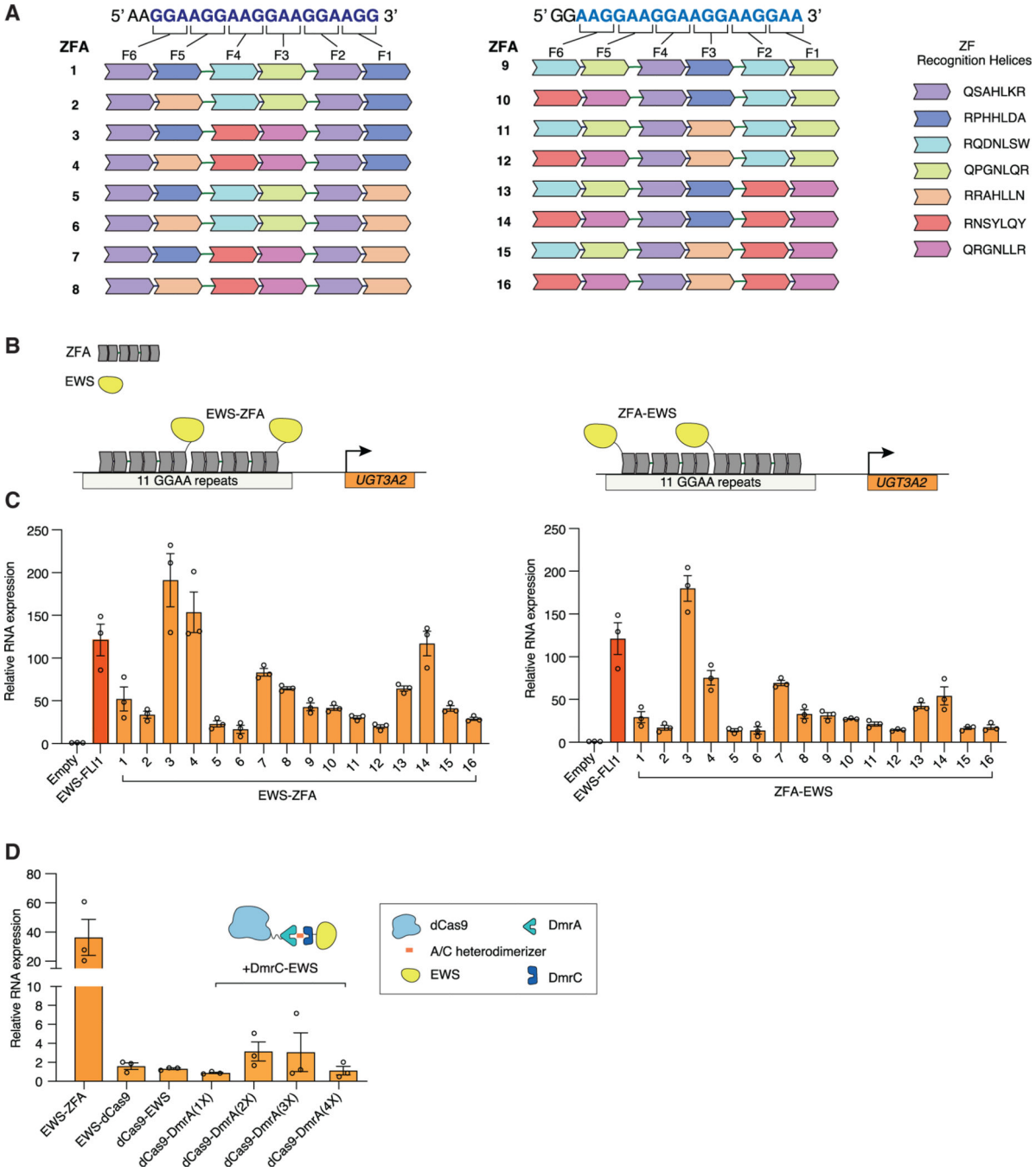



**Figure 1. Engineering ZFAs to bind GGAA microsatellites in the human genome and efficient activation of a target gene by engineered ZFAs fused to EWS**

(A) Schematic of 16 ZFAs, each engineered to bind ~4.5 GGAA microsatellites. The ZFAs have six zinc fingers, and each finger recognizes three nucleotides. The target sequences of ZFA 1 through 8 start with GGA, and ZFA 9 through 16 with AAG. The amino acid compositions of recognition helices for each zinc finger are shown on the right. Multiple zinc fingers with different recognition helices can recognize the same nucleotides.

