Charles University

Faculty of Science

Study programme: Molecular Biology and Biochemistry of Organisms

Adam Vobruba

Mammalian proteins carrying zinc finger arrays

Savčí proteiny se sadou zinkových prstů

Bachelor's thesis

Supervisor: prof. Mgr. Petr Svoboda, Ph.D.

Prague, 2024

Declaration:

I declare that I have written this thesis on my own and have appropriately cited all sources and literature used. This thesis, or any significant portion of it, has not been submitted to obtain any other academic degree.

In Prague, 30th April 2024

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 30. 4. 2024

Adam Vobruba

Acknowledgement:

My thanks go to prof. Petr Svoboda for his patient guidance and great advice given to me during the time this thesis has been taking shape. I am also very grateful to all my close ones for their ongoing support. Without them, this work could not have been achieved.

Poděkování:

Mé díky patří prof. Petru Svobodovi za jeho trpělivé vedení a dobré rady, které mi poskytoval v průběhu tvorby této práce. Velmi vděčný jsem také všem svým blízkým za jejich neutuchající podporu. Bez nich by tato práce nevznikla.

Abstract:

A zinc finger is a small peptide motif stabilised by a single zinc ion, best known for its capability to specifically bind a 3-nucleotide sequence of DNA, depending on the exact amino acids present in the DNA-binding positions. Zinc fingers are unique in their ability to freely link together and form longer tandem arrays, which can bind DNA targets of any length and sequence determined by the combination of individual fingers. These arrays can easily mutate and be rebuilt to change binding specificity, allowing great flexibility and helping zinc fingers to their widespread presence in numerous endogenous proteins of various functions. This property of zinc finger arrays also made them a suitable tool for the creation of custom DNA-binding domains for genetic engineering. This thesis provides an overview of the discovery, structure and function of these domains and then reviews and discusses selected naturally occurring mammalian zinc finger proteins and their properties, showcasing diverse uses zinc finger arrays have been adapted for throughout evolution. The history and future of zinc fingers in artificial proteins created for gene therapy and research are discussed as well.

Keywords: zinc finger, ZnF, KRAB, KZFP, CTCF, PRDM9, ZFN, mammals

Abstrakt:

Zinkový prst je malý peptidový motiv stabilizovaný jedním zinečnatým iontem, známý zejména pro svou schopnost specificky vázat třínukleotidovou sekvenci DNA v závislosti na konkrétních aminokyselinách přítomných v DNA-vazebných pozicích. Zinkové prsty jsou unikátní svou schopností spojovat se do delších tandemově uspořádaných sad, které mohou vázat DNA cíle o libovolné délce a sekvenci, určené kombinací jednotlivých prstů. Tyto sady mohou jednoduše měnit vazebnou specificitu díky mutacím a přestavbám, což jim dodává velkou flexibilitu a pomohlo zinkovým prstům rozšířit se do mnoha endogenních proteinů o rozličných funkcích. Tato vlastnost sad zinkových prstů je také činí vhodnými pro tvorbu DNA-vazebných domén na míru pro využití v genovém inženýrství. Tato práce podává přehled o objevu těchto domén, jejich struktuře a funkci a poté shrnuje a diskutuje vybrané přirozeně se vyskytující savčí proteiny se zinkovými prsty a jejich vlastnosti, představujíc různé funkce pro něž byly sady zinkových prstů v průběhu evoluce adaptovány. Diskutována je také minulost a budoucnost zinkových prstů v umělých proteinech vytvořených pro genovou terapii a výzkum.

Klíčová slova: zinkový prst, ZnF, KRAB, KZFP, CTCF, PRDM9, ZFN, savci

Glossary

ATF – artificial transcription factor

Cas9 – CRISPR-associated protein 9

CRISPR – clustered regularly interspaced short palindromic repeats

CTCF – CCCTC-binding factor

DNA – deoxyribonucleic acid

DSB – double-strand break

dsDNA – double-stranded DNA

DUF3669 – domain of unknown function 3669

ERV – endogenous retrovirus

ESC – embryonic stem cell

EXAFS – extended X-ray absorption fine structure

HAT – histone acetyltransferase

HIV – human immunodeficiency virus

HMT – histone methyltransferase

HP1 – heterochromatin protein 1

HR – homologous recombination

KAP1 – KRAB-associated protein 1

KRAB – Krüppel-associated box

KZFP – KRAB-zinc finger protein

LINE-1/L1 – long interspersed nuclear element

LTR – long terminal repeat

MHC II – major histocompatibility complex II

mRNA – messenger RNA

NHEJ – non-homologous end joining

NMR – nuclear magnetic resonance

NuRD – nucleosome remodelling and deacetylase complex

PRDM9 – PR/SET domain-containing protein 9

RNA- ribonucleic acid

RNP – ribonucleoprotein

SCAN - SRE-ZBP, CTfin51, AW-1 and Number 18 cDNA

SETDB1 – SET domain bifurcated 1

SINE – short interspersed nuclear element

TAD – topologically associated domain

TALE – transcription activator-like effector

TE – transposable element

TF – transcription factor

ZFN – zinc finger nuclease

ZnF – zinc finger

Table of contents

1. Introduction

Many fundamental biological processes within the cell, from transcriptional regulation to spacial organization of the genome, require a way for proteins to bind specific DNA sequences with high accuracy and affinity. Many different peptide domains have evolved to fulfil this role. This thesis will focus on zinc fingers, short peptide motifs that stand out among other DNA-binding motifs with their ability to join together with other zinc fingers to form longer arrays, targeting sites of varying lengths and sequences. This grants the zinc finger DNA-binding domains unique and interesting properties and allows them to be used in a broad range of proteins, some of which will be reviewed in this paper.

The zinc finger (abbreviated as ZnF) is a small structurally rigid protein motif stabilised by one or more zinc ions coordinated within the structure by cysteine and histidine residues. There are multiple different classes of zinc fingers, which will be briefly discussed below. However, in this paper attention will be paid primarily to classical zinc fingers, called also C2H2 zinc fingers. These short motifs, around 30 amino acids, encompass just one zinc ion bound generally by two cysteine and two histidine residues. Most of the motif sequence is highly conserved (J. Miller et al., 1985) and the most important changes occur in 4 positions – these 4 amino acid residues determine the zinc finger's binding specificity. Every individual C2H2 zinc finger can independently and specifically recognise a 3 bp segment of one strand of double-stranded DNA (dsDNA) plus one additional base in the opposite strand (Fairall et al., 1993).

The principle of DNA recognition by C2H2 ZnFs is distinctly different from most other DNA-binding domains. Whereas many proteins bind to DNA in the form of a dimer and make use of the symmetry of the DNA double-helix and of the usually palindromic sequence they recognise (helix-turn-helix motifs could be used here as an example), zinc finger domains can bind to any non-palindromic sequence in a linear way in the form of a monomer (Klug, 2010). This in part allows them to bear the most crucial feature, which is one of the reasons ZnFs are the most common protein domain in metazoan transcription factors (Weirauch & Hughes, 2011) – the ability to link together in tandem arrays, sometimes called "polydactyl" ZnFs because they contain many individual fingers. This grants the bond between the protein and its target DNA high affinity and specificity, as well as the capability to bind sequences of greatly varying lengths. Individual fingers in an array all have a similar structure, but exhibit affinity to a different base pair triplet, thus by arranging them in alternative ways, a multitude of combinatorial binding possibilities can be achieved.

These short motifs have been undergoing duplication and rearrangement by exon shuffling for hundreds of millions of years and have become associated with a wide selection of effector domains, generating multiple protein families with novel structures and functions. C2H2 zinc fingers have been found widespread in all eukaryotes including yeasts and plants (Table 1). The types of effector domains conjoined with polydactyl ZnF domains include transcriptional activators, repressors, nucleases and many others. Some of these families have remained exclusive to certain evolutionary clades and created lineage-specific groups of ZnF proteins (Stubbs et al., 2011).

The possibility to use zinc fingers as building blocks for the creation of designer domains specifically binding to a nucleic acid sequence of choice helped them receive a lot of attention in artificial protein engineering. Fusing these man-made domains to different effector domains gave rise to proteins which have been used in genome editing and gene therapy. Even though artificial ZnF proteins seem to have been surpassed in the last decade by other kinds of genome editing tools, there could still be more unique applications found for these custom polydactyl ZnF proteins in the future.

This thesis aims to provide a short introduction into how zinc fingers were initially discovered and what other types aside from C2H2 are there to be found and then review a selected few most notable naturally occurring ZnF array proteins and their biological roles. Attention will be also paid to artificially engineered ZnF proteins. In the end, I will discuss the future of ZnF array proteins, both short-term in their further use in genetic engineering and gene therapy and long-term from the evolutionary perspective.

Organism	Total protein-coding genes	Genes containing C2H2 motif
Homo sapiens	22 803	877 (3,85 %)
Mus musculus	21 517	749 (3,48 %)
Rattus norvegicus	23 098	697 (3,02 %)
Danio rerio	30 149	1 254 (4,16 %)
Drosophila melanogaster	13 986	303 (2,17 %)
Caenorhabditis elegans	19 985	193 (0,97 %)
Saccharomyces cerevisiae	6 600	52 (0,79 %)
Arabidopsis thaliana	27 628	165 (0,6 %)

Table 1 - Number of known genes containing the C2H2 motif compared with the number of all protein-coding genes in different model organisms. Data gathered using InterPro (Paysan-Lafosse et al., 2023) and Ensembl Biomart (Martin et al., 2023).

1.1. Zinc finger discovery

The first mention of a DNA-binding structural protein motif that encompasses a zinc atom in its structure comes from a 1985 paper by J. Miller, A.D. McLachlan and A. Klug (J. Miller et al., 1985). Previous studies of 5S RNA genes in *Xenopus laevis*, conducted by groups of D.D. Brown and R.G. Roeder, had shown that for RNA polymerase III to successfully bind and transcribe the gene, a transcription complex comprising at least 3 molecules is needed. Among these is transcription factor TFIIIA, a 38,5 kDa protein which directly interacts with a 50 bp region inside the coding sequence of the gene, called the "internal control region" (reviewed in Brown, 1984). 5S RNA molecules are being stored in immature *Xenopus* oocytes in the form of a 7S ribonucleoprotein (RNP) particle also containing TFIIIA, thus indicating that TFIIIA binds the 5S RNA gene as well as its product (Picard & Wegnez, 1979).

Upon experiments closer examining the 7S RNP, Miller et al. have found that each 7S particle contains between 7-11 Zn atoms and that proteolytic digestion of the 30 kDa domain of TFIIIA results in fragments of a limit size of ~3 kDa (J. Miller et al., 1985). McLachlan's computer analysis of the amino acid sequence of TFIIIA published by (Ginsberg et al., 1984) showed that the protein contains 9 tandemly repeated units of ~30 amino acids, each one with two pairs of cysteines and histidines at the same positions. Based on these results, Miller et al. suggested that these repeated units correspond in size to the 3 kDa fragments and each one of them is independently folded around a Zn cation, forming a small domain termed "zinc finger" for its shape and ability to "grasp" DNA and its need of the metal ion to preserve its structure (J. Miller et al., 1985).

The 30 amino acid domain consists of a 25 amino acid module forming the "finger" itself and 5 residues connecting the neighbouring zinc fingers in the form of a short linker. A zinc ion is coordinated in the middle of each finger by two invariant cysteines and two histidines, stabilising the structural motif. Zinc is not subject to redox chemistry, which makes it able to bind the two parts of the motif together even in the reducing conditions inside the cell, unlike disulphide bonds which would get reduced (Williams, 1984). This proposal of a zinc finger structure was supported in the following year by two other studies – an extended X-ray absorption fine structure (EXAFS) study confirmed that the Zn ion within each finger is indeed coordinated by two pairs of cysteines and histidines (Diakun et al., 1986) and the exon-intron composition of TFIIIA gene was shown to correspond with borders of the putative zinc finger domains (Tso et al., 1986). Aside from the conserved Cys and His pairs, three conserved hydro-

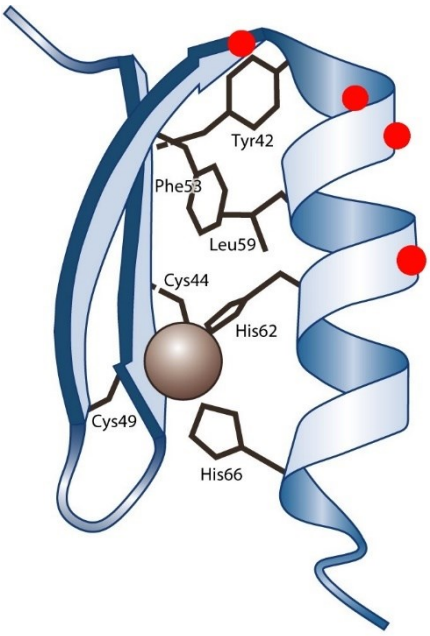

Figure 1 – Structure of a zinc finger domain obtained by a two-dimensional NMR study (Neuhaus et al., 1992). The zinc ion, represented by a grey sphere, is coordinated by amino acid residues Cys44 and Cys49 from the N-terminal β-sheet and His62 and His66 from the C-terminal α-helix. Tyr42, Phe53 and Leu59 form the hydrophobic core of the structure. The DNA-interacting amino acids are indicated by red circles. Adapted from (Klug, 2010).

phobic residues were noted in the middle region of the peptide between the two Zn coordinating pairs, supposedly forming a hydrophobic core of the domain (J. Miller et al., 1985). A structural model was proposed and later confirmed consisting of an antiparallel β-sheet, containing the two cysteines, and an α-helix with the two histidines (Figure 1) (Berg, 1988; Lee et al., 1989; Neuhaus et al., 1992).