(B) Schematic of ZFAs fused to EWS activating *UGT3A2* by binding to an 11-unit GGAA microsatellite located ~2 kb upstream of the transcription start site. EWS is fused to the N terminus (left) or C terminus (right) of ZFAs.
(C) Thirty-two fusions of EWS and ZFAs that target GGAA repeats were tested for *UGT3A2* gene activation in U2OS cells by nucleofection. EWS-ZFA7 closely mimicked the activation level of EWS-FLI1 and therefore was selected for further experiments.
(D) mRNA expression of *UGT3A2* in U2OS cells nucleofected with EWS-ZFA7, EWS-dCas9, dCas9-EWS, or dCas9-based bipartite EWS activators (dCas9-DmrA and DmrC-EWS). The bipartite system increases the density of EWS molecules recruited to a target site.
See also Figure S1 and Table S1.

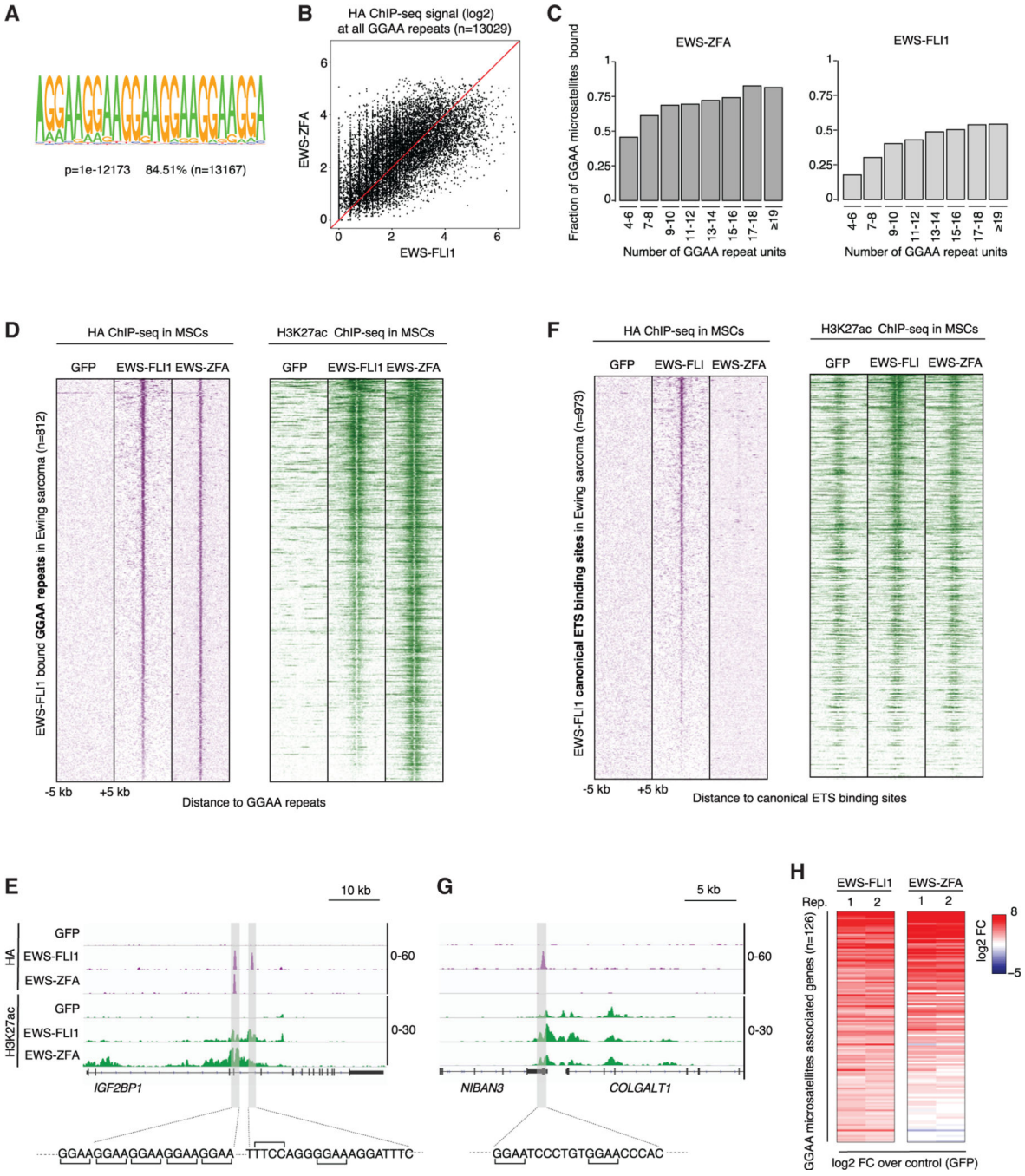



**Figure 2. Efficient and specific binding of EWS-ZFA at GGAA repeats in MSCs induces active chromatin and activation of GGAA-repeat-associated genes**

(A) GGAA-repeat motifs identified at sites bound by EWS-ZFA in MSCs.

(B) Scatterplot showing binding of 3xHA-tagged EWS-FLI1 and EWS-ZFA to GGAA repeats genome wide (n = 13,029) in MSCs determined using hemagglutinin (HA) ChIP-seq. ChIP-seq signals are on a $\log_2$ scale. The Spearman correlation coefficient is 0.68 with $p < 2.2 \times 10^{-16}$. Data from one of two biological replicate experiments is shown.

(C) Bar plots showing the fraction of GGAA repeats in the genome bound by EWS-ZFA (left) and EWS-FLI1 (right) upon lentiviral transduction in MSCs. Data from one of two