The mode of specific binding of a zinc finger to DNA was first uncovered with the solution of the crystal structure of the three-fingered DNA-binding domain of mouse transcription factor Zif268 bound to a DNA oligonucleotide (Pavletich & Pabo, 1991). Each finger's α-helix binds into the major groove of DNA to three consecutive bases of one DNA strand with amino acid residues at helical positions -1, 3 and 6 using specific hydrogen-bond interactions, one amino acid per one base, while the amino acid at helical position 2 contacts one base in the opposite strand (Fairall et al., 1993). Multiple variations of this configuration were later found among different zinc finger-DNA complexes (Elrod-Erickson et al., 1998).

1.2. Zinc finger classification

The zinc fingers of the protein TFIIIA fall into the category of so-called "classical" ZnFs, also known as Cys_2His_2 or just C2H2 because of the way they bind their Zn ion. There are however

many diverse classes of zinc fingers with various ways of binding the Zn ion and differing structures (Figure 2) and functions. Here I will review a few of the best-known and most-represented classes, showcasing their roles in the organism. The presented categorisation however is not strictly accepted and other attempts to classify zinc fingers in different ways have been made (Iuchi, 2001; Matthews & Sunde, 2002).

The first-discovered C2H2 zinc fingers are the best-characterized type of these zinc-binding domains. They are the most numerous class of ZnFs in the human genome, being present within the structure of more than 3 % of all human genes (Klug, 2010). In DNA-binding protein domains, they are usually found in groups of at least three fingers to provide enough strength and specificity to the bond (Stubbs et al., 2011). Aside from binding to DNA, they are known to interact with RNA and proteins as well. In protein-protein binding, a single ZnF can be sufficient, as is the case with e.g. FOG-1, which binds as a cofactor to transcriptional regulator GATA-1 using just one of its multiple CCHC zinc fingers, a variant of the C2H2 which holds the same classical ββα conformation (Fox et al., 1999). Some C2H2 proteins of the class termed multiple-adjacent-C2H2 (Iuchi, 2001) have multiple binding activities at once, like WT1 binding both DNA and RNA (Caricasole et al., 1996) or Ikaros binding DNA and proteins (Sun et al., 1996).

The treble clef finger class has been named for the structure's resemblance to the musical score sign. They are commonly composed of a zinc knuckle (β-hairpin containing the sequence Cys-x-x-Cys) on the N-terminus followed by a loop, another β-hairpin, and an α-helix. The Zn ligand-pairs are located in the zinc knuckle and the N-terminal part of the helix. Treble clef fingers are present in various groups of proteins with a wide array of functions and an often undetectable sequence similarity despite structural likeness. Among these groups are RING finger-like domains, termed C3HC4 fingers and typically containing two Zn binding sites; protein kinase cysteine-rich domains, contained in protein kinases C or RAF kinases; or the nuclear receptor-like fingers, found in GATA-1, nuclear receptor DNA-binding domains or the LIM domain (Grishin, 2001; Krishna et al., 2003).

The Gag knuckle fold comprises a short zinc knuckle followed by a loop or a truncated helix. Resemblance to the classical ZnF can be seen, but both the β-hairpin and the helix are significantly shorter. One zinc ion is coordinated within the fold by two ligands from the zinc knuckle and another two from the loop or both sides of the helix. This fold group was named the Gag knuckle because it is found in retroviral nucleocapsid gag proteins, which bind single-stranded RNA through this domain and play a major role in viral packaging (Laity et al., 2001).

A Gag knuckle is also included in the A subunit of RNA polymerase II or the reoviral outer capsid protein σ3 (Krishna et al., 2003).

In the zinc ribbon domain, zinc-binding ligands are provided by two zinc knuckles. The N-terminal primary β-hairpin together with the C-terminal secondary hairpin form the structural core of the fold (Krishna et al., 2003). Zinc ribbons are very diverse and are often only similar in the sequence of the zinc knuckles, otherwise showing great structural variability. The family of classical zinc ribbons includes domains primarily from proteins involved in transcription or translation, including transcription factors, RNA polymerases, topoisomerases, and ribosomal proteins. Among other families of the zinc ribbon fold are the rubredoxin family, found in proteins used in electron transfer, or the family of zinc ribbons of the adenovirus DNA-binding protein (summarized in Krishna et al., 2003).

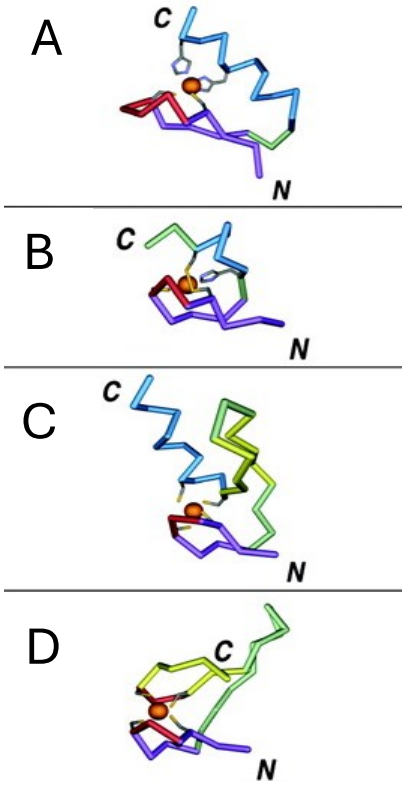

Figure 2 – Structures of chosen zinc finger fold groups. (A) C2H2-like finger (B) Gag knuckle (C) Treble clef (D) Zinc ribbon. Colour coding used: orange – zinc-binding residues, red – zinc knuckles, cyan – α-helices, purple – β-strands neighbouring the zinc knuckle, yellow – other β-strands, green – loops. Adapted from (Krishna et al., 2003).

2.Naturally occurring zinc finger array proteins

C2H2 zinc finger DNA-binding domains appear within a large number of metazoan proteins. This is due to their ability to tandemly link together in altering lengths and orders, which confers huge variability in the specificity of binding. Throughout evolution, zinc finger arrays have been brought together with different types of effector domains through exon shuffling (Stubbs et al., 2011). Many of these composite proteins function in gene regulation as site-specific transcription factors (TFs). Aside from those, ZnF array proteins with other

functions are also known, such as regulation of DNA-methylation by VEZF1 (Dickson et al., 2010), transcriptional repression by KRAB domain-containing proteins (Bellefroid et al., 1991), regulation of chromatin structure by CTCF (reviewed in Phillips & Corces, 2009) or meiotic recombination hotspots determination by PRDM9 (Baudat et al., 2010; Myers et al., 2010). Typically, a ZnF array is located on the C-terminus of a protein and provides specific DNA-binding capability while an effector domain on the N-terminus either interacts with and "recruits" other often more general regulatory proteins or provides some regulatory function by itself (Frietze & Farnham, 2011).

2.1. Zinc finger arrays

The length of a ZnF array is usually at least 3 fingers, but proteins with over 40 tandemly arranged fingers were documented in vertebrate genomes (Stubbs et al., 2011). In proteins containing only one or two fingers, additional structural elements are often present to facilitate DNA recognition (Bowers et al., 1999; Dutnall et al., 1996; Fairall et al., 1993). The average number of zinc fingers in arrays with more than 3 fingers is 8,5 in humans and 7,5 in mice. An average human ZnF array protein could thus recognise a DNA sequence of 25-26 bp which is way more than typical transcription factors do. The longest ZnF arrays could bind sequences over 100 bp long, which is unusual, considering a 3 Gbp genome on average contains a specific 16 bp sequence just once. (Emerson & Thomas, 2009). This makes recognising sequences of such length seem redundant for ensuring the specificity of binding. One possible explanation is that a long array could be functionally divided into multiple subsets of fingers, each binding a different target (Hata et al., 2000; Hoffmann et al., 2003). Apart from ZnFs binding to DNA, cases are known of them mediating protein-protein relations and interacting with other transcription factors, for example with the Sp1 protein (Gregory et al., 1996; J. S. Lee et al., 1993; Perkins et al., 1994).

2.2. Effector domains

Most of the effector domains contained in transcription factor proteins can generally be divided into three groups. Firstly, these domains can directly interact with parts of the basal transcriptional machinery and mediate their interaction with DNA or with one another. So-called transactivation domains make contact with general TFs - members of the preinitiation complex - to recruit and stabilise them. The second type of TF effector domains cooperatively interact with other site-specific factors. This can enable the binding of factors whose affinity to a given DNA sequence otherwise would not be high enough. Aside from stabilising weaker

protein-DNA interaction, this can also allow combinatorial and thus more complex regulation of transcription. Lastly, effector domains can often recruit chromatin-modifying enzymes. Among these are histone acetyltransferases (HATs) and histone methyltransferases (HMTs), whose activity largely determines the open or closed state of chromatin and ultimately the expression of an affected gene (reviewed in detail in Frietze & Farnham, 2011).

2.3. KRAB

There are many ZnF protein families with different effector domains. A few of the most notable include BTB/POZ, ZAD, SCAN and the largest one – the KRAB domain family. This domain was discovered in 1991 by Bellefroid et al. and named Krüppel-associated box (KRAB) for its relatedness to the Krüppel zinc finger protein, which is essential for the regulation of segmentation in *Drosophila* (Bellefroid et al., 1991). The KRAB-zinc finger protein (KZFP) family has expanded remarkably and comprises hundreds of known proteins which are found almost exclusively in tetrapod vertebrates (Huntley et al., 2006) – only one ortholog of a KRAB domain has been reported in sea urchin (Birtle & Ponting, 2006). KZFPs have been duplicating most rapidly in mammalian lineages where they are the dominant TF type. The human genome for example contains over 400 KRAB-zinc finger protein-coding loci, generating over 700 alternatively transcribed KZFP mRNAs (Huntley et al., 2006), which could constitute up to 25 % of estimated total human TF genes (Vaquerizas et al., 2009).

Zinc finger array genes, including KZFP genes, are often found on chromosomes together in large clusters due to being prone to frequent segmental duplications, creating copy-number variants (Cooper et al., 2007; Nowick et al., 2010). Two potential causes have been identified that could explain the rapid pace with which new KZFPs are emerging and diverging in function. First, the duplicated genes often possess differing expression patterns compared to their parental gene. This implies that genes are duplicated together with their regulatory elements and that neighbouring genes are protected from their influence, leading to different levels of expression or tissue-specific expression even between closely related KZFP genes with high sequence similarity (Nowick et al., 2010; Stubbs et al., 2011). Second, the target specificity of KZFPs readily changes either by non-synonymous mutations in individual amino acids in the binding regions of the zinc fingers or by duplications and deletions of whole fingers (Shannon et al., 2003). The latter happens by the same mechanisms as duplications and deletions of microsatellites or any other simple tandemly repeated sequence – by stochastic illegitimate recombination between different segments of the array or by a mechanism such as

slippage of the DNA polymerase during replication (Hardwick et al., 2009). Research has shown that the formation of genes with novel DNA-binding specificities is evolutionarily favourable and positive selection is acting in the process of rapid diversification of KZFPs and other ZnF array proteins (Emerson & Thomas, 2009).