biological replicate experiments is shown. The number of consecutive GGAA repeats in each category is shown on the x axis.

(D) Heatmaps showing HA and H3K27ac ChIP-seq signals in MSCs at EWS-FLI1-bound GGAA repeats identified in Ewing sarcoma (n = 812) upon lentiviral transduction of either 3xHA-tagged EWS-FLI1 or EWS-ZFA. 3xHA-tagged GFP was used as control. 10-kb windows in each panel are centered on EWS-FLI1 binding sites in Ewing sarcoma.

(E) Example showing the binding of 3xHA-tagged EWS-FLI1 or EWS-ZFA and accompanying H3K27ac levels in MSC at the *IGF2BP1* locus containing a GGAA repeats element and a canonical ETS binding site.

(F) Heatmaps showing HA and H3K27ac ChIP-seq signals in MSCs at EWS-FLI1-bound canonical ETS binding sites identified in Ewing sarcoma (n = 973) upon lentiviral transduction of either 3xHA-tagged EWS-FLI1 or EWS-ZFA. GFP was used as control. 10-kb windows in each panel are centered on EWS-FLI1 binding sites in Ewing sarcoma.

(G) Example showing the binding of 3xHA-tagged EWS-FLI1 or EWS-ZFA and accompanying H3K27ac levels in MSC at the *NIBAN3* and *COLGALT1* loci containing a canonical ETS binding site.

(H) Heatmaps of $\log_2$ fold changes in expression of GGAA-repeat-associated genes (n = 126) in MSCs treated with EWS-FLI1 or EWS-ZFA constructs compared with a GFP control, determined by RNA sequencing (RNA-seq). Two biological replicates are shown. Spearman correlation of $\log_2$ fold changes in EWS-FLI1 and EWS-ZNF is 0.58 ($p < 2.22 \times 10^{-16}$).

See also Figure S2 and Table S2.

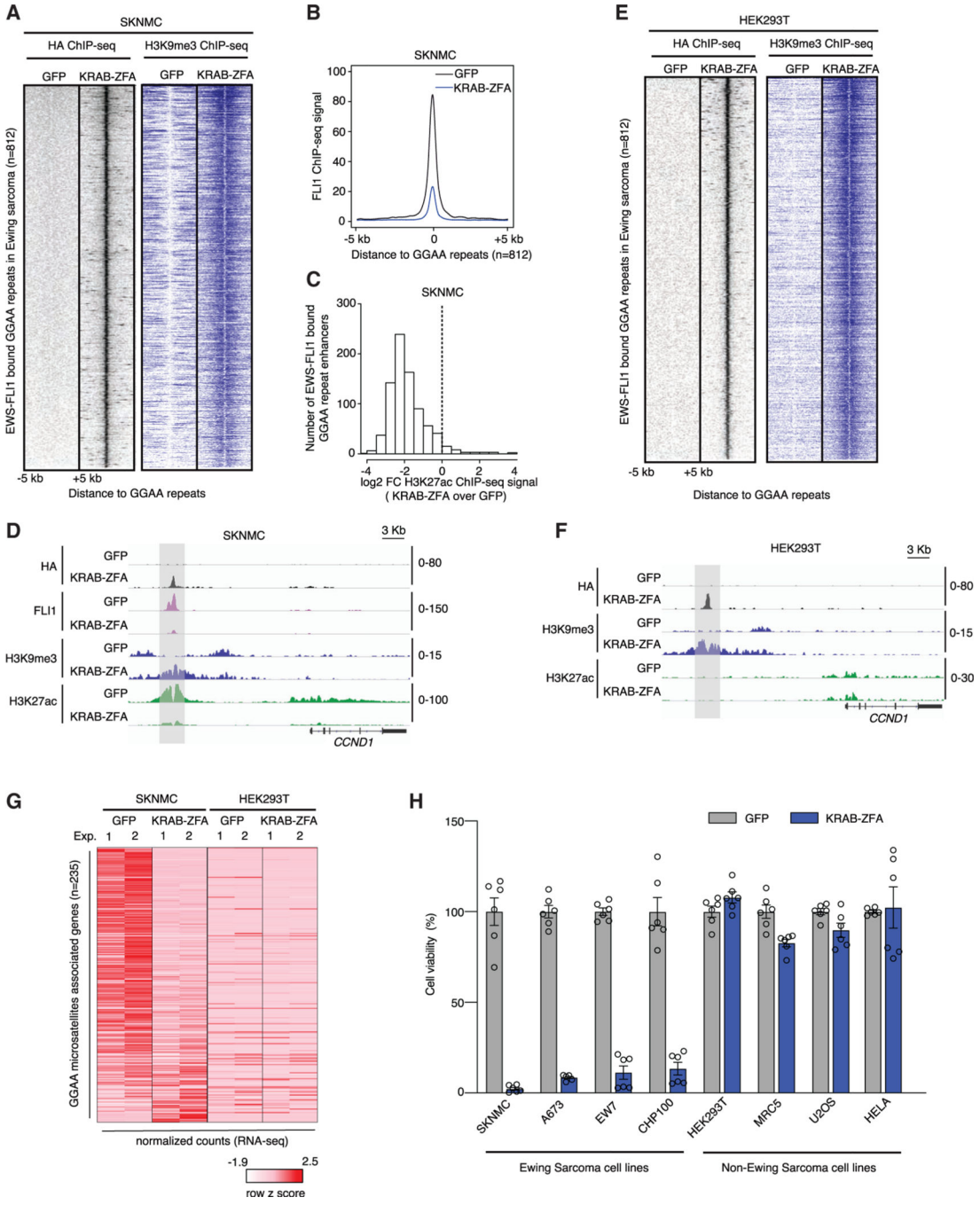



**Figure 3. Binding of KRAB-ZFA to GGAA repeats induces selective toxicity in Ewing sarcoma cell lines by repressing target gene expression**

(A) Heatmaps showing binding of 3xHA-tagged KRAB-ZFA and H3K9me3 deposition at EWS-FLI1 bound GGAA repeats (n = 812) in SKNMC cells as determined using ChIP-seq.