The KRAB is an approximately 75 amino acids long domain, composed of two parts called the KRAB A-box and B-box (Bellefroid et al., 1991). The A-box alone can interact with the domain's protein partners and mediate transcriptional regulation and the B-box is thought to potentiate the effectivity of these interactions (Vissing et al., 1995; Witzgall et al., 1994). KRAB recruits a ubiquitous scaffold KRAB-associated protein-1 (KAP1) which enables KZFPs' repressive activity (Friedman et al., 1996) by binding multiple different heterochromatin-inducing proteins. This silencing complex consists of heterochromatin protein 1 (HP1) isoforms (Nielsen et al., 1999; Sripathy et al., 2006), the histone3-lysine 9-specific methyltransferase SET domain bifurcated 1 (SETDB1) (Schultz et al., 2002), the nucleosome remodelling and deacetylase (NuRD) complex (Schultz et al., 2001) and DNA methyltransferases (Quenneville et al., 2012) (Figure 3). The KAP1 co-factor is abundant in all tissues and acts as a universal binding partner for KRAB, mediating its repressive function. This is partly also what allowed the KRAB family of proteins to grow to such numbers – new KRAB proteins with novel DNA-binding capabilities can arise without hindering the

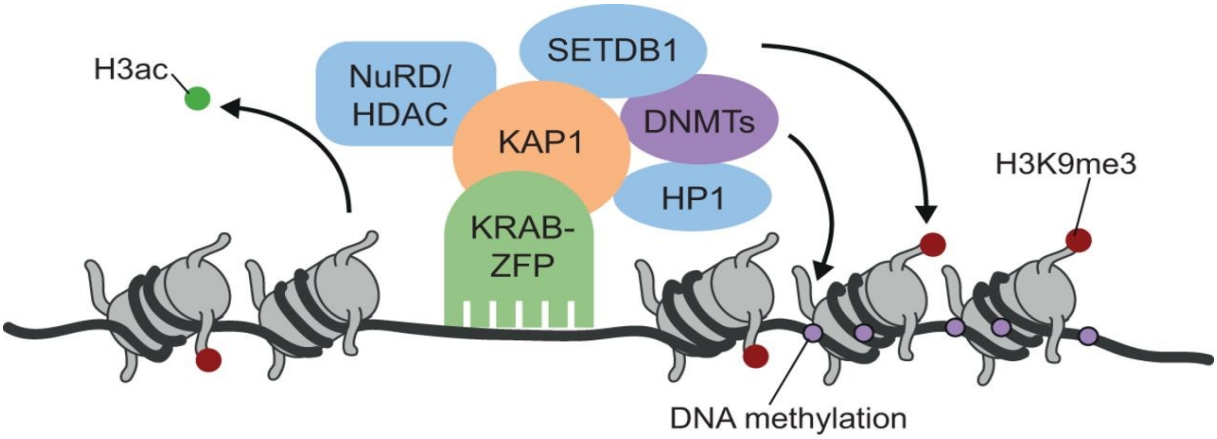

Figure 3 – The KZFP silencing complex. The KZFP binds to DNA with its array of zinc fingers and recruits KAP1 via its KRAB domain. KAP1 subsequently recruits a repressor complex, resulting in histone methylation (H3K9me3), histone deacetylation (H3ac) and DNA methylation and ultimately in transcriptional repression. KAP1 – KRAB-associated protein-1; HP1 – heterochromatin protein 1; SETDB1 – SET domain bifurcated 1; NuRD/HDAC – nucleosome remodelling and deacetylase complex/histone deacetylase; DNMT – DNA methyltransferase; H3K9me3 – trimethylated lysine 9 of histone 3; H3ac – acetylated histone 3; KRAB-ZFP – Krüppel-associated box- zinc finger protein. From (Ecco et al., 2017).

interactions of KRAB/KAP1 on the N-terminus. KZFPs generally do not dimerise with other proteins, nor do they need other interaction partners for the stabilisation of their bond to DNA and therefore need not be as conserved as other protein families that require tighter interaction with their partners (Stubbs et al., 2011). Several highly conserved lineages of KZFPs contain other domains at the N-terminus aside from KRAB, such as DUF3669 (domain of unknown function 3669) or SCAN (SRE-ZBP, CTfin51, AW-1 and Number 18 cDNA). The SCAN domain can mediate dimerisation with other proteins while the function of DUF3669 is, as the name suggests, still largely unknown (Ecco et al., 2017; Edelstein & Collins, 2005). Some KZFPs, often those containing these domains, do not recruit KAP1 and perform some other function than transcriptional repression. A few of these proteins are thought to be involved in transcriptional activation, but more data needs to be gathered on this subject (reviewed in Rosspopoff & Trono, 2023).

2.3.1. Control of transposable elements

KZFPs bind to diverse genomic targets, such as promoters, simple repeats or genes for other ZnF array genes, where they execute different tasks (Imbeault et al., 2017). Covering all of these is, however, beyond the scope of this thesis. Hence, I will only review the most numerous and (to me) interesting KZFP targets – transposable elements (TEs).

Transposable elements are mobile genetic elements - sequences of nucleic acid capable of replicating and changing their position within the genome. They were first discovered in maize in 1950 by Barbara McClintock (McClintock, 1950). By the mode of their transposition, TEs are classified either as DNA transposons or retrotransposons (Finnegan, 1989). DNA transposons move using a DNA intermediate and usually (but not always) with the cut-and-paste mechanism, essentially being excised from their original location by transposase enzymes and integrated elsewhere in the genome (Wicker et al., 2007). Retrotransposons make use of an RNA transcript which is then reverse transcribed into DNA and inserted into the genome, the so-called copy-and-paste mechanism (Weiner et al., 1986). Estimates of the abundance of TEs within the human genome reach 4,5 million TE-derived sequences, making up about 50 % of its DNA content (Ecco et al., 2017) or even more, because over time, TEs become unrecognisable due to mutations (de Koning et al., 2011). About 1,6 % of the human genome is comprised of DNA transposon fossils (Smit, 1996) while the rest of the contained TEs belong to the class of retrotransposons.

TEs can impact their genomic environment in many ways including, but not limited to, providing alternative promoters, alternative splice sites, novel binding sites for transcription factors or production of regulatory RNAs (Friedli & Trono, 2015). Even though no more than one in every ten thousand human mobile elements are still able to transpose (Hancks & Kazazian, 2016), they provide a huge and important regulatory platform and are often crucial for the evolution and biology of organisms (Chuong et al., 2017; Garcia-Perez et al., 2016; Trizzino et al., 2017). TEs have been documented for example as the source of modified tissue-specific gene expression (Ting et al., 1992) and providers of binding sites for pluripotency factors in embryonic stem cells (ESCs) (Bourque et al., 2008) or for the tumour suppressor protein p53 (T. Wang et al., 2007). At the same time, there are many known human pathologies caused by TEs, such as the insertion of a primate specific retrotransposon of the Alu SINE (short interspersed nuclear element) type into the genes *BRCA1* and *BRCA2* causing breast cancer (Miki et al., 1996; Puget et al., 1999) or the activation of an endogenous retrovirus (ERV) of the long terminal repeat (LTR) retrotransposon class causing oncogene expression in human Hodgkin's lymphoma (Lamprecht et al., 2010).

The majority of KZFPs, about two-thirds, bind to sites located within TEs, mostly in retrotransposons. Some KZFPs can associate with multiple TE families and some TE subfamilies are, on the contrary, targeted by multiple KZFPs (Imbeault et al., 2017). A well-documented function of KZFPs is the control of early development and embryogenesis by the formation of heterochromatin in specific loci. They allow for the silencing of TE-derived sequences to stop them from impeding the precisely orchestrated processes that take place during embryogenesis from zygotic genome activation onwards and to ensure transcriptional homeostasis. Upon the depletion of the scaffold protein KAP1 or the H3K9 methyltransferase SETDB1 in human or mice ESCs, several TEs become expressed (Corsinotti et al., 2013; Matsui et al., 2010; Rowe et al., 2010; Turelli et al., 2014).

Research has shown that there has been an ongoing coevolution of KZFPs and TEs for millions of years. When the genome has been invaded by a new family of ERVs, novel KZFP duplicates targeting these TEs have appeared subsequently, mirroring the waves of retroviral invasion into the genome (Najafabadi et al., 2015; Thomas & Schneider, 2011). These data have led to the proposition of the hypothesis of a perpetual "arms race" between TEs and the KZFPs that suppress them. This model suggests that KZFPs with novel binding specificities emerge in response to rapidly mutating TEs, driven by strong positive selection, to silence them and keep them from propagating further (Emerson & Thomas, 2009). Support for this model can be found

for example in the dynamic regulation in human ESCs of LINE-1 (also called L1, long interspersed nuclear element-1) lineage of retrotransposons, the only known still active autonomous retrotransposon. Any time a new L1 lineage appears, it is first suppressed by small RNA pathways inducing DNA methylation, namely the PIWI-interacting RNA pathway, before KZFPs capable of binding these TEs appear and take on the role of silencing them until mutational drift deprives them of any activity (Castro-Diaz et al., 2014).

More recent studies however suggest this model is too simplistic and the host-invader arms race is not the only driving force of the rapid KZFP selection and diversification. A so-called "domestication" model has been presented, proposing that KZFPs do not just silence TEs and obstruct their transposition potential, but also co-opt them for the benefit of the host (Imbeault et al., 2017). This is supported by several lines of evidence. First, tens of thousands of TEs, despite being long devoid of any transposition activity, are still regulated by KRAB/KAP1 complexes, thus the retainment of KZFP-binding sequences in these integrants cannot stem from the necessity to block their replication (Ecco et al., 2017). Second, many TEs have invaded the human genome or kept duplicating even after a KZFP capable of binding their sequence had appeared and recent data has even shown positive selection in some TEs for providing binding sites for KZFPs (Imbeault et al., 2017). Third, the expression patterns of KZFPs are often highly regulated and tissue- and lineage-specific, suggesting tight control over the interactions of KZFPs with TEs (Ecco et al., 2017).

It was previously accepted that most TEs were irreversibly repressed during early embryogenesis, thus exhibiting no need for further suppression (Maksakova et al., 2008). However, more and more research shows this is not true. KZFPs have been shown to regulate TEs in adult tissues, establishing transcriptional networks that affect not only early development but many other physiological processes. TEs contribute to proper human embryonic genome activation while being controlled by KZFPs, but later the same TEs also act as tissue-specific enhancers (Pontis et al., 2019). The murine KZFPs ZFP932 and Gm15446 were demonstrated to bind ERV sequences in ESCs, but also in a range of somatic cells where they subsequently modulate the expression of neighbouring genes via promoter and enhancer effects (Ecco et al., 2016). One part of the human body seems to be especially prone to dynamic evolution due to TE and KZFP activity – the central nervous system. A wide range of psychological and cognitive phenotypes in humans don't remarkably lower life expectancy and biological fitness (Ecco et al., 2017), thereby allowing for more diversity and rapid evolution. In line with this, a higher number of KZFPs have been found expressed in brain tissues than in the majority of

other adult human organs (Imbeault et al., 2017) and a higher degree of TE activity has been noted in the brain (Erwin et al., 2014). KZFPs ZNF417 and ZNF587 were shown to control regulatory sequences based on TEs in human neurons and influence their differentiation and neurotransmission profile. Interindividual polymorphism in these KZFPs and/or their TE targets has been consequently proposed to introduce variability in brain function and development (Turelli et al., 2020). Other studies have also confirmed the widespread correlation of KZFP expression in the human brain with the expression of their target genes, suggesting the involvement of KZFPs in gene regulatory networks during brain development (Farmiloe et al., 2020). These findings imply that KZFPs together with their TE targets could have had a crucial role in the development of higher brain functions in humans (Ecco et al., 2017), offering an exciting field for future research.

2.4. PRDM9

Another remarkable protein that demonstrates the unique properties of zinc finger arrays is PRDM9 (PR/SET domain-containing protein 9), the meiotic chromosome recombination-facilitating protein present in many vertebrates, including most examined mammals except canids (Paigen & Petkov, 2018). In 2010, three research groups proposed PRDM9 as a major component of the meiotic recombination hotspot-forming machinery (Baudat et al., 2010; Myers et al., 2010; Parvanov et al., 2010). Chromosomal recombination, also known as crossing-over, is a key process occurring in sexually reproducing organisms during gametogenesis that ensures proper chromosome segregation in the first meiotic reductional division and creation of genetic diversity by reciprocal exchange of genetic material between two homologous chromosomes (reviewed in Petronczki et al., 2003). The meiotic recombination involves the highly regulated induction of double-strand breaks (DSBs) in the DNA, which are then repaired using the homologous chromosome. These DSBs don't occur randomly along the chromosomes but cluster into short (1-2 kb in humans) segments of the genome called recombination hotspots (Jeffreys et al., 2001). The sequence-specific binding of PRDM9 determines the location of these hotspots (Baker et al., 2014; Brick et al., 2012) and facilitates the success of subsequent events during the reparation of the DSBs (Brick et al., 2012; Hayashi et al., 2005). PRDM9 was also identified as the first and so far the only mammalian gene involved in speciation and hybrid sterility (Mihola et al., 2009).

PRDM9 has a KRAB domain on the N-terminus, which however in this case does not recruit KAP1 but is rather involved in interactions with other proteins required for the proper

advancement of meiosis (Imai et al., 2017; Parvanov et al., 2017). Aside from KRAB, PRDM9 comprises an SSXRD nuclear localization signal, a SET methyltransferase domain (Hayashi et al., 2005; Powers et al., 2016) and a C-terminal C2H2 ZnF array. Most of the protein's structure is highly conserved across different species (Paigen & Petkov, 2018) except for the ZnF array, which is highly polymorphic in the binding-specifying amino acids of the fingers and even in the number of ZnF repeats (I. L. Berg et al., 2010). PRDM9 binds with its ZnF array to a recognition sequence motif where it initiates recombination. However, a phenomenon called the hotspot conversion paradox occurs in meiosis, where the recombination-initiating hotspot allele is replaced with a copy of its homologue by gene conversion (Boulton et al., 1997) and thereby slowly introducing mutations into the PRDM9 recognition sequence and PRDM9 consequently losing its binding affinity to it. This would eventually lead to the loss of all PRDM9 binding sequences and PRDM9-facilitated crossing-over could no longer occur, resulting either in the development of a different recombination mechanism or in sterility and extinction of the species. For this reason, there is a strong positive selection acting on the DNA-binding domain of PRDM9 so that new alleles with different hotspot binding sequences but otherwise the same conserved protein structure are appearing rapidly and thus overcoming the problem of PRDM9 binding target erosion (Baker et al., 2015). PRDM9 allele variations have been studied extensively in mice (Buard et al., 2014) and humans (I. L. Berg et al., 2010; Ponting, 2011) and very high numbers of them have been discovered, prompting claims that *Prdm9* may even be the most rapidly evolving protein-coding mammalian sequence known (Powers et al., 2016).

The complete mechanism by which PRDM9 affects the course of meiosis has not been elucidated yet. It was initially thought to act as a TF, this is however refuted by the rapid evolution of the DNA-binding ZnFs (Oliver et al., 2009). It is known that aside from the KRAB domain enabling interactions with other important protein players, a major role in PRDM9's function is played by the PR/SET methyltransferase domain. This enzymatic domain was shown to epigenetically mark histones within the recombination hotspots – it is capable of trimethylating both H3K4 and H3K36 *in vivo* (Powers et al., 2016). The methylated nucleosomes then move apart, creating a nucleosome-lacking region centred over the PRDM9 binding motif (Baker et al., 2014) where a DSB is subsequently formed after the recruitment of topoisomerase SPO11 (Lange et al., 2016). This nucleosome-depleted region of DNA also limits the extent of Holliday junction migration during the crossing-over (Baker et al., 2014). Before the DSB is made, thus marked part of the DNA loop needs to be brought to the

chromosomal axis – this likely happens through the protein complex that associates with PRDM9's KRAB domain (Parvanov et al., 2017).

Before PRDM9 was recognized as a major part of the recombination machinery, Mihola et al. have labelled it as a gene responsible for hybrid sterility in crosses between two subspecies of *Mus musculus* (Mihola et al., 2009). As the only candidate gene for the hybrid sterility locus *Hst1*, described in mice in 1974 by J. Forejt and P. Iványi (Forejt & Iványi, 1974), *Prdm9* was found to be involved in meiotic failure in sterile male mice hybrids between the subspecies *M. m. musculus* and *M. m. domesticus*. Mihola et al. suggested that as a gene reducing hybrid fitness, *Prdm9* could be the first known speciation gene in mammals, creating reproductive isolation between subspecies and facilitating divergent evolution. The reason for this is likely the incompatibility of different PRDM9 alleles caused by the rapid evolution of ZnF arrays in both subspecies (Oliver et al., 2009).

PRDM9 greatly illustrates the versatility of ZnF DNA-binding arrays, as the rapidly mutating ZnF domain tethered to otherwise highly conserved effector domains allows the preservation of functional meiosis and therefore reproduction in mammals and could make PRDM9 one of the genetic drivers of speciation.

2.5. CTCF

The last endogenous ZnF protein I chose to review is the CCCTC-binding factor (CTCF). It was first described in 1990 as a novel DNA-binding protein found in the 5' flanking region of the chicken *c-myc* gene and implicated a role of a transcriptional repressor (Lobanenkov et al., 1990). In later studies, CTCF was suggested to partake in many diverse processes, such as transcriptional regulation (Filippova et al., 1996; Vostrov & Quitschke, 1997), barrier insulation and enhancer blocking (Bell et al., 1999), genomic imprinting (Bell & Felsenfeld, 2000; Szabó et al., 2000) and long-range chromatin interactions and maintenance of genome topology (Splinter et al., 2006; Zhao et al., 2006), with some of these attributed roles even contradicting each other. With the advancement of technology and research techniques, more recent studies have proposed the various functions of CTCF are in fact the results of its ability to mediate inter- and intrachromosomal interactions between distant parts of the genome and are fully context-dependent – meaning that the outcome depends on the genomic environment of the CTCF-binding site and the other specific proteins CTCF interacts with (discussed in Ong & Corces, 2014).

The CTCF protein contains a highly conserved set of 11 DNA- and protein-binding ZnFs which are almost 100% homologous between humans, chickens and mice (Filippova et al., 1996) and shorter motifs on the N- and C-termini, such as an AT-hook with a proposed role in DNA-binding or protein-protein interactions or a SKKEDSSDSE motif where CTCF can be phosphorylated (summarized in Ohlsson et al., 2001). The sequence identity between avian and mammalian ZnFs is 100 %, yet the sequences they recognize are not identical. This is due to the ability of CTCF to use different subsets of fingers for binding of each target (Filippova et al., 1996; Nakahashi et al., 2013), thus greatly increasing the number of possible target sequences and protein partners even with a highly conserved set of ZnFs. Other factors impacting the binding of CTCF to specific sites and its function are differential CpG methylation of target sequences (H. Wang et al., 2012), post-translational covalent modification of CTCF (MacPherson et al., 2009; Yu et al., 2004) or interaction with other proteins, the most prominent being cohesin (Wendt et al., 2008; Xiao et al., 2011). A subset of CTCF-binding sites was indicated to be cell-type specific (Cuddapah et al., 2009), further supporting the hypothesis that CTCF function is largely context-dependent.

Genome-wide studies have shown a strong correlation between positions of enhancer elements and cell-type specific CTCF-binding sites (Sanyal et al., 2012; Shen et al., 2012), showing that one of the main functions of CTCF may be to facilitate interactions between promoters and distant regulatory sequences. This is illustrated by the role of CTCF in the activation of expression of the major histocompatibility complex II (MHC II) locus (Majumder et al., 2008) or in the regulation of lineage-specific V(D)J recombination at antigen receptor loci in B and T cells (C. Guo et al., 2011). CTCF was also demonstrated to control the pausing of the transcription elongation process of RNA polymerase II and to affect alternative mRNA splicing (Shukla et al., 2011), showcasing another role of differential cell-type specific binding of CTCF in the regulation of gene expression.

As I already mentioned earlier, a major role attributed to CTCF is that of genome topology maintenance. Higher eukaryotic genomes are arranged into topologically associating domains (TADs) which are characterized by high intradomain and low interdomain frequency of interactions (summarized in Fudenberg et al., 2016). TAD borders in mammals were found to be enriched for CTCF-binding sites, however, only 15 % of these sites are found at TAD borders while the rest are found inside TADs (Dixon et al., 2012). Experiments have shown that CTCF together with cohesin play a part in TAD border organization, which is loosened upon their depletion (Sofueva et al., 2013; Zuin et al., 2014), consistent with the reports that CTCF acts as

an insulator. These two proteins alone, however, probably are not sufficient for TAD border maintenance and other protein partners are needed, two likely candidates being condensin and TFIIIC (Ong & Corces, 2014). The majority of CTCF that is bound within TADs likely participates in mediating short-range interactions (Lin et al., 2012) together with cohesin through a mechanism of controlled chromatin loop extrusion (Fudenberg et al., 2016; Sanborn et al., 2015), thus rationalising the claims of CTCF being an enhancer/silencer facilitator.

All of these results suggest that CTCF acts as an architectural protein affecting genome topology (Ong & Corces, 2014) and performs diverse functions based on the combination of its interaction partners, covalent modifications and varying binding specificity, although its whole set of 11 ZnFs remains unchanged. CTCF is thus an important example of a ZnF protein driven by selection pressure to retain a widely conserved ZnF domain. This contrasts with the earlier described property of ZnF arrays allowing them to easily mutate and change binding specificity.

3. Artificial zinc finger array proteins

The unique properties of zinc finger arrays that I described earlier, being their modular design and many combinatorial binding possibilities that stem from it, were recognized as an opportunity for the creation of designer ZnF proteins that could with proper assembly specifically bind any DNA sequence needed. These custom DNA-binding domains could then be connected to various effector domains and provide a tool for specifically targeted genome engineering or gene regulation, providing potential for use in biomedicine and other fields. Here, I will summarize the process and limitations of ZnF engineering and provide examples of achievements in genome manipulation and biomedical research.

To be able to rationally select ZnFs for the creation of proteins binding to a desired sequence, it was first needed to generate a library of ZnFs binding to ideally all of the 64 possible base pair triplets and to understand the code behind which combinations of amino acids in the recognition positions within the fingers bind which targets. A method that proved to be useful for this purpose is phage display, where a zinc finger is cloned and fused to a protein of bacteriophage fd's capsid, expressed together with the capsid and subsequently purified by the affinity of binding to a target oligonucleotide (Choo & Klug, 1994b). Building on these data, these researchers were able to uncover some broad rules of a DNA-recognition code and allow for the prediction of a binding site for a specific finger (Choo & Klug, 1994a). As a proof of concept, they engineered a protein comprising 3 ZnFs that was able to specifically bind a 9 bp sequence of a *bcr-abl* oncogene and block its transcription (Choo et al., 1994). Subsequently,

an almost complete library of predefined ZnF modules was developed by the team of C.F. Barbas III (Dreier et al., 2001, 2005; Segal et al., 1999) and a different one by the biotech company ToolGen under the leadership of Kim Jin-soo (Bae et al., 2003). Using these libraries, a method of ZnF protein engineering called the modular assembly became easier and less time-consuming, allowing the construction of sets of predefined ZnFs connected by conserved linkers without the need to lengthily select the needed ZnFs *de novo* every time (described in detail in Bhakta & Segal, 2010). A different approach called oligomerized pool engineering (OPEN) was devised by the team of J.K. Joung that also accounts for the possible interactions between neighbouring fingers during their selection, which could affect their binding specificity (Maeder et al., 2009).

3.1. Zinc finger nucleases

Upon the successful creation of a DNA-binding domain of choice, it can be fused to various effector domains. The most widely used type of artificial ZnF constructs are zinc finger nucleases (ZFNs). These proteins are used for the formation of site-specific DSBs which are then fixed by the cellular DNA damage response mechanisms either by non-homologous end joining (NHEJ) or homologous recombination (HR). The presence of a DSB without a homologous template available usually triggers NHEJ which is error-prone and often leads to short insertions or deletions in the site of the DSB resulting in frameshift mutations and loss of function of the targeted gene. ZFNs thus greatly increased the efficiency of gene knockout, a crucial genetic engineering technique (Santiago et al., 2008). Alternatively, if donor DNA with locus-specific homology arms on both ends is provided, HR response is triggered and the DSB is repaired with the use of the donor DNA as a template. This leads to the replacement of the original gene with a modified version or insertion of an entirely new gene (Moehle et al., 2007). When DNA is cleaved by two ZFNs simultaneously, a larger chromosomal segment can be deleted as well (H. J. Lee et al., 2010) (Figure 4).

The most commonly used ZFN effector domain responsible for the generation of the DSB in the target site is the cleavage domain of the type II restriction enzyme *Fok*I, which is in itself sequence non-specific and can be joined to a different DNA-binding domain, such as a ZnF array, as first reported by Kim et al. (Y. G. Kim et al., 1996). The *Fok*I domain has to dimerize in order to cleave DNA (Bitinaite et al., 1998). For this reason, the optimal setup was found to comprise two monomeric ZFNs, each with 3 ZnFs connected to *Fok*I by a short linker, binding

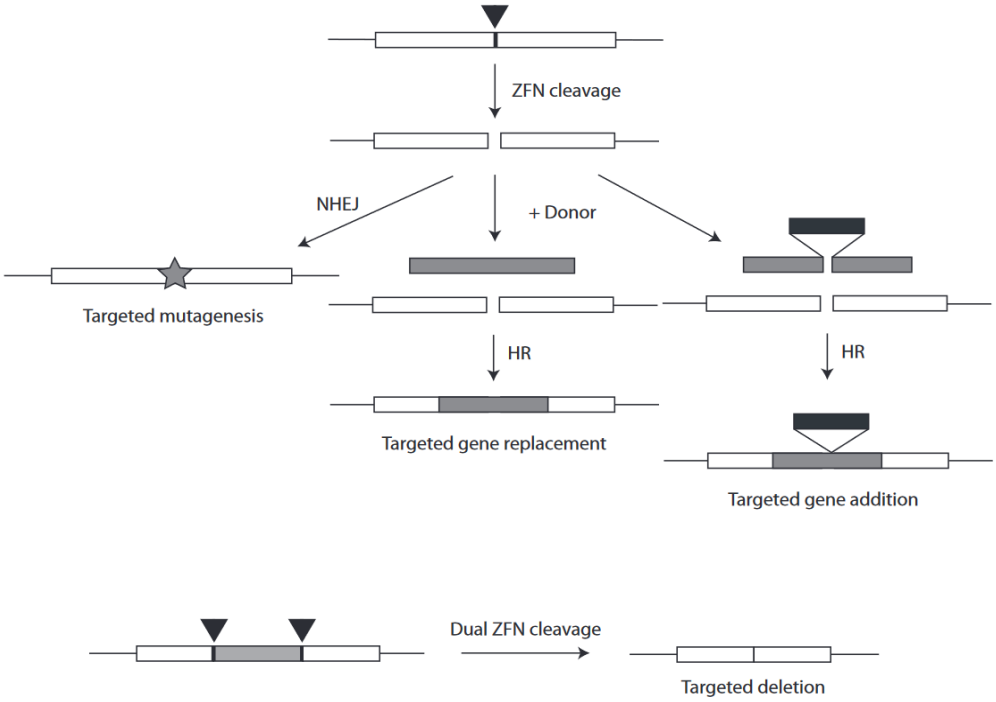

Figure 4 – HR and NHEJ repair pathways for a ZFN-induced DSB. With no homologous template available, NHEJ occurs and often introduces a mutation into the target gene. By HR with donor DNA present, the original gene can be replaced, or an entirely new gene sequence can be added. By dual ZFN cleavage, gene deletion can be achieved. From (Carroll, 2011b).

to 9 bp sites on opposite DNA strands with a 6 bp distance between them (Bibikova et al., 2001). The need for dimerization of the two proteins helps with limiting off-target cleaving in sites where only one of the two ZFNs would bind. The requirement for the binding of both 3 ZnF-containing monomers in the same locus demands specific binding to 18 bp of DNA, which in theory is enough to target only one locus even in large genomes (Carroll, 2011a). To refine the cleavage specificity even further and enhance the efficacy of DSB forming, modified *Fok*I domains were engineered that have to bind as a heterodimer to cleave their target (J. C. Miller et al., 2007; Szczepek et al., 2007) and exhibit higher cleavage activity (J. Guo et al., 2010). Zinc finger nickases have also been developed which only cut one strand of DNA, stimulating the HR but not the NHEJ repair pathway, leading to reduced frequency of mutagenic effects (Ramirez et al., 2012). Higher success rates were also reported when using four-fingered monomers instead of ones containing just three ZnFs (H. J. Kim et al., 2009).