(B) Composite plot showing EWS-FLI1 occupancy of GGAA repeats after introduction of KRAB-ZFA or GFP (control) in SKNMC. The x axis represents a 10-kb window centered on 812 GGAA repeats.

(C) Histograms showing changes in H3K27ac at 812 EWS-FLI1-bound GGAA repeats upon treatment of SKNMC cells with KRAB-ZFA.

(D) Example showing the binding of KRAB-ZFA (3xHA tagged), endogenous EWS-FLI, H3K9me3, and H3K27ac, a GGAA-repeat element associated with the *CCND1* locus, after treatment of SKNMC cells with KRAB-ZFA constructs. GFP was used as control.
(E) Heatmaps showing binding of KRAB-ZFA (3xHA tagged) and H3K9me3 deposition in HEK293T cells at GGAA repeats bound by EWS-FLI1 in Ewing sarcoma (n = 812) as determined using ChIP-seq.
(F) Example showing the binding of KRAB-ZFA (3xHA tagged), H3K9me3, and H3K27ac at a GGAA-repeat element associated with the *CCND1* locus, after treatment of HEK293T cells with KRAB-ZFA construct. GFP was used as control.
(G) Heatmaps showing expression (row-normalized counts) of GGAA-repeat-associated genes (n = 235) in SKNMC and HEK293T cells treated with KRAB-ZFA or GFP (control) determined by RNA-seq. Data are from two biological replicates.
(H) Viability of Ewing sarcoma and non-Ewing cell lines 8 days post lentiviral transduction of KRAB-ZFA and GFP (control). Open circles indicate two biological replicates with three technical replicates; error bars show the SEM.
See also Figures S3 and S4; Tables S3–S5.

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Rabbit Anti-FLI1 Polyclonal Antibody, Unconjugated | Abcam | Cat# ab15289; RRID: AB_301825 |
| Anti-HA High Affinity; Rat monoclonal antibody (clone 3F10) | Roche | Cat# 11867423001; RRID: AB_390918 |
| Mouse Anti-Glyceraldehyde-3-PDH (GAPDH) Monoclonal antibody, Unconjugated | Millipore | Cat# MAB374; RRID: AB_2107445 |
| Rabbit Anti-Histone H3, trimethyl (Lys9) ChIP Grade Polyclonal Antibody, Unconjugated | Abcam | Cat# ab8898; RRID: AB_306848 |
| Histone H3K27ac antibody | Active Motif | Cat# 39134; RRID: AB_2722569 |
| Goat Anti-Rabbit IgG (H L)-HRP Conjugate antibody | Bio Rad | Cat# 1706515; RRID: AB_11125142 |
| Goat Anti-Mouse IgG (H L)-HRP Conjugate antibody | Bio Rad | Cat# 1706516; RRID: AB_11125547 |
| Goat anti-Rat IgG (H + L) Secondary Antibody, HRR | Invitrogen | Cat# 62–9520; RRID: AB_2533965 |
| Chemicals, peptides, and recombinant proteins | | |
| HALT PROTEASE AND PHOSPHATASE inhibitors | Pierce | Cat#RI78445 |
| Dynabeads® Protein G | LIFE TECHNOLOGIES | Cat#10004D |
| TranslT-LTI Transfection Reagent | MIRUS BIO LLC | Cat#MIR 2305 |
| Polybrene (HEXADIMETHRINE BROMIDE) | Sigma-Aldrich | Cat#H9268–50G |
| Puromycin | FISHER SCIENTIFIC | Cat#NC9138068 |
| Western Lightning Western Blot Chemiluminescence Reagent Plus | PERKINELMER | Cat#NEL104001EA |
| AUTORAD BLUE FILM | FISHER SCIENTIFIC | Cat#NC9648989 |
| RNase A | ROCHE | Cat#11119915001 |
| Rroteinase K | LIFE TECHNOLOGIES | Cat#25530049 |
| LentiX concentrator | Takara | Cat#631232 |
| AMP Pure beads | Beckman Coulter | Cat#A63881 |
| Critical commercial assays | | |
| TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold Set A | Illumina | Cat#RS-122–2301 |
| TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold Set B | Illumina | Cat#RS-122–2302 |
| Nextera DNA Sample Preparation Kit | Illumina | Cat#FC-121–1030 |
| Nucleospin RNA Plus | Clontech | Cat#740984.50 |