ZFNs were found to have great potential in biomedicine and therapeutic use. The targeted gene disruption or replacement enables the correction of underlying genetic causes of many diseases, providing not just symptomatic but permanent cure for the condition. ZFNs have been applied to fix mutations related to sickle cell disease (Sebastiano et al., 2011), haemophilia B (Li et al., 2011), X-linked severe combined immune deficiency (Urnov et al., 2005) or Parkinson's

disease (Soldner et al., 2011). ZFN-mediated knockout of the CCR5 receptor in $CD4^+$ T cells was shown to potentially grant heritable immunity to human immunodeficiency virus (HIV) and even entered clinical trials (Perez et al., 2008; Tebas et al., 2014).

3.2. Other custom zinc finger proteins

Aside from the *Fok*I nuclease domain, artificial ZnF arrays have been fused to many other effector domains with diverse functions. To address the downsides of ZFNs, such as the generation of potentially toxic off-target DSBs and the reliance on the endogenous cellular DNA repair mechanisms dependent on cell type and cell cycle (Gaj et al., 2013), different ZnF proteins have been developed, such as ZnF recombinases and transposases. ZnF recombinases, as the name suggests, directly catalyze recombination between two specific binding sites (Gordley et al., 2007) whereas ZnF transposases are capable of site-specific integration of a transposon expression cassette (Voigt et al., 2012). ZnF proteins have also been engineered to regulate the expression of selected genes, forming a class of artificial transcription factors (ATFs). These proteins contain transcriptional activator or repressor effector domains, such as VP64 and KRAB (Beerli et al., 1998). ATFs were demonstrated to successfully downregulate an HIV-1 promoter (Segal et al., 2004) or stimulate the expression of γ-globin in hematopoietic stem cells as a treatment for β-thalassemia and sickle cell disease (Wilber et al., 2010). Systems for sequence-specific detection of methylated CpG dinucleotides utilizing ZnF DNA-binding domains have also been devised (Stains et al., 2006).

Artificial ZnF proteins unfortunately have drawbacks as well. Even with big advancements in the simplicity and accessibility of this method to all researchers, the protein engineering needed is still quite laborious, time-consuming and expensive (discussed in Chandrasegaran & Carroll, 2016). An important and limiting step of their application is also the delivery of the protein construct into the cell. This can be done either through DNA, mRNA or directly with purified protein, with each of these methods having significant upsides but also downsides (reviewed in Zhang et al., 2019). With the emergence of CRISPR/Cas9 systems in 2012, artificial ZnF proteins have slowly started to fade away from common use in genetic engineering.

4. Conclusion

Since their discovery in 1985, zinc fingers have been recognized as widespread and versatile protein motifs providing specific binding primarily to DNA. The main “selling point” of these peptide structures is their ability to form arrays of varying lengths and composed of individual fingers with different triplet binding specificities, granting them the potential to bind in monomeric form to virtually any DNA sequence. ZnFs are prone to frequent target sequence changes due to mutations in individual DNA-binding amino acid residues or due to duplications, deletions, and recombination of whole fingers. This, coupled with the capacity to associate with various effector domains, enabled throughout evolution the emergence of myriad functionally different protein families utilizing ZnF arrays. These proteins use ZnFs to specifically bind DNA, RNA or proteins. The versatility of proteins with longer ZnF arrays can be further increased by the usage of distinct finger subsets for different targets.

ZnF proteins are often thought of as predominantly transcriptional regulators. While this is largely true, there are many ZnF proteins with different functions as well, as I illustrated with the examples of PRDM9 and CTCF. The three types of endogenous ZnF proteins I reviewed all vary in their preceding as well as prospective evolution. CTCF is highly conserved in its entirety across multiple species but can nonetheless bind different partners (Filippova et al., 1996) and perform diverse functions due to utilization of distinctive ZnF groups, covalent modifications and other factors. For this reason, selection pressure is acting on CTCF to retain its conserved structure. In contrast, PRDM9 contains a set of highly conserved effector domains which perform a specialized role in forming meiotic recombination hotspots but its DNA-binding ZnF array needs to evolve very rapidly to overcome the constant erosion of its target sequences and maintain the organism’s fertility (Baker et al., 2015). Possibly as a by-product of this accelerated rate of evolution of PRDM9’s ZnF array, this protein might also be the first known endogenous protein directly implicated in speciation (Mihola et al., 2009).

The third ZnF protein type I covered is the family of KZFPs. The function of this massively expanded class of mammalian proteins is not entirely immutable, as they mostly act as transcriptional repressors but can also assume other roles. What probably drove this KZFP multiplication and the expansion of their class was the constant influx of mobile genetic elements into the genome and their ongoing mutation and duplication (Emerson & Thomas, 2009). In response, new KZFPs with novel ZnF array configurations keep evolving to bind, silence and tame the TEs. Other mechanisms that have evolved the capability to suppress

uncontrolled TE transposition are the PIWI-interacting small RNA (piRNA) pathways. These small single-strand RNAs are formed by processing long RNA precursor transcripts and later form complexes with members of the PIWI class of Argonaute proteins, which are then employed in transcriptional or posttranscriptional regulation of targets complementary to the piRNA. piRNA cascades are present predominantly in the germline, where they keep TEs in check and help preserve the germline's integrity (reviewed in Ozata et al., 2019). Extensive studies have been carried out on piRNA pathways in invertebrates, mainly in *Drosophila*, and these results indicate a strong positive selection acting on the piRNA machinery, suggesting a correlation between TE diversification and piRNA evolution (reviewed in Blumenstiel et al., 2016). Strong positive selection of piRNA genes was also demonstrated in fish, but their mammalian homologues were shown to evolve much slower (Yi et al., 2014). This corresponds with the expansion of KZFPs, a different mechanism of TE silencing, in mammalian genomes, suggesting that these proteins have largely replaced piRNA pathways in mammals and further solidifying KZFPs' proposed role in TE regulation.

KZFPs can sometimes even start harnessing the potential of TEs as a regulatory platform for the host's own benefit (Imbeault et al., 2017). These recent discoveries are particularly interesting as they suggest a substantial role of KZFP/TE interplay in establishing transcriptional networks throughout the organism. An especially exciting proposal is that judging by the increased presence of KZFPs and TEs in the central nervous system, these proteins could have played a major role in the evolution of higher brain functions in primates (Ecco et al., 2017) which opens intriguing new directions of further study.

Soon after their first description, ZnFs have become a subject of intense research with the intention of utilizing them for the creation of custom DNA-binding domains coupled to various effector domains, suited for use in research and gene therapy. In 2009, a competing design emerged under the name transcription activator-like effector (TALE) following its discovery in *Xanthomonas* bacteria, offering even greater design flexibility with easier engineering (Boch et al., 2009; Moscou & Bogdanove, 2009). However, a giant leap in genome editing technology came in 2012 when the CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated) system was introduced (Jinek et al., 2012). This simple bacterial system of adaptive immunity against exogenous nucleic acids uses an RNA-guided nuclease to introduce a DSB into the target sequence. All that is needed for its efficient use in targeting any DNA sequence is the programming of a single RNA oligonucleotide complementary to the target sequence (Jinek et al., 2012). Since then, it has taken the scientific world by storm for its

efficiency, ease of use and financial affordability. After catalytical inactivation of the Cas9 nuclease, it can even be combined with other effector domains and used as a general platform for transcriptional regulation (Gilbert et al., 2013). These groundbreaking results led to the gradual disappearance of artificial ZnF proteins from the commonly used tool repertoire of many research groups in favour of CRISPR/Cas9.

Although the future of synthetic ZnF proteins currently doesn't seem to involve their widespread utilization, several niche uses might still stay relevant even with new technologies present. An innovative application may be invented, creating a new field for designer ZnF proteins to shine. In any case, their natural counterparts will probably stay in widespread use across the eukaryotic domain at least for another few million years.

5. References

Reviews are marked with an asterisk.

Bae, K.-H., Do Kwon, Y., Shin, H.-C., Hwang, M.-S., Ryu, E.-H., Park, K.-S., Yang, H.-Y., Lee, D., Lee, Y., Park, J., Sun Kwon, H., Kim, H.-W., Yeh, B.-I., Lee, H.-W., Hyung Sohn, S., Yoon, J., Seol, W., & Kim, J.-S. (2003). Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nature Biotechnology*, *21*(3), 275–280. https://doi.org/10.1038/nbt796

Baker, C. L., Kajita, S., Walker, M., Saxl, R. L., Raghupathy, N., Choi, K., Petkov, P. M., & Paigen, K. (2015). PRDM9 Drives Evolutionary Erosion of Hotspots in Mus musculus through Haplotype-Specific Initiation of Meiotic Recombination. *PLoS Genetics*, *11*(1), e1004916. https://doi.org/10.1371/journal.pgen.1004916

Baker, C. L., Walker, M., Kajita, S., Petkov, P. M., & Paigen, K. (2014). PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Research*, *24*(5), 724–732. https://doi.org/10.1101/GR.170167.113

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., & De Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*, *327*(5967), 836–840. https://doi.org/10.1126/SCIENCE.1183439

Beerli, R. R., Segal, D. J., Dreier, B., & Barbas, C. F. (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. *Proceedings of the National Academy of Sciences*, *95*(25), 14628–14633. https://doi.org/10.1073/pnas.95.25.14628

Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, *405*(6785), 482–485. https://doi.org/10.1038/35013100

Bell, A. C., West, A. G., & Felsenfeld, G. (1999). The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. *Cell*, *98*(3), 387–396. https://doi.org/10.1016/S0092-8674(00)81967-4

Bellefroid, E. J., Poncelet, D. A., Lecocq, P. J., Revelant, O., & Martial, J. A. (1991). The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proceedings of the National Academy of Sciences*, *88*(9), 3608–3612. https://doi.org/10.1073/PNAS.88.9.3608

Berg, I. L., Neumann, R., Lam, K.-W. G., Sarbajna, S., Odenthal-Hesse, L., May, C. A., & Jeffreys, A. J. (2010). PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature Genetics*, *42*(10), 859–863. https://doi.org/10.1038/ng.658

Berg, J. M. (1988). Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(1), 99–102. https://doi.org/10.1073/PNAS.85.1.99

*Bhakta, M. S., & Segal, D. J. (2010). The Generation of Zinc Finger Proteins by Modular Assembly. In J. P. Mackay & D. J. Segal (Eds.), *Engineered Zinc Finger Proteins* (pp. 3–30). Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-60761-753-2_1

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y.-G., & Chandrasegaran, S. (2001). Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Molecular and Cellular Biology*, *21*(1), 289–297. https://doi.org/10.1128/MCB.21.1.289-297.2001

Birtle, Z., & Ponting, C. P. (2006). Meisetz and the birth of the KRAB motif. *Bioinformatics*, *22*(23), 2841–2845. https://doi.org/10.1093/BIOINFORMATICS/BTL498

Bitinaite, J., Wah, D. A., Aggarwal, A. K., & Schildkraut, I. (1998). *Fok*I dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences*, *95*(18), 10570–10575. https://doi.org/10.1073/pnas.95.18.10570

*Blumenstiel, J. P., Erwin, A. A., & Hemmer, L. W. (2016). What Drives Positive Selection in the Drosophila piRNA Machinery? The Genomic Autoimmunity Hypothesis. *The Yale Journal of Biology and Medicine*, *89*(4), 499–512. http://www.ncbi.nlm.nih.gov/pubmed/28018141

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*, *326*(5959), 1509–1512. https://doi.org/10.1126/science.1178811

Boulton, A., Myers, R. S., & Redfield, R. J. (1997). The hotspot conversion paradox and the evolution of meiotic recombination. *Proceedings of the National Academy of Sciences*, *94*(15), 8058–8063. https://doi.org/10.1073/pnas.94.15.8058

Bourque, G., Leong, B., Vega, V. B., Chen, X., Yen, L. L., Srinivasan, K. G., Chew, J. L., Ruan, Y., Wei, C. L., Huck, H. N., & Liu, E. T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research*, *18*(11), 1752–1762. https://doi.org/10.1101/GR.080663.108

Bowers, P. M., Schaufler, L. E., & Klevit, R. E. (1999). A folding transition and novel zinc finger accessory domain in the transcription factor ADR1. *Nature Structural Biology 1999 6:5*, *6*(5), 478–485. https://doi.org/10.1038/8283

Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R. D., & Petukhova, G. V. (2012). Genetic recombination is directed away from functional genomic elements in mice. *Nature*, *485*(7400), 642–645. https://doi.org/10.1038/nature11089

*Brown, D. D. (1984). The role of stable complexes that repress and activate eucaryotic genes. *Cell*, *37*(2), 359–365. https://doi.org/10.1016/0092-8674(84)90366-0

Buard, J., Rivals, E., Dunoyer de Segonzac, D., Garres, C., Caminade, P., de Massy, B., & Boursot, P. (2014). Diversity of Prdm9 Zinc Finger Array in Wild Mice Unravels New Facets of the Evolutionary Turnover of this Coding Minisatellite. *PLoS ONE*, *9*(1), e85021. https://doi.org/10.1371/journal.pone.0085021

Caricasole, A., Duarte, A., Larsson, S. H., Hastie, N. D., Little, M., Holmes, G., Todorov, I., & Ward, A. (1996). RNA binding by the Wilms tumor suppressor zinc finger proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(15), 7562–7566. https://doi.org/10.1073/PNAS.93.15.7562

*Carroll, D. (2011a). Genome Engineering With Zinc-Finger Nucleases. *Genetics*, *188*(4), 773–782. https://doi.org/10.1534/genetics.111.131433

*Carroll, D. (2011b). Zinc-Finger Nucleases: A Panoramic View. *Current Gene Therapy*, *11*(1), 2–10. https://doi.org/10.2174/156652311794520076

Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S. M., Turelli, P., & Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes & Development*, *28*(13), 1397–1409. https://doi.org/10.1101/gad.241661.114

*Chandrasegaran, S., & Carroll, D. (2016). Origins of Programmable Nucleases for Genome Engineering. *Journal of Molecular Biology*, *428*(5), 963–989. https://doi.org/10.1016/j.jmb.2015.10.014

Choo, Y., & Klug, A. (1994a). Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proceedings of the National Academy of Sciences*, *91*(23), 11168–11172. https://doi.org/10.1073/pnas.91.23.11168

Choo, Y., & Klug, A. (1994b). Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proceedings of the National Academy of Sciences*, *91*(23), 11163–11167. https://doi.org/10.1073/pnas.91.23.11163

Choo, Y., Sánchez-García, I., & Klug, A. (1994). In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature*, *372*(6507), 642–645. https://doi.org/10.1038/372642a0

*Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics*, *18*(2), 71–86. https://doi.org/10.1038/nrg.2016.139

*Cooper, G. M., Nickerson, D. A., & Eichler, E. E. (2007). Mutational and selective effects on copy-number variants in the human genome. *Nature Genetics 2007 39:7*, *39*(7), S22–S29. https://doi.org/10.1038/NG2054

Corsinotti, A., Kapopoulou, A., Gubelmann, C., Imbeault, M., Santoni de Sio, F. R., Rowe, H. M., Mouscaz, Y., Deplancke, B., & Trono, D. (2013). Global and Stage Specific Patterns of Krüppel-Associated-Box Zinc Finger Protein Gene Expression in Murine Early Embryonic Cells. *PLOS ONE*, *8*(2), e56721. https://doi.org/10.1371/JOURNAL.PONE.0056721

Cuddapah, S., Jothi, R., Schones, D. E., Roh, T.-Y., Cui, K., & Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Research*, *19*(1), 24–32. https://doi.org/10.1101/gr.082800.108

de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLOS Genetics*, *7*(12), e1002384. https://doi.org/10.1371/JOURNAL.PGEN.1002384

Diakun, G. P., Fairall, L., & Klug, A. (1986). EXAFS study of the zinc-binding sites in the protein transcription factor IIIA. *Nature*, *324*(6098), 698–699. https://doi.org/10.1038/324698a0

Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G., & West, A. G. (2010). VEZF1 elements mediate protection from DNA methylation. *PLoS Genetics*, *6*(1). https://doi.org/10.1371/JOURNAL.PGEN.1000804

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D., & Barbas, C. F. (2001). Development of Zinc Finger Domains for Recognition of the 5′-ANN-3′ Family of DNA Sequences and Their Use in the Construction of Artificial Transcription Factors. *Journal of Biological Chemistry*, *276*(31), 29466–29478. https://doi.org/10.1074/JBC.M102604200

Dreier, B., Fuller, R. P., Segal, D. J., Lund, C. V., Blancafort, P., Huber, A., Koksch, B., & Barbas, C. F. (2005). Development of Zinc Finger Domains for Recognition of the 5′-CNN-3′ Family DNA Sequences and Their Use in the Construction of Artificial Transcription Factors. *Journal of Biological Chemistry*, *280*(42), 35588–35597. https://doi.org/10.1074/jbc.M506654200

Dutnall, R. N., Neuhaus, D., & Rhodes, D. (1996). The solution structure of the first zinc finger domain of SWI5: a novel structural extension to a common fold. *Structure*, *4*(5), 599–611. https://doi.org/10.1016/S0969-2126(96)00064-0

Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H. M., Turelli, P., & Trono, D. (2016). Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Developmental Cell*, *36*(6), 611–623. https://doi.org/10.1016/J.DEVCEL.2016.02.024

*Ecco, G., Imbeault, M., & Trono, D. (2017). KRAB zinc finger proteins. *Development*, *144*(15), 2719–2729. https://doi.org/10.1242/dev.132605

*Edelstein, L. C., & Collins, T. (2005). The SCAN domain family of zinc finger transcription factors. *Gene,* 359, 1-17. https://doi.org/10.1016/j.gene.2005.06.022

Elrod-Erickson, M., Benson, T. E., & Pabo, C. O. (1998). High-resolution structures of variant Zif268–DNA complexes: implications for understanding zinc finger–DNA recognition. *Structure*, *6*(4), 451–464. https://doi.org/10.1016/S0969-2126(98)00047-1

Emerson, R. O., & Thomas, J. H. (2009). Adaptive Evolution in Zinc Finger Transcription Factors. *PLOS Genetics*, *5*(1), e1000325. https://doi.org/10.1371/JOURNAL.PGEN.1000325

*Erwin, J. A., Marchetto, M. C., & Gage, F. H. (2014). Mobile DNA elements in the generation of diversity and complexity in the brain. *Nature Reviews Neuroscience*, *15*(8), 497–506. https://doi.org/10.1038/nrn3730

Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T., & Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature*, *366*(6454), 483–487. https://doi.org/10.1038/366483a0

Farmiloe, G., Lodewijk, G. A., Robben, S. F., van Bree, E. J., & Jacobs, F. M. J. (2020). Widespread correlation of KRAB zinc finger protein binding with brain-developmental gene expression patterns. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1795). https://doi.org/10.1098/RSTB.2019.0333

Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P. E., Collins, S. J., & Lobanenkov, V. V. (1996). An Exceptionally Conserved Transcriptional Repressor, CTCF, Employs Different Combinations of Zinc Fingers To Bind Diverged Promoter Sequences of Avian and Mammalian c-myc Oncogenes. *Molecular and Cellular Biology*, *16*(6), 2802–2813. https://doi.org/10.1128/MCB.16.6.2802

*Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends in Genetics*, *5*(C), 103–107. https://doi.org/10.1016/0168-9525(89)90039-5

Forejt, J., & Iványi, P. (1974). Genetic studies on male sterility of hybrids between laboratory and wild mice (Mus musculus L.). *Genetics Research*, *24*(2), 189–206. https://doi.org/10.1017/S0016672300015214

Fox, A. H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., & Crossley, M. (1999). Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. *The EMBO Journal*, *18*(10), 2812–2822. https://doi.org/10.1093/emboj/18.10.2812

*Friedli, M., & Trono, D. (2015). The Developmental Control of Transposable Elements and the Evolution of Higher Species. *Annu. Rev. Cell Dev. Biol*, *31*, 429–451. https://doi.org/10.1146/annurev-cellbio-100814-125514

Friedman, J. R., Fredericks, W. J., Jensen, D. E., Speicher, D. W., Huang, X. P., Neilson, E. G., & Rauscher, F. J. (1996). KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes & Development*, *10*(16), 2067–2078. https://doi.org/10.1101/GAD.10.16.2067

*Frietze, S., & Farnham, P. J. (2011). Transcription Factor Effector Domains. In T. R. Hughes (Ed.), *A Handbook of Transcription Factors* (52nd ed., Vol. 52, pp. 261–277). https://doi.org/10.1007/978-90-481-9069-0

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, *15*(9), 2038–2049. https://doi.org/10.1016/j.celrep.2016.04.085

*Gaj, T., Gersbach, C. A., & Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, *31*(7), 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004

*Garcia-Perez, J. L., Widmann, T. J., & Adams, I. R. (2016). The impact of transposable elements on mammalian development. *Development*, *143*(22), 4101–4114. https://doi.org/10.1242/DEV.132639

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, *154*(2), 442–451. https://doi.org/10.1016/j.cell.2013.06.044

Ginsberg, A. M., King, B. 0, & Roeder, R. G. (1984). Xenopus 5S Gene Transcription Factor, TFIIIA : Characterization of a cDNA Clone and Measurement of RNA Levels throughout Development. *Cell*, *39*, 79–489.

Gordley, R. M., Smith, J. D., Gräslund, T., & Barbas, C. F. (2007). Evolution of Programmable Zinc Finger-recombinases with Activity in Human Cells. *Journal of Molecular Biology*, *367*(3), 802–813. https://doi.org/10.1016/j.jmb.2007.01.017

Gregory, R., Taxman, D., Seshasayee, D., Kensinger, M., Bieker, J., & Wojchowski, D. (1996). Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. *Blood*, *87*(5), 1793–1801. https://doi.org/10.1182/blood.V87.5.1793.bloodjournal8751793

Grishin, N. V. (2001). Treble clef finger-a functionally diverse zinc-binding structural motif. *Nucleic Acids Research*, *29*(8), 1703–1714.

Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H.-L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., Bates, J. G., Richards, N., Myers, D., Patel, H., Gallagher, M., Schlissel, M. S., Murre, C., Busslinger, M., Giallourakis, C. C., & Alt, F. W. (2011). CTCF-binding elements mediate control of V(D)J recombination. *Nature*, *477*(7365), 424–430. https://doi.org/10.1038/nature10495

Guo, J., Gaj, T., & Barbas, C. F. (2010). Directed Evolution of an Enhanced and Highly Efficient FokI Cleavage Domain for Zinc Finger Nucleases. *Journal of Molecular Biology*, *400*(1), 96–107. https://doi.org/10.1016/j.jmb.2010.04.060

*Hancks, D. C., & Kazazian, H. H. (2016). Roles for retrotransposon insertions in human disease. *Mobile DNA*, *7*(1), 9. https://doi.org/10.1186/s13100-016-0065-9

Hardwick, R. J., Tretyakov, M. V., & Dubrova, Y. E. (2009). Age-Related Accumulation of Mutations Supports a Replication-Dependent Mechanism of Spontaneous Mutation at Tandem Repeat DNA Loci in Mice. *Molecular Biology and Evolution*, *26*(11), 2647–2654. https://doi.org/10.1093/MOLBEV/MSP182

Hata, A., Seoane, J., Lagna, G., Montalvo, E., Hemmati-Brivanlou, A., & Massagué, J. (2000). OAZ Uses Distinct DNA-and Protein-Binding Zinc Fingers in Separate BMP-Smad and Olf Signaling Pathways. *Cell*, *100*, 229–240.