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Celltiter-Glo Luminescent cell viability assay | Promega | Cat#07570 |
| fast SYBR Green Master Mix | ThermoFisher | Cat#4385618 |
| High Capacity cDNA reverse transcription kit | ThermoFisher | Cat#4387406 |
| Ovation Ultralow System V2 kit | Nugen | Cat#0344NB-A01 |
| SE Cell Line Kit | Lonza | Cat# V4SC-1096 |
| Deposited data | | |
| Western blot (Figures S2 and S4) | This study | Mendeiey Data: http://doi.org/10.17632/zbrd8fgdn5.1 |
| ChIR-seq and RNA-seq data | This study | GEO: GSE163886 |
| Experimental models: Cell lines | | |
| HEK293 | ATCC | CRL-1573 |
| A673 | ATCC | CRL-1598 |
| SKNMC | ATCC | HTB-10 |
| U2OS | ATCC | HTB-96 |
| HEK293-T | ATCC | CRL-11268 |
| HEK293-T LentiX | Clontech | Cat#632180 |
| Human pediatric Mesenchymal Stem Cells | This study | N/A |
| EW7 | Lausanne | Submitted sample, STRB5754 (EW7), is not a match to any cell line in either the ATCC or DSMZ STR database. However, the profile for the submitted sample is an exact match to the STR profile for the (EW-7) cell line that is listed on the ExPASy website |
| MRC5 | ATCC | CCL-171 |
| HELA | ATCC | Submitted sample, STRB5759 (HeLa), is a similar match to ATCC cell line CCL-2 (HeLa) |
| CHP100 | Lausanne | Submitted sample, STRB5755 (CHP100), shows similarities to ATCC cell line CRL-5918 (NCI-H2073) however the cell lines appear to be unrelated, see addendum. The profile for the submitted sample is a similar match to the STR profile for the (CHP100) cell line that is listed on the ExPASy website |
| Oligonucleotides | | |
| See Table S6 | | N/A |
| Recombinant DNA | | |

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| pLiv-3xHA-EWS-FLI1 | This paper | YET3443 |
| pLiv-3xHA-GFP | This paper | YET3442 |
| pLiv-3xHA-EWS-ZFA | This paper | YET3444 |
| pLiv-3xHA-ZFA-EWS | This paper | YET3445 |
| pLiv-3xHA-KRAB-ZFA | This paper | YET3446 |
| pCAG-EWS-3XHA-dCas9 | This paper | YET3486 |
| pCAG-dCas9–3xHA-EWS | This paper | NP173 |
| pCAG-dCas9-DmrA (x1- x4) | This paper | BPK1019, BPK1033, BPK1140, BPK1179 |
| pCAG-DmrC-EWS | This paper | YET3386 |
| Software and algorithms | | |
| Prism | Version 9 | N/A |
| Bwa version 0.7.12-r1039 | Li and Durbin, 2009 | https://github.com/lh3/bwa |
| Bwtool version 1.0 | Pohl and Beato, 2014 | https://github.com/CRG-Barcelona/bwtool |
| STAR version STAR_2.4.0h | Dobin et al., 2013 | https://github.com/alexdobin/STAR |
| R | version 3.6.2 | https://cran.r-project.org |
| IGV | IGV_2.3.60 | https://github.com/igvteam/igv |
| HOMER version 4.7 | Heinz et al., 2010 | http://homer.salk.edu/homer/ |
| DESeq2 version 1.26.0 | Love et al., 2014 | https://bioc.ism.ac.jp/packages/3.1/bioc/html/DESeq2.html |
| Bedtools version v2.26.0 | Quinlan and Hall, 2010 | https://github.com/arq5x/bedtools |
| featureCounts Version 1.5.0-p2 | Liao et al., 2014 | http://subread.sourceforge.net |
| MASC2 version macs2 2.0.10.20120913 | Zhang et al., 2008 | https://github.com/taoliu/MACS |