Hayashi, K., Yoshida, K., & Matsui, Y. (2005). A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature*, *438*(7066), 374–378. https://doi.org/10.1038/nature04112

Hoffmann, A., Ciani, E., Boeckardt, J., Holsboer, F., Journot, L., & Spengler, D. (2003). Transcriptional Activities of the Zinc Finger Protein Zac Are Differentially Controlled by DNA Binding. *Molecular and Cellular Biology*, *23*(3), 988–1003. https://doi.org/10.1128/MCB.23.3.988-1003.2003

Huntley, S., Baggott, D. M., Hamilton, A. T., Tran-Gyamfi, M., Yang, S., Kim, J., Gordon, L., Branscomb, E., & Stubbs, L. (2006). A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Research*, *16*(5), 669–677. https://doi.org/10.1101/GR.4842106

Imai, Y., Baudat, F., Taillepierre, M., Stanzione, M., Toth, A., & de Massy, B. (2017). The PRDM9 KRAB domain is required for meiosis and involved in protein interactions. *Chromosoma*, *126*(6), 681–695. https://doi.org/10.1007/s00412-017-0631-z

Imbeault, M., Helleboid, P. Y., & Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*, *543*(7646), 550–554. https://doi.org/10.1038/nature21683

*Iuchi, S. (2001). Three classes of C2H2 zinc finger proteins. *Cellular and Molecular Life Sciences*, *58*(4), 625–635. https://doi.org/10.1007/PL00000885

Jeffreys, A. J., Kauppi, L., & Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics*, *29*(2), 217–222. https://doi.org/10.1038/ng1001-217

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Kim, H. J., Lee, H. J., Kim, H., Cho, S. W., & Kim, J.-S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Research*, *19*(7), 1279–1288. https://doi.org/10.1101/gr.089417.108

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, *93*(3), 1156–1160. https://doi.org/10.1073/pnas.93.3.1156

*Klug, A. (2010). The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. *Annual Review of Biochemistry*, *79*(1), 213–231. https://doi.org/10.1146/annurev-biochem-010909-095056

*Krishna, S. S., Majumdar, I., & Grishin, N. V. (2003). Structural classification of zinc fingers: survey and summary. *Nucleic Acids Research*, *31*(2), 532–550. https://doi.org/10.1093/nar/gkg161

*Laity, J. H., Lee, B. M., & Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology*, *11*(1), 39–46. https://doi.org/10.1016/S0959-440X(00)00167-6

Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K., Bouhlel, M. A., Richter, J., Soler, E., Stadhouders, R., Jöhrens, K., Wurster, K. D., Callen, D. F., Harte, M. F., Giefing, M., Barlow, R., Stein, H., Anagnostopoulos, I., … Mathas, S. (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nature Medicine*, *16*(5), 571–579. https://doi.org/10.1038/nm.2129

Lange, J., Yamada, S., Tischfield, S. E., Pan, J., Kim, S., Zhu, X., Socci, N. D., Jasin, M., & Keeney, S. (2016). The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair. *Cell*, *167*(3), 695-708.e16. https://doi.org/10.1016/j.cell.2016.09.035

Lee, H. J., Kim, E., & Kim, J.-S. (2010). Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Research*, *20*(1), 81–89. https://doi.org/10.1101/gr.099747.109

Lee, J. S., Galvin, K. M., & Shi, Y. (1993). Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proceedings of the National Academy of Sciences*, *90*(13), 6145–6149. https://doi.org/10.1073/PNAS.90.13.6145

Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., & Wright, P. E. (1989). Three-Dimensional Solution Structure of a Single Zinc Finger DNA-Binding Domain. *Science*, *245*(4918), 635–637. https://doi.org/10.1126/SCIENCE.2503871

Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S. Y., Bhagwat, A. S., Malani, N., Anguela, X. M., Sharma, R., Ivanciu, L., Murphy, S. L., Finn, J. D., Khazi, F. R., Zhou, S., Paschon, D. E., Rebar, E. J., Bushman, F. D., Gregory, P. D., Holmes, M. C., & High, K. A. (2011). In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature*, *475*(7355), 217–221. https://doi.org/10.1038/nature10177

Lin, Y. C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C. K., & Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nature Immunology*, *13*(12), 1196–1204. https://doi.org/10.1038/ni.2432

Lobanenkov, V. V, Nicolas, R. H., Adler, V. V, Paterson, H., Klenova, E. M., Polotskaja, A. V, & Goodwin, G. H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene*, *5*, 1743–1753.

MacPherson, M. J., Beatty, L. G., Zhou, W., Du, M., & Sadowski, P. D. (2009). The CTCF Insulator Protein Is Posttranslationally Modified by SUMO. *Molecular and Cellular Biology*, *29*(3), 714–725. https://doi.org/10.1128/MCB.00825-08

Maeder, M. L., Thibodeau-Beganny, S., Sander, J. D., Voytas, D. F., & Joung, J. K. (2009). Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nature Protocols*, *4*(10), 1471–1501. https://doi.org/10.1038/nprot.2009.98

Majumder, P., Gomez, J. A., Chadwick, B. P., & Boss, J. M. (2008). The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *The Journal of Experimental Medicine*, *205*(4), 785–798. https://doi.org/10.1084/jem.20071843

*Maksakova, I. A., Mager, D. L., & Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cellular and Molecular Life Sciences*, *65*(21), 3329–3347. https://doi.org/10.1007/s00018-008-8494-3

Martin, F. J., Amode, M. R., Aneja, A., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Bignell, A., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Charkhchi, M., Cockburn, A., Da Rin Fiorretto, L., … Flicek, P. (2023). Ensembl 2023. *Nucleic Acids Research*, *51*(D1), D933–D941. https://doi.org/10.1093/NAR/GKAC958

Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M. C., & Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, *464*(7290), 927–931. https://doi.org/10.1038/nature08858

*Matthews, J. M., & Sunde, M. (2002). Zinc fingers - Folds for many occasions. *IUBMB Life*, *54*(6), 351–355. https://doi.org/10.1080/15216540216035

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences*, *36*(6), 344–355. https://doi.org/10.1073/pnas.36.6.344

Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C., & Forejt, J. (2009). A Mouse Speciation Gene Encodes a Meiotic Histone H3 Methyltransferase. *Science*, *323*(5912), 373–375. https://doi.org/10.1126/science.1163601

Miki, Y., Katagiri, T., Kasumi, F., Yoshimoto, T., & Nakamura, Y. (1996). Mutation analysis in the BRCA2 gene in primary breast cancers. *Nature Genetics*, *13*(2), 245–247. https://doi.org/10.1038/ng0696-245

Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y.-L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A., Gregory, P. D., Pabo, C. O., & Rebar, E. J. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnology*, *25*(7), 778–785. https://doi.org/10.1038/nbt1319

Miller, J., McLachlan, A. D., & Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO Journal*, *4*(6), 1609–1614.

Moehle, E. A., Rock, J. M., Lee, Y.-L., Jouvenot, Y., DeKelver, R. C., Gregory, P. D., Urnov, F. D., & Holmes, M. C. (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences*, *104*(9), 3055–3060. https://doi.org/10.1073/pnas.0611478104

Moscou, M. J., & Bogdanove, A. J. (2009). A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science*, *326*(5959), 1501–1501. https://doi.org/10.1126/science.1178817

Myers, S., Bowden, R., Tumian, A., Bontrop, R. E., Freeman, C., MacFie, T. S., McVean, G., & Donnelly, P. (2010). Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*, *327*(5967), 876–879. https://doi.org/10.1126/SCIENCE.1182363

Najafabadi, H. S., Mnaimneh, S., Schmitges, F. W., Garton, M., Lam, K. N., Yang, A., Albu, M., Weirauch, M. T., Radovani, E., Kim, P. M., Greenblatt, J., Frey, B. J., & Hughes, T. R. (2015). C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nature Biotechnology*, *33*(5), 555–562. https://doi.org/10.1038/nbt.3128

Nakahashi, H., Kwon, K. R. K., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., Qian, J., Dubois, W., Welsh, S., Phair, R. D., Pugh, B. F., Lobanenkov, V., Hager, G. L., & Casellas, R. (2013). A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. *Cell Reports*, *3*(5), 1678–1689. https://doi.org/10.1016/J.CELREP.2013.04.024

Neuhaus, D., Nakaseko, Y., Schwabe, J. W. R., & Klug, A. (1992). Solution structures of two zinc-finger domains from SWI5 obtained using two-dimensional 1H nuclear magnetic resonance spectroscopy. *Journal of Molecular Biology*, *228*(2), 637–651. https://doi.org/10.1016/0022-2836(92)90846-C

Nielsen, A. L., Ortiz, J. A., You, J., Oulad-Abdelghani, M., Khechumian, R., Gansmuller, A., Chambon, P., & Losson, R. (1999). Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *The EMBO Journal*, *18*(22), 6385–6395. https://doi.org/10.1093/EMBOJ/18.22.6385

Nowick, K., Hamilton, A. T., Zhang, H., & Stubbs, L. (2010). Rapid Sequence and Expression Divergence Suggest Selection for Novel Function in Primate-Specific KRAB-ZNF Genes. *Molecular Biology and Evolution*, *27*(11), 2606–2617. https://doi.org/10.1093/MOLBEV/MSQ157

*Ohlsson, R., Renkawitz, R., & Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends in Genetics*, *17*(9), 520–527. https://doi.org/10.1016/S0168-9525(01)02366-6

Oliver, P. L., Goodstadt, L., Bayes, J. J., Birtle, Z., Roach, K. C., Phadnis, N., Beatson, S. A., Lunter, G., Malik, H. S., & Ponting, C. P. (2009). Accelerated Evolution of the Prdm9 Speciation Gene across

Diverse Metazoan Taxa. *PLoS Genetics*, *5*(12), e1000753. https://doi.org/10.1371/journal.pgen.1000753

*Ong, C.-T., & Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nature Reviews Genetics*, *15*(4), 234–246. https://doi.org/10.1038/nrg3663

*Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D., & Zamore, P. D. (2019). PIWI-interacting RNAs: small RNAs with big functions. In *Nature Reviews Genetics* (Vol. 20, Issue 2, pp. 89–108). Nature Publishing Group. https://doi.org/10.1038/s41576-018-0073-3

*Paigen, K., & Petkov, P. M. (2018). PRDM9 and Its Role in Genetic Recombination. *Trends in Genetics*, *34*(4), 291–300. https://doi.org/10.1016/J.TIG.2017.12.017

Parvanov, E. D., Petkov, P. M., & Paigen, K. (2010). *Prdm9* Controls Activation of Mammalian Recombination Hotspots. *Science*, *327*(5967), 835–835. https://doi.org/10.1126/science.1181495

Parvanov, E. D., Tian, H., Billings, T., Saxl, R. L., Spruce, C., Aithal, R., Krejci, L., Paigen, K., & Petkov, P. M. (2017). PRDM9 interactions with other proteins provide a link between recombination hotspots and the chromosomal axis in meiosis. *Molecular Biology of the Cell*, *28*(3), 488–499. https://doi.org/10.1091/mbc.e16-09-0686

Pavletich, N. P., & Pabo, C. O. (1991). Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. *Science*, *252*(5007), 809–817. https://doi.org/10.1126/SCIENCE.2028256

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Azaro Pinto, B. L. ´, Salazar, G. A., Bileschi, M. L., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letuní, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J. A., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, *51*. https://doi.org/10.1093/nar/gkac993

Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V, Lee, Y.-L., Guschin, D. Y., Rupniewski, I., Waite, A. J., Carpenito, C., Carroll, R. G., S Orange, J., Urnov, F. D., Rebar, E. J., Ando, D., … June, C. H. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature Biotechnology*, *26*(7), 808–816. https://doi.org/10.1038/nbt1410

Perkins, N. D., Agranoff, A. B., Pascal, E., & Nabel, G. J. (1994). An interaction between the DNA-binding domains of RelA(p65) and Sp1 mediates human immunodeficiency virus gene activation. *Molecular and Cellular Biology*, *14*(10), 6570–6583. https://doi.org/10.1128/MCB.14.10.6570

*Petronczki, M., Siomos, M. F., & Nasmyth, K. (2003). Un Ménage à Quatre: The Molecular Biology of Chromosome Segregation in Meiosis. *Cell*, *112*(4), 423–440. https://doi.org/10.1016/S0092-8674(03)00083-7

*Phillips, J. E., & Corces, V. G. (2009). CTCF: Master Weaver of the Genome. *Cell*, *137*(7), 1194–1211. https://doi.org/10.1016/j.cell.2009.06.001

Picard, B., & Wegnez, M. (1979). Isolation of a 7S particle from Xenopus laevis oocytes: a 5S RNA-protein complex. *Proceedings of the National Academy of Sciences*, *76*(1), 241–245. https://doi.org/10.1073/pnas.76.1.241

*Ponting, C. P. (2011). What are the genomic drivers of the rapid evolution of PRDM9? *Trends in Genetics*, *27*(5), 165–171. https://doi.org/10.1016/j.tig.2011.02.001

Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., Theunissen, T. W., Jaenisch, R., & Trono, D. (2019). Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell*, *24*(5), 724-735.e5. https://doi.org/10.1016/j.stem.2019.03.012

Powers, N. R., Parvanov, E. D., Baker, C. L., Walker, M., Petkov, P. M., & Paigen, K. (2016). The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLoS Genetics*, *12*(6). https://doi.org/10.1371/JOURNAL.PGEN.1006146

Puget, N., Sinilnikova, O. M., Stoppa-Lyonnet, D., Audoynaud, C., Pages, S., Lynch, H. T., Goldgar, D., Lenoir, G. M., & Mazoyer, S. (1999). An Alu-Mediated 6-kb Duplication in the BRCA1 Gene: A New Founder Mutation? *The American Journal of Human Genetics*, *64*(1), 300–302. https://doi.org/10.1086/302211

Quenneville, S., Turelli, P., Bojkowska, K., Raclot, C., Offner, S., Kapopoulou, A., & Trono, D. (2012). The KRAB-ZFP/KAP1 System Contributes to the Early Embryonic Establishment of Site-Specific DNA Methylation Patterns Maintained during Development. *Cell Reports*, *2*(4), 766–773. https://doi.org/10.1016/J.CELREP.2012.08.043

Ramirez, C. L., Certo, M. T., Mussolino, C., Goodwin, M. J., Cradick, T. J., McCaffrey, A. P., Cathomen, T., Scharenberg, A. M., & Joung, J. K. (2012). Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Research*, *40*(12), 5560–5568. https://doi.org/10.1093/nar/gks179

*Rosspopoff, O., & Trono, D. (2023). Take a walk on the KRAB side. *Trends in Genetics*, *39*(11), 844–857. https://doi.org/10.1016/J.TIG.2023.08.003

Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D. B., & Trono, D. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, *463*(7278), 237–240. https://doi.org/10.1038/nature08674

Sanborn, A. L., Rao, S. S. P., Huang, S.-C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochkov, I. D., Chinnappan, D., Cutkosky, A., Li, J., Geeting, K. P., Gnirke, A., Melnikov, A., McKenna, D., Stamenova, E. K., Lander, E. S., & Aiden, E. L. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*, *112*(47), E6456–E6465. https://doi.org/10.1073/pnas.1518552112

Santiago, Y., Chan, E., Liu, P.-Q., Orlando, S., Zhang, L., Urnov, F. D., Holmes, M. C., Guschin, D., Waite, A., Miller, J. C., Rebar, E. J., Gregory, P. D., Klug, A., & Collingwood, T. N. (2008). Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proceedings of the National Academy of Sciences*, *105*(15), 5809–5814. https://doi.org/10.1073/pnas.0800940105

Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature*, *489*(7414), 109–113. https://doi.org/10.1038/nature11279

Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G., & Rauscher, F. J. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development*, *16*(8), 919–932. https://doi.org/10.1101/GAD.973302

Schultz, D. C., Friedman, J. R., & Rauscher, F. J. (2001). Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2α subunit of NuRD. *Genes & Development*, *15*(4), 428–443. https://doi.org/10.1101/GAD.869501

Sebastiano, V., Maeder, M. L., Angstman, J. F., Haddad, B., Khayter, C., Yeo, D. T., Goodwin, M. J., Hawkins, J. S., Ramirez, C. L., Batista, L. F. Z., Artandi, S. E., Wernig, M., & Joung, J. K. (2011). In Situ Genetic Correction of the Sickle Cell Anemia Mutation in Human Induced Pluripotent Stem Cells Using Engineered Zinc Finger Nucleases. *Stem Cells*, *29*(11), 1717–1726. https://doi.org/10.1002/stem.718

Segal, D. J., Dreier, B., Beerli, R. R., & Barbas, C. F. (1999). Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. *Proceedings of the National Academy of Sciences*, *96*(6), 2758–2763. https://doi.org/10.1073/pnas.96.6.2758

Segal, D. J., Gonçalves, J., Eberhardy, S., Swan, C. H., Torbett, B. E., Li, X., & Barbas, C. F. (2004). Attenuation of HIV-1 Replication in Primary Human Cells with a Designed Zinc Finger Transcription Factor. *Journal of Biological Chemistry*, *279*(15), 14509–14519. https://doi.org/10.1074/jbc.M400349200

Shannon, M., Hamilton, A. T., Gordon, L., Branscomb, E., & Stubbs, L. (2003). Differential Expansion of Zinc-Finger Transcription Factor Loci in Homologous Human and Mouse Gene Clusters. *Genome Research*, *13*(6a), 1097–1110. https://doi.org/10.1101/GR.963903

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., & Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, *488*(7409), 116–120. https://doi.org/10.1038/nature11243

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., & Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, *479*(7371), 74–79. https://doi.org/10.1038/nature10442

*Smit, A. F. (1996). The origin of interspersed repeats in the human genome. *Current Opinion in Genetics & Development*, *6*(6), 743–748. https://doi.org/10.1016/S0959-437X(96)80030-X

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S. M., Schroth, G. P., Tanay, A., & Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. *The EMBO Journal*, *32*(24), 3119–3129. https://doi.org/10.1038/emboj.2013.237

Soldner, F., Laganière, J., Cheng, A. W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L. I., Myers, R. H., Lindquist, S., Zhang, L., Guschin, D., Fong, L. K., Vu, B. J., Meng, X., Urnov, F. D., Rebar, E. J., Gregory, P. D., Zhang, H. S., & Jaenisch, R. (2011). Generation of Isogenic Pluripotent Stem Cells Differing Exclusively at Two Early Onset Parkinson Point Mutations. *Cell*, *146*(2), 318–331. https://doi.org/10.1016/j.cell.2011.06.019

Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., & de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes & Development*, *20*(17), 2349–2354. https://doi.org/10.1101/gad.399506

Sripathy, S. P., Stevens, J., & Schultz, D. C. (2006). The KAP1 Corepressor Functions To Coordinate the Assembly of De Novo HP1-Demarcated Microenvironments of Heterochromatin Required for KRAB Zinc Finger Protein-Mediated Transcriptional Repression. *Molecular and Cellular Biology*, *26*(22), 8623–8638. https://doi.org/10.1128/MCB.00487-06

Stains, C. I., Furman, J. L., Segal, D. J., & Ghosh, I. (2006). Site-Specific Detection of DNA Methylation Utilizing mCpG-SEER. *Journal of the American Chemical Society*, *128*(30), 9761–9765. https://doi.org/10.1021/ja060681j

*Stubbs, L., Sun, Y., & Caetano-Anolles, D. (2011). Function and Evolution of C2H2 Zinc Finger Arrays. In T. R. Hughes (Ed.), *A Handbook of Transcription Factors* (Vol. 52, pp. 75–94). Springer Netherlands. https://doi.org/10.1007/978-90-481-9069-0

Sun, L., Liu, A., & Georgopoulos, K. (1996). Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. *The EMBO Journal*, *15*(19), 5358–5369. https://doi.org/10.1002/j.1460-2075.1996.tb00920.x

Szabó, P. E., Tang, S.-H. E., Rentsendorj, A., Pfeifer, G. P., & Mann, J. R. (2000). Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. *Current Biology*, *10*(10), 607–610. https://doi.org/10.1016/S0960-9822(00)00489-9

Szczepek, M., Brondani, V., Büchel, J., Serrano, L., Segal, D. J., & Cathomen, T. (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature Biotechnology*, *25*(7), 786–793. https://doi.org/10.1038/nbt1317

Tebas, P., Stein, D., Tang, W. W., Frank, I., Wang, S. Q., Lee, G., Spratt, S. K., Surosky, R. T., Giedlin, M. A., Nichol, G., Holmes, M. C., Gregory, P. D., Ando, D. G., Kalos, M., Collman, R. G., Binder-Scholl, G., Plesa, G., Hwang, W.-T., Levine, B. L., & June, C. H. (2014). Gene Editing of *CCR5* in Autologous CD4 T Cells of Persons Infected with HIV. *New England Journal of Medicine*, *370*(10), 901–910. https://doi.org/10.1056/NEJMoa1300662

Thomas, J. H., & Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. *Genome Research*, *21*(11), 1800–1812. https://doi.org/10.1101/GR.121749.111

Ting, C. N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C., & Meisler, M. H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes & Development*, *6*(8), 1457–1465. https://doi.org/10.1101/GAD.6.8.1457

Trizzino, M., Park, Y. S., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., Perry, G. H., Lynch, V. J., & Brown, C. D. (2017). Transposable elements are the primary source of novelty in primate gene regulation. *Genome Research*, *27*(10), 1623–1633. https://doi.org/10.1101/GR.218149.116

Tso, J. Y., Berg, D. J. van Den, & Korn, L. J. (1986). Structure of the gene for Xenopus transcription factor TFIIIA. *Nucleic Acids Research*, *14*(5), 2187–2200. https://doi.org/10.1093/NAR/14.5.2187

Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., & Trono, D. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Research*, *24*(8), 1260–1270. https://doi.org/10.1101/GR.172833.114

Turelli, P., Playfoot, C., Grun, D., Raclot, C., Pontis, J., Coudray, A., Thorball, C., Duc, J., Pankevich, E. V., Deplancke, B., Busskamp, V., & Trono, D. (2020). Primate-restricted KRAB zinc finger proteins and target retrotransposons control gene expression in human neurons. *Science Advances*, *6*(35). https://doi.org/10.1126/sciadv.aba3200

Urnov, F. D., Miller, J. C., Lee, Y.-L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., & Holmes, M. C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, *435*(7042), 646–651. https://doi.org/10.1038/nature03556

*Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., & Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics 2009 10:4*, *10*(4), 252–263. https://doi.org/10.1038/NRG2538

Vissing, H., Meyer, W. K. H., Aagaard, L., Tommerup, N., & Thiesen, H. J. (1995). Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. *FEBS Letters*, *369*(2–3), 153–157. https://doi.org/10.1016/0014-5793(95)00728-R

Voigt, K., Gogol-Döring, A., Miskey, C., Chen, W., Cathomen, T., Izsvák, Z., & Ivics, Z. (2012). Retargeting Sleeping Beauty Transposon Insertions by Engineered Zinc Finger DNA-binding Domains. *Molecular Therapy*, *20*(10), 1852–1862. https://doi.org/10.1038/mt.2012.126

Vostrov, A. A., & Quitschke, W. W. (1997). The Zinc Finger Protein CTCF Binds to the APBβ Domain of the Amyloid β-Protein Precursor Promoter. *Journal of Biological Chemistry*, *272*(52), 33353–33359. https://doi.org/10.1074/jbc.272.52.33353

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., Thurman, R. E., Kaul, R., Myers, R. M., & Stamatoyannopoulos, J. A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research*, *22*(9), 1680–1688. https://doi.org/10.1101/gr.136101.111

Wang, T., Zeng, J., Lowe, C. B., Sellers, R. G., Salama, S. R., Yang, M., Burgess, S. M., Brachmann, R. K., & Haussler, D. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proceedings of the National Academy of Sciences*, *104*(47), 18613–18618. https://doi.org/10.1073/pnas.0703637104

*Weiner, A. M., Deininger, P. L., & Efstratiadis, A. (1986). Nonviral retrotransposons: genes, pseudogenes and transposable elements generated by the reverse flow of genetic information. *Annual Review of Biochemistry*, *55*(1), 631–661. https://doi.org/10.1146/annurev.bi.55.070186.003215

*Weirauch, M. T., & Hughes, T. R. (2011). A Catalogue of Eukaryotic Transcription Factor Types, Their Evolutionary Origin, and Species Distribution. In T. R. Hughes (Ed.), *A Handbook of Transcription Factors* (Vol. 52, pp. 25–73). Springer Netherlands. https://doi.org/10.1007/978-90-481-9069-0

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K., & Peters, J.-M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, *451*(7180), 796–801. https://doi.org/10.1038/nature06634

*Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973–982. https://doi.org/10.1038/nrg2165

Wilber, A., Tschulena, U., Hargrove, P. W., Kim, Y.-S., Persons, D. A., Barbas, C. F., & Nienhuis, A. W. (2010). A zinc-finger transcriptional activator designed to interact with the γ-globin gene promoters enhances fetal hemoglobin production in primary human adult erythroblasts. *Blood*, *115*(15), 3033–3041. https://doi.org/10.1182/blood-2009-08-240556

*Williams, R. J. P. (1984). Zinc: what is its role in biology? *Endeavour*, *8*(2), 65–70. https://doi.org/10.1016/0160-9327(84)90040-1

Witzgall, R., O'Leary, E., Leaf, A., Önaldi, D., & Bonventre, J. V. (1994). The Krüppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proceedings of the National Academy of Sciences*, *91*(10), 4514–4518. https://doi.org/10.1073/PNAS.91.10.4514

Xiao, T., Wallace, J., & Felsenfeld, G. (2011). Specific Sites in the C Terminus of CTCF Interact with the SA2 Subunit of the Cohesin Complex and Are Required for Cohesin-Dependent Insulation Activity. *Molecular and Cellular Biology*, *31*(11), 2174–2183. https://doi.org/10.1128/MCB.05093-11

Yi, M., Chen, F., Luo, M., Cheng, Y., Zhao, H., Cheng, H., & Zhou, R. (2014). Rapid evolution of piRNA pathway in the Teleost fish: Implication for an adaptation to transposon diversity. *Genome Biology and Evolution*, *6*(6), 1393–1407. https://doi.org/10.1093/gbe/evu105

Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C., Oshimura, M., Feinberg, A. P., Lobanenkov, V., Klenova, E., & Ohlsson, R. (2004). Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nature Genetics*, *36*(10), 1105–1110. https://doi.org/10.1038/ng1426

*Zhang, H.-X., Zhang, Y., & Yin, H. (2019). Genome Editing with mRNA Encoding ZFN, TALEN, and Cas9. *Molecular Therapy*, *27*(4), 735–746. https://doi.org/10.1016/j.ymthe.2019.01.014

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Singh Sandhu, K., Singh, U., Pant, V., Tiwari, V., Kurukuti, S., & Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics*, *38*(11), 1341–1347. https://doi.org/10.1038/ng1891

Zuin, J., Dixon, J. R., van der Reijden, M. I. J. A., Ye, Z., Kolovos, P., Brouwer, R. W. W., van de Corput, M. P. C., van de Werken, H. J. G., Knoch, T. A., van IJcken, W. F. J., Grosveld, F. G., Ren, B., & Wendt, K. S. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences*, *111*(3), 996–1001. https://doi.org/10.1073/pnas.1317788111